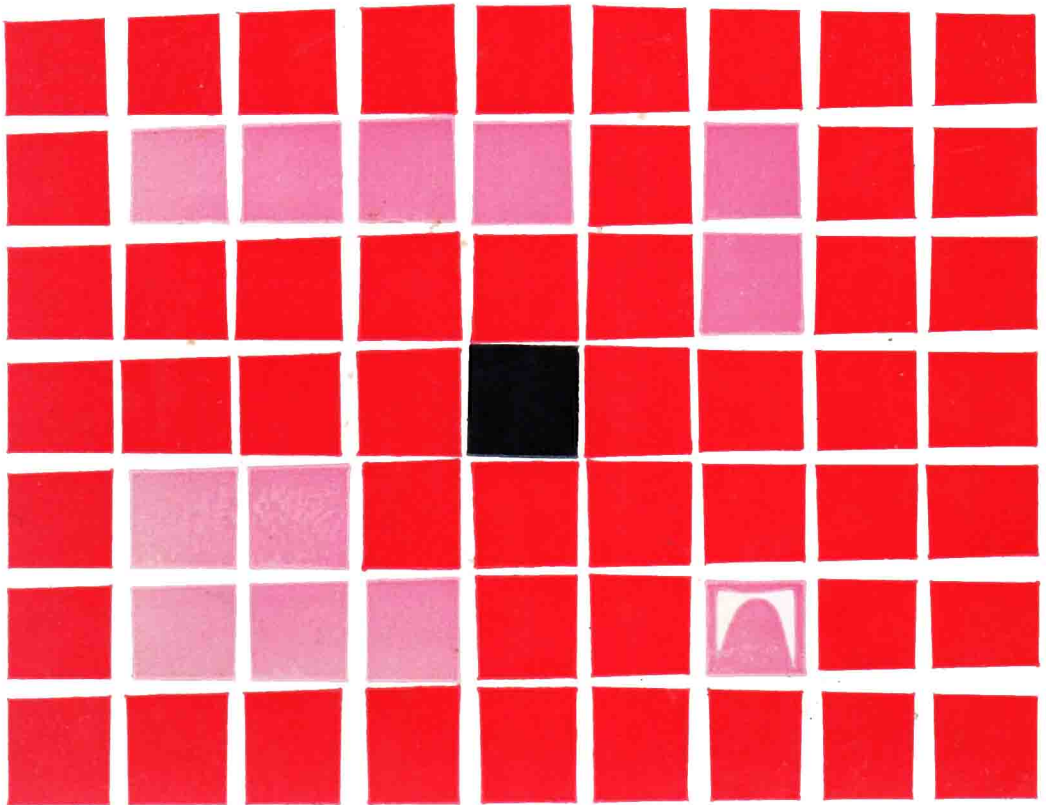


Introducción en la lingüística matemática

S. Marcus/E. Nicolau/S. Stati

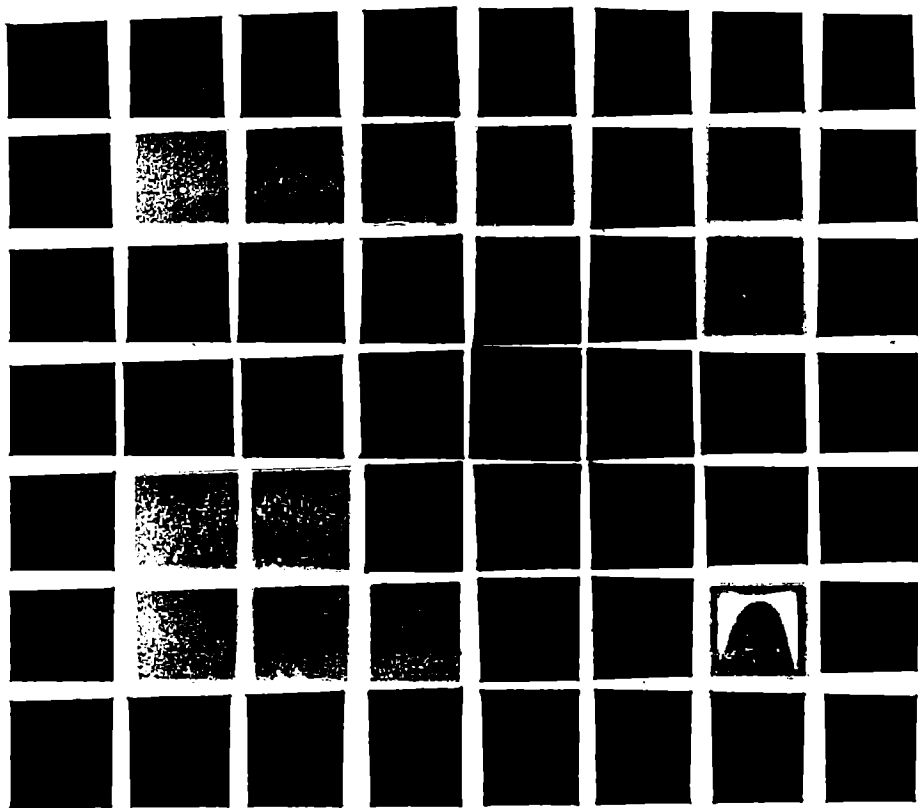
teide



Introducción en la lingüística matemática

S. Marcus/E. Nicolau/S. Stati

telde



INTRODUCCIÓN
EN LA
LINGÜÍSTICA MATEMÁTICA

Solomon Marcus
Profesor de la Universidad
de Bucarest

Edmond Nicolau
Profesor del Instituto
Politécnico de Bucarest

Sorin Stati
Profesor de la Universidad
de Bucarest

INTRODUCCIÓN en la LINGÜÍSTICA MATEMÁTICA

Versión castellana

Ramon Cerdà

Profesor Agregado de Lengua Española
Universidad de Barcelona

EDITORIAL TEIDE/BARCELONA



Primera edición en castellano: 1978

(Los autores han introducido, para la edición en castellano, cambios sustanciales y numerosas actualizaciones)

Título original:

Introducere în lingvistica matematică

Editura Științifică - București, 1966

Derechos de la edición en lengua castellana:

© Editorial Teide, S. A. - Viladomat, 291 - Barcelona-6

ISBN 84-307-7352-5

**Gràfiques Universitat, S. A. - Arquímedes, 3 - Sant Adrià de Besòs
Dipòsit Legal: B. 1978/78**

Prefacio

Igual que muchas otras disciplinas de reciente aparición, la lingüística matemática es una ciencia fronteriza, situada en la intersección de la lingüística con la matemática y la técnica. La finalidad del presente libro consiste en introducir al lector en la problemática general de la lingüística matemática sin penetrar en prolijidades técnicas demasiado minuciosas. En consecuencia, la obra se dirige no sólo a los lingüistas, ingenieros y matemáticos que deseen conocer las nuevas implicaciones de sus respectivas ciencias, sino también a todos los hombres de cultura.

El carácter heterogéneo de la lingüística matemática ha propiciado la colaboración de unos autores de formación diferente (un lingüista, un matemático y un ingeniero en cibernética). Por ello, los capítulos del libro resultan bastante diversos en su estilo y estructura y manifiestan una independencia notable. El lector puede insistir sobre unos capítulos y pasar más de prisa sobre otros de acuerdo con sus preocupaciones.

El capítulo 1 presenta una exposición sobre la lingüística estructural, disciplina cuya desarrollo se halla en la base de la lingüística matemática. En el segundo, se introducen las generalidades de la lingüística matemática. Cada uno de los capítulos subsiguientes reanuda y analiza con mayor profundidad uno de los aspectos o dominios sumariamente descritos en el capítulo 2. La repetición de ciertas cuestiones a lo largo del libro ha sido imprescindible justamente para revelar sus diferentes vínculos e implicaciones.

Los capítulos del libro se han redactado como sigue: el capítulo 1, por S. Statti; los capítulos 2, 3, 4 y 10, por S. Marcus; los capítulos 5, 6, 7, 9 y 11, por E. Nicolau, y el capítulo 8, por S. Marcus y S. Statti (§§ 1-13 y 14-16, respectivamente, 17 por ambos).

LOS AUTORES

marzo de 1964

NOTA A LA EDICION ESPAÑOLA

En la presente versión se han introducido diversas mejoras y actualizaciones con respecto no sólo a la edición rumana de 1966, sino también a la italiana de 1971 (Bologna, Ed. Pàtron). En especial, merecen mencionarse las adiciones incluidas en los capítulos 2 y 3 por Solomon Marcus y en los capítulos 7 y 9 por parte de Edmond Nicolau.

*Como traductor del texto rumano al español, el profesor Ramon Cerdà Massó de la Universidad de Barcelona ha realizado mucho más de lo que cabía requerir; gracias a una lectura competente y minuciosa, nos ha advertido sobre determinados puntos en que la exposición resultaba insuficiente. Por la dedicación y la elevada profesionalidad con que ha trabajado le debemos un particular reconocimiento. Agradecemos igualmente a la Editorial Teide por haber hecho posible la publicación de este libro.**

LOS AUTORES

Bucarest, marzo de 1976

* [Me limito a traducir fielmente el texto que los autores han redactado para esta versión. Sin embargo, el conocimiento que tengo de las ediciones anteriores me insta a subrayar que el esfuerzo y el interés de los autores ha sido muy superior a lo que ellos mismos dan, modestamente, a entender. En este punto, habría que consignar también la copiosa actualización del profesor S. Stati al capítulo primero. Todo ello se refiere únicamente a las adiciones de mayor entidad y no a las abundantísimas notas o modificaciones que se han esparcido a lo largo de toda la obra].

Prólogo a la versión española

“Si la pensée est fausse ou absurde, les symboles et les chiffres ne nous permettent pas de la rendre exacte et cohérente en elle-même: ils nous permettent seulement d’être d’une fausseté et d’une absurdité mathématiquement parfaites.”

COSERIU, “Critique”, 95.

Al proponer la presente versión, esperaba cubrir muchos más objetivos de los que pudieran inspirarse tan sólo en la experiencia docente de la Universidad de Barcelona, donde, por lo demás, poco hay de peculiar frente al resto universitario del país. Ahora, a la vista del trabajo en marcha, las esperanzas se robustecen, por fortuna, porque es incuestionable pensar ya en una utilidad más extensa, como la de una presentación recíproca entre estudiosos de lingüística, de informática, de ingeniería, de matemática, de crítica literaria... con el mejor ánimo de que cunda el ejemplo mismo de este libro.

Pienso que algunos aspectos de la problemática recién vertida, ya que no resuelta, ofrecen un interés suficiente para que los confiese aquí. En primer lugar, el título: Introducción en la lingüística matemática. Al margen del asunto preposicional (que, a lo sumo, suscitaría un pleito homeopático), el libro constituye sobre todo una auténtica y adecuada introducción en el tratamiento matemático de la lengua. Auténtica, porque los autores no-lingüistas se muestran bien conscientes de que se dirigen a un público específicamente “humanístico”, al que jamás intentan asombrar o atosigar. Con ello, logran la adecuación; pero también con el propósito fundamental de mostrar, por encima de cualquier otro cuidado, la profunda significación de la matemática en las ciencias actuales, marcando pautas y limitaciones sin dejar de referirse a la ciencia sustantiva del tratado, la lingüística. Nunca pierde el lector proclive a las humanidades la familiaridad de los planteamientos; por ello, se cumple con naturalidad el principal cometido del libro: el descubrimiento de sugerencias.

Un apartado, no menor, de la problemática general se caracteriza por el hecho de estar escrito el original en una lengua románica. Ello favorece, qué duda cabe, la comprensión de los ejemplos, puros pretextos, por lo común, para ilustrar las cuestiones teóricas. Allí donde era posible, el rumano ha quedado sustituido o acompañado por el español; y allí donde no lo era, se han volcado cuantiosas energías para evitar los cabos sueltos. Lo que no puede tolerarse es

que los ejemplos oscurezcan, en lugar de iluminar, las posibles opacidades de la teoría.

He intentado conservar a toda costa el ambiente original de la obra, aunque a veces, para conseguirlo, tuviese que recurrir a la paradójica medida de quitar, poner o modificar algún que otro punto. Ciertos retoques estilísticos son obligados para mantener la naturalidad de la exposición, ya que, pongamos por caso, el español se resiste más que el rumano a la iteración de términos idénticos de frecuencia media en espacios relativamente reducidos. Junto a esto, toda intervención enteramente mía queda limitada por paréntesis cuadrados.

He unificado la técnica tipográfica en la introducción de conceptos, ejemplos y palabras especiales a base de una administración coherente de espaciados, cursivas y comillas. Los párrafos, al numerarse doblemente, permiten rápidas referencias internas, y la bibliografía, reunida al final, va siendo citada por envíos abreviados al pie de página. No era cuestión, evidentemente, de interpolar una prolija hermenéutica matemática, pero, en algunos casos, me he permitido aclarar determinados signos, e incluso nociones más o menos básicas, que, a mi juicio, podían entorpecer un poco la fluidez de la lectura. De otro modo, bastará una consulta a la tabla de signos. Confío en que así se habrán cubierto, al menos, las dificultades más típicas con que suele tropezar el lector profano en cálculos lógicos, aun cuando no quede excluida la necesidad de actuar sobre ciertos párrafos con las entendederas mucho más en vilo que sobre otros. Los títulos escritos en rumano, ruso o checo van acompañados de traducción. Para transliterar el ruso, he adoptado los criterios de J. Calonge, Transcripción del ruso al español, Madrid, 1969.

He tenido a la vista la espléndida traducción italiana de Filippo Franciosi (Bologna, 1971), así como las anotaciones magistrales de Carlo Tagliavini. A ambos les debo algunas sugerencias, pero, en general, he aplicado una interpretación más libre que la de ellos sobre la fidelidad a la letra impresa.

Mención aparte merecen los autores, cuya prodigalidad en atenciones ha convertido en goce toda la tarea. En primer lugar, por el cariño que han dispensado a la presente versión, introduciendo cambios, actualizaciones y muy sustanciales mejoras por doquier. Y, también, porque han soportado con absoluta entereza —sin duda, por comunión de entusiasmo— las numerosas solicitudes que les he formulado en diversos momentos.

Y a Mónica, mi mujer, supremo diccionario en esta ocasión, debo no sólo el consejo y la seguridad que he recibido con su respaldo de hablante nativa, sino también la peculiar eficacia en depararme un sosiego indispensable para llevar adelante el empeño.

Tabla de símbolos

Signos fonéticos¹

i	esp. <i>sí</i> [si]; rum. <i>fi</i> [fi] (ser, estar).
e	esp. <i>té</i> [te]; rum. <i>pe</i> [pe] (a, sobre).
a	esp. <i>mar</i> [mar]; rum. <i>cal</i> [kal] (caballo).
o	esp. <i>sol</i> [sol]; rum. <i>dor</i> [dor] (añoranza).
u	esp. <i>su</i> [su]; rum. <i>tu</i> [tu] (tú).
î	rum. <i>riu</i> [riũ] (río); ruso <i>mî</i> [mî] (nosotros).
ə	rum. <i>rău</i> [rəũ] (malo); ing. <i>alone</i> [əlóũn] (solo).
ï	esp. <i>baile</i> [báïle], <i>cauto</i> [káũto]; rum. <i>fier</i> [fier] (hierro), <i>soare</i> [soãre] (sol). Toda vocal con este diacrítico es primer o segundo elemento de grupo vocálico en diptongo.
p	esp. <i>paz</i> [pəθ]; rum. <i>pur</i> [pur] (puro).
b	esp. <i>ámbar</i> [ámbar]; rum. <i>bun</i> [bun] (bueno).
f	esp. <i>fe</i> [fe]; rum. <i>fir</i> [fir] (hilo).
v	rum. <i>viu</i> [viũ] (vivo); fr. <i>verre</i> [ver] (vaso).

1. Los signos en transcripción, basados en el Alfabeto Fonético Internacional (cf. *The Principles*), vienen aclarados, allí donde es posible, sólo por vía de ejemplo mediante dos lenguas asequibles al texto. A pesar del cuidado en evitarlo, la correspondencia es a menudo aproximada (o, si se quiere, muy aproximada) entre los elementos gráficos que resultan de la intersección de cada pareja de ejemplos transcritos. En la lista, además, sólo se exponen los signos que aparecen en el libro, lo que no obsta para que a lo largo de las ejemplificaciones surjan otros inéditos sin ulterior aclaración.

m	esp. <i>mal</i> [mal]; rum. <i>mic</i> [mik] (pequeño).
t	esp. <i>tal</i> [tal]; rum. <i>tot</i> [tot] (todo, también).
d	esp. <i>andar</i> [andár]; rum. <i>cod</i> [kod] (código, bacalao).
θ	esp. <i>luz</i> [luθ]; ing. <i>thin</i> [θɪn] (delgado).
ð	esp. <i>codo</i> [kòðo]; ing. <i>then</i> [ðɛn] (entonces).
n	esp. <i>no</i> [no]; rum. <i>nuc</i> [nuk] (nogal).
s	esp. <i>sal</i> [sal]; rum. <i>suc</i> [suk] (jugo).
z	rum. <i>zero</i> [zéro] (cero); ing. <i>zip</i> [zɪp] (zumbar).
ʃ	rum. <i>și</i> [ʃi] (y, también); fr. <i>cher</i> [ʃɛr] (caro, querido).
ʒ	rum. <i>juca</i> [ʒuká] (jugar, bailar); fr. <i>jamais</i> [ʒamé] (nunca).
c	esp. <i>chopo</i> [còpo]; rum. <i>ceai</i> [caj] (té).
ʒ	rum. <i>geam</i> [ʒam] (cristal); ing. <i>gin</i> [ʒɪn] (ginebra).
y	esp. <i>mayo</i> [máyo]; rum. <i>cheie</i> [kéiye] (llave).
k	esp. <i>casa</i> [kása]; rum. <i>cînd</i> [kînd] (cuando).
g	esp. <i>inglés</i> [inglés]; rum. <i>gaz</i> [gaz] (gas).
x	esp. <i>caja</i> [káxa]; rum. <i>hotel</i> [xotél] (hotel).
l	esp. <i>la</i> [la]; rum. <i>leu</i> [leu] (león).
r	esp. <i>coro</i> [kóro]; rum. <i>mare</i> [mâre] (mar, grande).
m ⁷	rum. <i>pomi</i> [pom ⁷] (árboles), <i>lucrezi</i> [lukrész ⁷] (trabajos). Símbolo fonético que expresa una oposición fonológica en el consonantismo rumano (cf. en ruso) en posición final entre mojado (o diesizado) y no-mojado (no-diesizado) y que comporta, por lo común, una oposición morfológica entre plural y singular (cf. <i>pom</i> [pom] (árbol)) en los nombres masculinos y entre segunda y primera persona singular (cf. <i>lucrez</i> [lukrész] ((yo) trabajo)) en los verbos.

Signos gráficos especiales (en rumano)

î	[i] <i>mîine</i> [mîine] (mañana).
ă	[ə] <i>păr</i> [pər] (pelo, peral).
c ^{e.i}	[c] <i>ciur</i> [cur] (colador).
ch ^{e.i}	[k] <i>chema</i> [kemá] (llamar).
g ^{e.i}	[ʒ] <i>gem</i> [ʒem] (mermelada).
gh ^{e.i}	[g] <i>ghem</i> [gem] (ovillo).

- j [ʒ] *jura* [ʒurá] (jurar).
 § [ʃ] *şinã* [ʃinə] (rail).
 † [ts] *ţigan* [tsigán] (gitano).

Otros signos (se indica sólo la primera página en que aparecen).

- * *genuculus; forma no documentada (p. 5).
 *la hombre; expresión no gramatical o no aceptable (p. 14).
 { 1, * }; símbolo auxiliar de un alfabeto (p. 160).
- $X \rightarrow Y$; rescribir X como Y (p. 35).
 $x \rightarrow y$; x implica y (p. 107) (cf. \supset).
 $2 \rightarrow 4$; transferir el resultado de la operación 2 a la operación 4 (p. 174).
 $i + 1 \rightarrow i$; sustituir todos los índices i por los índices $i + 1$ (p. 180).
- \nrightarrow $a \nrightarrow b$; a no implica b (p. 156).
- ↑ $C \uparrow$; indicador de transferencia entre operadores lógicos (p. 145).
 $h \uparrow 2$; elevar h a la potencia de 2, h^2 (p. 189).
- ↓ $\downarrow x(x \in A)$; el x que pertenece al conjunto $A \dots$ (p. 124).
 $\downarrow O$; indicador de transferencia entre operadores lógicos (p. 145).
- \leftrightarrow $x \leftrightarrow y$; implicación bilateral (p. 125) (cf. \equiv).
- \Rightarrow $x \stackrel{H}{\Rightarrow} y$; derivación directa a través de una matriz (p. 96).
- \in $a \in B$; pertenencia a un conjunto (p. 84).
- \notin $a \notin B$; no pertenencia a un conjunto (p. 84).
- \subset $A \subset B$; A es un subconjunto del conjunto B (p. 84).
- \supset $A \supset B$; implicación lógica (p. 187).
- \equiv $A \equiv B$; implicación bilateral (p. 187).
- \neg $\neg p$; negación lógica: no- p (p. 123).
- \cup $A \cup B$; unión de conjuntos (p. 86).

XII TABLA DE SIGNOS

- \cap $A \cap B$; intersección de conjuntos (p. 89).
- \vee $p \vee q$; disyunción lógica (p. 123).
 $\vee x$; hay un tal x en particular... (p. 124).
- \wedge $p \wedge q$; conjunción lógica (p. 123).
 $\wedge x$; todo x en general... (p. 124).
- \sqsupset Fit $\lceil pxq \rceil$; ocurrencia del evento pxq , donde x es una variable (p. 127).
 $y \in \lceil x \rceil \leftrightarrow x = y$; transformación en un conjunto de un solo elemento (p. 200).
- $>$ $a > b$; a es mayor que b (p. 186).
- \geq $a \geq b$; a es mayor o igual que b (p. 176).
- $<$ $a < b$; a es anterior, en una cadena, a b (p. 85).
 $a < b$; a es menor que b (p. 97).
- \leq $a \leq b$; a es anterior o concomitante, en una cadena, a b (p. 85).
 $a \leq b$; a es igual o menor que b (p. 97).
- $/$ a/b ; oposición entre dos términos (p. 13).
 a/b ; combinación sintáctica a la derecha (p. 107).
- \backslash $a \backslash b$; combinación sintáctica a la izquierda (p. 107).
- \emptyset $A \cap B = \emptyset$; conjunto vacío (p. 89).
- $=$ $a = b$; igualdad (p. 19).
- \neq $a \neq b$; desigualdad (p. 89).
- $+$ $a + b$; concatenación en una secuencia (p. 18).
 $a + b$; adición (p. 46).
- $-$ $-a$; variable negativa (p. 176).
- \approx $a \approx b$; a equivale aproximadamente a b (p. 274).
- \vdash $a \vdash b$; intervalo entre dos miembros de una ecuación (p. 183).
- $|$ $|a|$; módulo de una variable (p. 176).

- Σ $\Sigma = a_n \cdot x$; suma indicada (p. 46).
 $\sum_{n=1}^{\infty} \frac{1}{n}$; orden de suma para una serie de valores (p. 240).
- Π $\prod_{i=1} A_i$; orden de multiplicación para una serie de valores (p. 185).
- \int cálculo integral (p. 334).
- ! C!; final de una sucesión de cálculo (p. 185).
- ? ?1x; presentación de una incógnita (p. 124).
- # #ab#; límite inicial y final de un texto dado (p. 194).

1. La lingüística estructural

1.1. Orígenes de la lingüística

La preocupación por el estudio de la lengua nació en la Antigüedad, concretándose al principio en forma de obras gramaticales compuestas por los indios desde el siglo v al iv antes de nuestra Era, y de ciertas ideas y teorías formuladas por algunos grandes filósofos griegos¹.

En la India, las primeras investigaciones lingüísticas respondían al propósito de explicar los textos de los himnos védicos (compuestos, probablemente, entre los siglos xviii y xv antes de n. E.), en los cuales el lector indio tropezaba siempre con la dificultad de comprender determinadas palabras y formas gramaticales arcaicas y fuera de uso. La lengua *sánscrita*, en la que estaban escritas las grandes epopeyas del *Rāmāyana* y *Mahābharata*, era relativamente extraña a las lenguas indias habladas alrededor de los años 500 y 400 a. de n. E., y requería aclaraciones. La más importante obra de gramática india pertenece a Pānini (s. iv a. de n. E.).

En la antigua Grecia, las preocupaciones lingüísticas tuvieron desde el principio un pronunciado carácter especulativo en el marco de dos problemas teóricos fundamentales: *a*) la naturaleza del lenguaje, el origen de los nombres de las cosas y la relación entre palabras y cosas; y *b*) la relación entre lengua y pensamiento, y la estructura lógica del lenguaje. Con relación a estos problemas cabe citar aquí a filósofos como Heráclito, Demócrito, Platón, Aristóteles y a los sofistas, estoicos y epicúreos, sobre todo. Pero también en Grecia nacieron otras actividades consagradas al estudio de la lengua como consecuencia de necesidades prácticas, bien de cara al adiestramiento de los oradores (recuérdese la importancia de la retórica en la Antigüedad), o bien con objeto de explicar

1. Cf. GRAUR-WALD, *Istorie*; ROBINS, *Breve historia*; TAGLIAVINI, *Panorama*; [MOUNIN, *Historia*, y LE ROY, *Grandes corrientes*].

la lengua de las epopeyas homéricas, especialmente estudiada por los filósofos de Alejandría, entre los siglos III y II a. de n. E.

Durante los siglos que siguieron a esta etapa primitiva, y hasta el s. XIX, en que la lingüística se constituye en ciencia autónoma, las investigaciones sobre la lengua se orientaron en tres direcciones.

1) La primera consiste en la descripción de una lengua, en un determinado período de su historia. Al principio la atención se concentró sobre la lengua escrita de los textos literarios, hasta que luego la lengua hablada penetró también en la esfera de las preocupaciones gramaticales. (El primer trabajo destacado de este tipo pertenece a un autor anónimo [al parecer] del siglo III y es comúnmente conocido por el nombre de *Appendix Probi*. En él se consignan ciertas diferencias entre el latín literario y el latín hablado, denominado "latín vulgar".)

Como la descripción presupone, a su vez, una agrupación de los fenómenos de la lengua, su estudio dio lugar a tres ramas científicas: la fonética, la gramática y la lexicología (estudio de las palabras)². En lo que concierne a la fonética, conviene señalar que hasta la época moderna (s. XIX), los trabajos más detallados y rigurosos se debieron a los antiguos indios (cuya obra sirvió de modelo a los romanos y, en la Edad Media, a los árabes). Ellos clasificaron los sonidos desde el punto de vista de la articulación, estudiaron las influencias recíprocas de los sonidos en la cadena hablada, la sílaba, e intuyeron la noción de *fonema*, al que dieron la denominación de "sphota". Para los estudiosos de la Antigüedad (indios y, en especial, griegos y romanos) es característica una cierta confusión entre sonidos y letras.

Por lo que se refiere a la gramática, apareció la distinción entre morfología y sintaxis (sólo el segundo de ambos términos pertenece a la antigüedad)³, y se concedió, entre griegos y romanos, una atención especial al establecimiento de modelos de declinación y de conjugación (los llamados "paradigmas") y a la distinción entre formas y construcciones "correctas" e "incorrectas" (de ahí el carácter normativo de las antiguas gramáticas y de las que se han compuesto posteriormente a partir de aquel modelo). Como ilustración al punto de vista lógico con que se abordaban los problemas de la descripción lingüística cabe aducir las clasificaciones de palabras en partes del discurso. Así, Platón distingue el *nombre* (palabra que puede figurar como sujeto de un juicio) del *verbo* (palabra que puede ser predicado de un juicio). Más tarde, los filósofos estoicos diferenciaron cinco partes del discurso: nombre propio, nombre común, verbo, conjunción y artículo. Otra clasificación, muy utilizada incluso recientemente, pertenece al filólogo alejandrino Dionisio Tracio (s. II a. de n. E.): sustantivo, adje-

2. Estos nombres corresponden a la moderna terminología.

3. Hasta hace poco, se han utilizado los términos "analogía" o "etimología", en lugar de "morfología".

tivo, pronombre, numeral, artículo, verbo, adverbio, preposición, conjunción e interjección. (Para Dionisio Tracio, el sustantivo y el adjetivo formaban una sola parte del discurso, mientras existía, como categoría aparte, el participio; la interjección fue añadida por los gramáticos latinos.)

Los estudios sobre el vocabulario dieron lugar a unas listas de palabras (llamadas glosarios) que incluían términos arcaicos, fuera de uso, o dialectales, y a la redacción de obras consagradas a los homónimos y sinónimos (en ambos casos, el mérito corresponde a los griegos del período clásico y helenístico). Durante varios siglos, en Grecia y en Italia, numerosos filósofos y filólogos tomaron parte en una polémica que ha alcanzado la celebridad: la forma sonora de las palabras está condicionada por las cualidades de las cosas aludidas por ellas (teoría “naturalista”), o bien, por el contrario, no tiene relación con las cosas, es arbitraria y se ha convertido en su simbolización como consecuencia de una convención establecida entre los hombres (teoría “convencionalista”).

2) La segunda dirección, que asumieron con preferencia los estudios de la lengua hasta la constitución de la lingüística científica, consiste en la adopción de una perspectiva histórica.

La idea de que las lenguas evolucionan y cambian apareció también como resultado de los estudios filológicos, los cuales, en la tarea de explicar los textos literarios, debían comparar las formas arcaicas con las contemporáneas. Teniendo en cuenta la movilidad de la lengua, los filósofos estoicos sostuvieron que, al principio, el lenguaje se componía de palabras cuya apariencia fonética estaba determinada por las cosas denominadas por ellas y que, con el tiempo, esta relación se había perdido. Al gramático romano Varrón (116-27 a. de n. E.) se deben las primeras formulaciones claras sobre la evolución continua de la lengua.

Durante 1800 años el estudio evolutivo de la lengua no experimentó progresos destacables, con lo que permaneció muy atrás con respecto al desarrollo de la lingüística descriptiva. El único dominio en cuyo estudio se utilizó más el punto de vista histórico con resultados apreciables fue el de la etimología, que ya había interesado a griegos y romanos. Sin embargo, la mayoría de las explicaciones que éstos dieron sobre el modo de evolucionar los sonidos y los significados de las palabras consideradas resulta ingenua o fantástica.

3) La tercera dirección viene determinada por el uso de un criterio comparativo. La Antigüedad ignoró casi por completo el problema de las relaciones entre las lenguas, de sus semejanzas y diferencias; la idea misma de que existen lenguas emparentadas no aparece ni siquiera en forma embrionaria. Algunas observaciones de Varrón constituyen un caso excepcional, ya que traspasan los límites del latín y se aplican, entre otras, a las lenguas etrusca y griega. De la comparación entre el vocabulario latino y el griego, Varrón llega a la con-

clusión, importante y fecunda, de que las palabras pasan prestadas de un pueblo a otro.

Hasta los siglos xvii y xviii no aparecen aplicaciones destacables del punto de vista comparativo. Abordando los problemas lingüísticos desde un prisma lógico —con arreglo a una concepción procedente, como hemos dicho, de los filósofos griegos—, una serie de estudiosos, especialmente franceses, postuló que la gramática es la forma de expresión de las reglas de la lógica. Y como éstas son idénticas a todos los pueblos, los distintos idiomas no constituyen más que aplicaciones particulares a unos principios universales e inmutables. En consecuencia, se pueden construir gramáticas válidas para todas las lenguas sobre criterios *racionales o universales*. Como modelo se utilizó la famosa gramática de Port Royal (1660), basada en la comparación de las lenguas griega, latina, hebrea, francesa, italiana, española, alemana e inglesa⁴.

Otra aplicación importante, científicamente más exacta y fructífera por el desarrollo y perfeccionamiento ulterior, deriva del estudio comparativo fundamentado en el supuesto de que existen lenguas emparentadas, es decir que provienen de una fuente común que se ha diversificado con el tiempo. Esta idea nació por la necesidad de explicar las semejanzas que hay, por ejemplo, entre las lenguas eslavas (comparadas por el sabio ruso Lomonosov) o entre el latín, el griego y el sánscrito (como observó, entre los primeros, el jesuita francés Coeurdoux).

1.2. Constitución de la lingüística científica

La lingüística científica nace durante el primer cuarto del siglo xix con la creación del método histórico-comparativo⁵. Viene a ser el resultado de una síntesis de los datos obtenidos en estudios anteriores, en la cual se hermana el punto de vista descriptivo con el histórico y comparativo: las *d e s c r i p c i o n e s* cada vez más minuciosas (en forma de gramáticas y diccionarios) de las lenguas antiguas y modernas de Europa y Asia confirman el supuesto de que las lenguas *e v o l u c i o n a n* constantemente, mientras la *c o m p a r a c i ó n* entre lenguas distintas demuestra que algunas se parecen notablemente entre sí. La descripción, combinada con el estudio histórico, lleva a la conclusión de que estas hablas semejantes no son más que el resultado del proceso de diversificación de otras lenguas más primitivas, conocidas y analizadas a partir de los textos.

En la base de esta síntesis se hallan dos principios utilizados hasta la actualidad en la investigación lingüística: 1) la apariencia sonora de las palabras no

4. [Cf. DONZÉ, *Gramática*].

5. Cf. GRAUR-WALD, *Istorie*; TAGLIAVINI, *Panorama*; [además de las adiciones a la n. 1, MALMBERG, *Nuevos caminos*].

viene determinada por el significado que expresan (en la terminología moderna este principio recibe el nombre de “arbitrariedad del signo lingüístico”)⁶. Con arreglo a esto, en cuanto comprobamos, en dos o más lenguas, que el mismo significado se realiza por medio de formas relativamente semejantes, esta semejanza se debe bien al hecho de que tales lenguas tienen un “antepasado” común, o bien al préstamo entre ellas de las palabras respectivas. La elección de una explicación u otra se basa en el segundo principio, o sea 2): la regularidad de los cambios fonéticos. Este principio indica que, por lo común, un determinado sonido se modifica del mismo modo en todas las palabras donde ocupe la misma posición (sea inicial, final, intervocálica, interconsonántica, acentuada o no, etc.). Por ejemplo, *l* latina cambia en *r* en rumano siempre que se halle precedida y seguida de vocal (lat. *sole*, ‘sol’, pasa a rum. *soare*; *sale*, ‘sal’, a *sare*, etc.). En toda lengua existen reglas específicas para los cambios fonéticos: así, *i* latina situada en principio de palabra y seguida de vocal se transforma en *j* [ʃ] para el rumano, *gi* [ʒ] para el italiano y *j* [x] para el español (lat. *iocari* pasa a rum. *jucare*, it. *giocare*, esp. *jugar*, igual que lat. *iudicare* pasa a rum. *judecare*, it. *giudicare*, esp. *juzgar*)⁷. De ahí que cuando las palabras comparadas respetan estas reglas específicas de evolución lleguemos a la conclusión de que están emparentadas, y en cambio, si no cumplen las reglas, es probable que nos hallemos ante un caso de palabras prestadas.

El descubrimiento de las reglas de correspondencia entre los sonidos de las lenguas emparentadas (por ejemplo, que *j* rumana corresponda a *gi* italiana y *j* española, y que las tres correspondan a *i* inicial seguida de vocal en latín) posibilita la reconstrucción de palabras desaparecidas, no documentadas en textos. Así, pongamos por caso, aunque no conociéramos la palabra latina *nocte*, ‘noche’, podríamos presumir su existencia comparando entre sí las formas *noapte* del rumano, *notte* del italiano y *noche* del español y aplicando la regla de correspondencia entre el rumano *pt*, el italiano *tt* y el español *ch* con el latín *ct*, como lo atestiguan numerosas equivalencias del tipo rum. *opt*, it. *otto*, esp. *ocho* frente al latín *octo*; rum. *fapt*, it. *fatto*, esp. *hecho*, lat. *factum*, etc.

Cuando una palabra latina no se encuentra documentada, los lingüistas la reconstituyen a base de las reglas de correspondencia fonética partiendo de las formas románicas. Por ejemplo, los textos latinos no nos han conservado la palabra *genuculus*, ‘rodilla’, pero presuponemos su existencia basándonos en *genunchi* del rumano, *genou* del francés, *ginocchio* del italiano [o *genoll* del catalán]. Las formas así reconstruidas se escriben precedidas del signo * (asterisco): **genuculus*. La exactitud de la reconstrucción —y, por tanto, la validez del método— ha sido confirmada a veces gracias al descubrimiento, en los textos lati-

6. Cf. SAUSSURE, *Curso*.

7. [Para la transcripción fonética (AFI), consultar la Tabla de símbolos. También *The Principles*].

nos, de la palabra reconstruida por los lingüistas. Así ha ocurrido con *muttum*, reconstruido por comparación de las formas respectivas del francés, provenzal y catalán (*mot*, 'palabra'). Ahora *muttum* se escribe, pues, sin asterisco.

En la segunda mitad del siglo XIX, el método histórico-comparativo fue empleado de un modo casi exclusivo para deducir, por comparación de las lenguas indoeuropeas (latín, griego antiguo, sánscrito, armenio clásico, antiguo eslavo, etcétera), los sonidos y las formas de la lengua originaria de todos estos idiomas emparentados, para el conocimiento de la cual no disponemos de ningún texto. La principal utilidad del método reside, sin embargo, en el hecho de haber estimulado la investigación histórica de las lenguas y de haber permitido la explicación científica de su evolución.

Junto al nombre de Franz Bopp, que publicó en 1816 la primera obra de gramática histórico-comparativa⁸, conviene mencionar a lingüistas como J. Grimm, Fr. Diez, W. von Humboldt, A. Schleicher y a los representantes más ilustres de la escuela neo-gramática (H. Osthoff, K. Brugmann, H. Paul, W. Meyer-Lübke), a la que se debe un considerable perfeccionamiento del método histórico-comparativo. Su actividad se extiende durante más de un siglo. Y gracias a la obra de lingüistas contemporáneos como A. Meillet (1866-1936), J. Kuryłowicz (n. 1895) y E. Benveniste (n. 1902), el estudio de las lenguas desde este punto de vista accede a un grado superior.

Cabe decir que el método histórico-comparativo domina los estudios lingüísticos aproximadamente desde 1816 (año en que apareció el libro de Fr. Bopp) hasta 1916 (cuando se publica el *Curso de lingüística general* de F. de Saussure). Aunque ciertos trabajos importantes aparecidos durante este intervalo adoptan un criterio básicamente descriptivo, la nota dominante de estos cien años corresponde al *historicismo*, en que la atención de los especialistas se ha dedicado a la explicación, lo más detallada y exacta posible, sobre el modo como han evolucionado las lenguas emparentadas. La descripción de la lengua en un determinado período, por ejemplo de las lenguas modernas en su estado actual, se realizaba —y, a veces, aún se realiza— a través de su historia respectiva, es decir en función de estados anteriores ampliamente superados. Así, en numerosas gramáticas de la lengua francesa se habla de declinación y de casos para los sustantivos (nominativo *le père*, 'el padre', genitivo *du père*, dativo *au père*, acusativo *le père*) por el mero hecho de que en la flexión nominal latina existían declinaciones y casos y, en menor grado, en francés antiguo. En realidad, en francés moderno los nombres ya no flexionan de acuerdo con el caso ni en singular ni en plural, sino que van caracterizados por una forma única respectiva: p. ej., *cheval*, 'caballo'/*chevaux*, 'caballos', coincidiendo a menudo el singular y el plural, aunque se escriban de un modo diferente: *père/pères*, 'padre'/'padres'.

8. Titulada *Sobre el sistema de la conjugación sánscrita en comparación con el de las lenguas griega, latina, persa y alemana*.

En esta presentación sumaria e inevitablemente incompleta de la lingüística histórica hasta la aparición del estructuralismo, conviene mencionar todavía dos métodos de investigación que han allegado resultados particularmente valiosos:

1) La *geografía lingüística*, es decir el método de estudio de las variantes de una lengua con la ayuda de mapas: las ramificaciones territoriales (dialectos, subdialectos, hablas) se investigan, habitualmente, a base de encuestas en los mismos lugares, y luego se confeccionan mapas del área estudiada sobre los cuales se anota la distribución geográfica de las palabras, de la pronunciación y de las formas gramaticales. Las ventajas de este procedimiento favorecen a la lingüística descriptiva y, en especial, a la lingüística histórica. El suizo J. Gilliéron (1854-1926) está considerado como el creador del método. En rigor, el procedimiento de cartografiar los fenómenos de la lengua ya fue utilizado antes de él, pero ello sin que la representación en mapas de las ramificaciones lingüísticas llegaran a cuajar en una concepción o en una teoría como la que elaboró el mismo Gilliéron a partir del Atlas Lingüístico de Francia, que compuso y publicó durante el primer decenio del siglo actual⁹.

2) La *fonética instrumental*, es decir el método de investigación de los sonidos con ayuda de aparatos. También nació a principios del presente siglo y ha progresado constantemente a medida que se han perfeccionado los instrumentos puestos a disposición de los técnicos. El estudio detallado del modo como se producen los sonidos (fonética articuladora o genética) y de sus cualidades acústicas (fonética acústica) —inestimable auxiliar de la lingüística descriptiva— ha demostrado que la misma palabra no es pronunciada exactamente igual por dos hablantes y que ni siquiera el mismo hablante pronuncia dos veces del mismo modo una misma palabra (cf. 1.7). Esta conclusión debe considerarse como un poderoso estimulante para la creación de una nueva rama de la fonética (la fonética funcional o fonología), cuya aparición coincide con el nacimiento de la lingüística estructural.

1.3. Breve historia de la lingüística estructural

La publicación, en 1916, del *Curso de lingüística general* de Ferdinand de Saussure ha señalado, para la ciencia de la lengua, el comienzo de una nueva etapa. Naturalmente, esta fecha tiene un valor relativo, ya que las teorías básicas del lingüista suizo no pertenecen al mérito exclusivo de una sola persona. El terreno había sido propiciado ya desde finales del siglo precedente para este cambio de perspectiva, que se caracteriza por dos rasgos esenciales:

9. Inspirados en el de Gilliéron, aparecen en España atlas lingüísticos como el *Atlas Lingüístico de Catalunya* (ALC) de GRIERA y el *Atlas Lingüístico de la Península Ibérica* (ALPI) de MENÉNDEZ PIDAL-NAVARRO TOMÁS.

1) La lengua no es un conglomerado de elementos, sino un sistema de elementos solidarios. Desde el punto de vista metodológico, esto implica el desplazamiento del centro de gravedad en el análisis lingüístico de los elementos a las relaciones entre los elementos. Como en otras ciencias, también en la lingüística toma cuerpo la concepción según la cual el objeto de estudio es una estructura compleja, en la que el conjunto es más que la suma de las partes. Esta nueva orientación, teorizada por Saussure, da lugar a una corriente lingüística, de aspecto variado y a veces contradictorio, conocida por el nombre de *estructuralismo*¹⁰.

2) El segundo rasgo esencial, que deriva lógicamente del primero, consiste en que el estudio de la lengua como estructura compleja no es posible si no se hace abstracción del factor evolutivo. Para descubrir el conjunto de relaciones, hay que considerar la interdependencia de los elementos en un momento dado (donde "momento" significa, de hecho, un periodo caracterizado por la existencia de un modo determinado de organización, relativamente estable, de las unidades lingüísticas). Sin duda, las lenguas se hallan en continuo movimiento y en evolución, pero su mecanismo funcional no puede captarse si no es considerándolo como estructuras de términos coexistentes. Desde el punto de vista metodológico, también aquí cabe hablar de un desplazamiento del centro de gravedad en el análisis del fenómeno lingüístico, que va precisamente del historicismo a una tendencia hacia la descripción.

Tal como decíamos antes, la aparición del estructuralismo no se explica sólo como un eco de las teorías de Saussure, sino como consecuencia de unas investigaciones que, en cierto modo, prepararon el terreno para aquel cambio de orientación. Queremos referirnos a la aportación de los fundamentos de una nueva rama fonética, la fonología, o estudio de la función de los sonidos en el proceso comunicativo. A este respecto, el mérito corresponde en primer lugar a la escuela de Kazan y a su principal representante, Baudoin de Courtenay. Él estableció la distinción entre sonido y fonema y mostró que cada fonema se realiza en el habla a través de muchas variantes, concebidas por los hablantes como una sola unidad sonora desde el momento en que aparecen en el mismo morfema.

En la evolución del estructuralismo destacamos otras escuelas, entre las que existen bastantes diferencias tanto en teoría como en problemática. Todas ellas, sin embargo, se vinculan por los dos rasgos esenciales que hemos mencionado más arriba, a los que, en todo caso, no se concede una importancia siempre igual. Las más significativas son la escuela de Praga, el descriptivismo americano y la glosemática, a las cuales se añade, en los dos últimos decenios, una variante del estructuralismo, conocida por el nombre de "escuela generativista" o

10. Cf. APRESIAN, *Principles y Éléments*; LEPSCHY, *Lingüística* y MALMBERG, *Nuevos caminos*.

“transformativista”, que se opone a las demás variantes del estructuralismo, reunidas, así, bajo la denominación genérica de “estructuralismo clásico”.

La escuela de Praga apareció el año 1926 gracias a sus fundadores los lingüistas V. Mathesius, N. S. Trubetsky y R. Jakobson¹¹. Por su obra, los miembros de la escuela praguense pueden considerarse los padres de la fonología. El aspecto sonoro del signo lingüístico fue el primer compartimento de la lengua estudiado con una perspectiva estructural y sus resultados fueron extraordinariamente fructíferos; con algunos cambios y adiciones mantienen todavía su validez. Los estructuralistas de Praga expusieron e ilustraron especialmente la idea de que la lengua consiste en un sistema de sistemas (fonológico, morfológico, sintáctico) y, al menos en teoría, mostraron la necesidad del estudio histórico junto al descriptivo, combatiendo así a Saussure, que consideraba que el sistema es inmutable. Afirmaban que la evolución de la lengua puede comprenderse mejor si no se intenta explicar cada cambio por separado (como habían hecho, generalmente, sus predecesores), sino planteando las causas y los efectos de los cambios con relación al sistema de la lengua. En este espíritu está concebida, por ejemplo, la obra fundamental de A. Martinet¹².

El descriptivismo americano se ha desarrollado principalmente después de la publicación del libro de L. Bloomfield, *Language* (1933), por medio de sus principales representantes Z. S. Harris, Ch. F. Hockett, K. L. Pike y E. A. Nida¹³. Su principal mérito estriba en la elaboración de *gramáticas* estructurales mediante la aplicación al estudio de la morfología y de la sintaxis de los métodos verificados en la fonología (p. ej., la conmutación, la reducción de las variantes a invariantes) o de métodos nuevos (análisis en constituyentes inmediatos). Una rama especial del descriptivismo, representada sobre todo por Z. S. Harris, promueve la descripción de las unidades de la lengua sin apelar al sentido. Una tal lingüística asemántica ha encontrado justificación en una serie de cometidos prácticos, especialmente en la traducción automática (donde, en realidad, no cabe pretender que una máquina destinada a encontrar los equivalentes entre dos lenguas “comprenda” el texto que debe traducir).

Glosemática es la denominación que dio a la escuela estructuralista danesa su propio creador, L. Hjelmslev, y su discípulo y colaborador H. Uldall. La concepción de la escuela de Copenhague se expone en una obra básica, cuya versión original en danés aparece en 1943¹⁴. Los dos rasgos esenciales del estructuralismo aparecen en su obra mucho más destacados que en la de cualquier otro inves-

11. Cf. TRUBETSKOJ, *Principios*; VACHEK, *School of Prague* [y TRNKA et al., *Círculo*].

12. Cf. MARTINET, *Economía*.

13. Cf., sobre todo, HARRIS, *Structural Linguistics*; HOCKETT, *Course*; PIKE, “Taxemes” y NIDA, *Morphology*.

14. Cf. HJELMSLEV, *Prolegómenos*.

tigador. La glosemática expone una teoría del signo lingüístico considerado exclusivamente como forma y estudia tanto la fonología como la gramática. En el análisis y clasificación de las relaciones se hace especial hincapié en la semejanza que hay entre los tipos de relación que aparecen en todos los compartimentos de la lengua. La idea de que existe un paralelismo evidente en la organización sistemática de fonemas, morfemas, sintagmas, etc. es conocida por el nombre de *isomorfismo*, el cual ha encontrado una amplia aplicación en lingüística matemática (cf. 3.25). Por lo demás, Hjelmslev se propuso hacer de su teoría un “álgebra de la lengua”. Asimismo, el lingüista danés se planteó el problema de analizar los sentidos lexicales en el marco del estructuralismo.

La teoría de las gramáticas generativas nació, después de la segunda guerra mundial, en los Estados Unidos, paralelamente al desarrollo de la teoría de los modelos y de la cibernética. N. Chomsky, en sus *Syntactic Structures* (1957), expuso sus principios fundamentales¹⁵. Una gramática generativa es un conjunto de símbolos de las unidades de la lengua y de reglas de funcionamiento que se aplican a estos símbolos con el propósito de producir las proposiciones correctas de la lengua en cuestión. Tal mecanismo gramatical debe estar compuesto de modo que la aplicación de las reglas de funcionamiento sobre los símbolos tenga el mismo efecto que la construcción de enunciados por parte de un hablante (“native speaker”)¹⁶. La lingüística generativa es un modelo que se propone explicar el hecho por el cual cualquier hablante es capaz de formular o de comprender un número infinito de frases inéditas.

Se conocen hasta ahora tres tipos de gramáticas generativas: las de un número de estados finito, las de constituyentes inmediatos y las gramáticas transformacionales (cf. 3.7 y ss.). La teoría de las gramáticas generativas participa tanto de la lingüística estructural como de la lingüística matemática. Sus principales méritos residen en:

- a) haber dado una formalización de la sintaxis más rigurosa que la allegada por los descriptivistas o por otros estructuralistas¹⁷;
- b) haber reconocido los derechos de la semántica distinguiendo en la gramática un componente fonológico y otro semántico¹⁸;
- c) haber superado las preocupaciones del estructuralismo propiamente dicho (pregenerativo), que se limitaba a la segmentación y a la clasificación de las unidades de la lengua y resultaba todavía tributario de la visión tradicional de la

15. [Cf. Bibliografía, para la versión española].

16. Cf. CHOMSKY, *Estructuras y Aspectos*, y RUWET, *Introducción*.

17. Cf., entre los “descriptivistas”, BLOCH-TRAGER, *Outline*; BLOOMFIELD, *Lenguaje*; HARRIS, *String Analysis* y NIDA, *Synopsis*, además de otros, como MUJIN, *Funkcionalnyi analiz*; MARTINET, *El lenguaje* y TESNIÈRE, *Éléments*.

18. Cf. CHOMSKY, *Aspects*; KATZ, *Semantic Theory* y KATZ-POSTAL, *Integrated Theory*.

lengua como jerarquía de los niveles (estratos) fonético, morfológico, léxico y sintáctico.

La gramática, en términos generativos, es una descripción completa de las reglas fonéticas, sintácticas y lexicales de una lengua. Hasta el momento, se han elaborado numerosos estudios de gramática transformacional consagrados al inglés, y, menos, a las lenguas románicas¹⁹. Han aparecido, incluso, los primeros ensayos sobre la historia de la teoría generativa²⁰.

Después de esta historia breve de la lingüística estructural, pasaremos en lo que sigue a exponer los principios estructuralistas que nos parecen más significativos, aun con el riesgo de dejar al lector con la impresión de que esta corriente lingüística constituye un cuerpo doctrinal bastante unitario²¹. Las referencias a una escuela se reducirán a aquellos puntos en que las características son demasiado destacadas, respecto a otras escuelas, para que puedan pasar inadvertidas.

1.4. La crítica a la lingüística tradicional

Los estructuralistas formularon sus tesis partiendo de una serie de críticas contra la lingüística precedente, en especial la de los neogramáticos y la gramática tal como se enseña en las escuelas, tributaria de una tradición que se inicia en la antigüedad greco-latina y continúa en toda la Edad Media. Reuniendo estas dos tendencias bajo una sola denominación, los estructuralistas aluden a la lingüística "tradicional".

A esta lingüística tradicional se le reprochan las siguientes deficiencias:

1) No constituye una disciplina autónoma, en el sentido de que no ha conseguido delimitar el objeto y los métodos de estudio propios del objeto y los métodos de otras ciencias (lógica, psicología, fisiología, física, etc.). La lingüística debe, pues, estudiar la lengua "en sí misma y para sí misma".

2) Tiene un carácter sustancialista, al colocar en primer plano el análisis físico y fisiológico (la sustancia de la envoltura sonora) y el análisis lógico-semántico (el contenido de ideas y sentimientos expresados a través del discurso). Las clasificaciones fonéticas y gramaticales se basan, respectivamente, en las cualidades físicas y fisiológicas de los sonidos y en los significados que expre-

19. Cf., para estas últimas, HADLICH, *Gramática* y VASILIU-GOLOPENȚIA, *Transformational Syntax*.

20. Cf. KING, *Historical Linguistics*. La bibliografía de la escuela transformativista es inmensa (cf. una parte representativa en KRENN-MÜLLNER, *Bibliographie*). Para la crítica de la teoría, cf. COSERIU, *Einführung* y UHLENBECK, *Critical Comments*, entre otros.

21. Para un conocimiento más detallado de las principales escuelas estructuralistas, remitimos a GRAUR-WALD, *Istorie*; HAUGEN, *Directions*; LEPSCHY, *Lingüística*; APRESIAN, *Principles y Éléments*; MALMBERG, *Nuevos caminos* y VACHEK, *School of Prague*.

san las formas y las construcciones. La atención de los lingüistas debe pasar de la sustancia a la forma, esto es al estudio del “revestimiento” de las relaciones. Como ciencia de las formas, la lingüística debe formalizarse, adquirir un carácter más abstracto y, a este propósito, acercarse a la matemática (se ha hablado de la lingüística como de un “álgebra de la lengua”)²².

3) Está dominada por el “atomismo” en el sentido de que describe cada elemento e investiga su evolución sin tener en cuenta los vínculos de interdependencia con los demás elementos. Así, el objeto de estudio queda “pulverizado” y el conocimiento de su esencia desaparece. Cuando, por el contrario, las relaciones y la interdependencia entre los elementos se convierten en centro de atención, la lingüística “atomística” cede ante la lingüística “estructural”.

4) Como una derivación del mismo “atomismo”, se reprocha a la lingüística tradicional la excesiva separación entre los distintos compartimentos de la lengua (fonética, morfología, sintaxis, léxico y semántica). La metodología estructural ha de llegar a una comprensión y a una presentación unitaria de los hechos propios de todos los compartimentos, prestando una atención preferente al isomorfismo y no a la especificidad. Numerosos autores tratan sobre todo de cancelar los límites, considerados artificiosos, entre la morfología y la sintaxis²³.

5) Está supeditada al historicismo. La mezcla del punto de vista descriptivo con el histórico, en la que prevalece éste, vicia la comprensión del fenómeno lingüístico. No aplica el principio de que ambos puntos de vista requieren una delimitación precisa en tanto que la lingüística descriptiva (“sincrónica”) se opone netamente a la lingüística histórica (“diacrónica”). Y como la primera es la que nos ayuda a comprender la esencia de la lengua (la estructura, el carácter sistemático), ella debe ocupar un lugar prioritario.

6) Depende de las clasificaciones propuestas ya en la antigüedad y en el medioevo (p. ej., en la distribución de las palabras en partes del discurso), sobre las cuales cabe admitir, en el mejor de los casos, que tienen validez para las lenguas que estudiaron los mismos autores primitivos: el griego y el latín. De esta crítica resulta la necesidad de proceder a otras clasificaciones (formales, relacionales, puramente lingüísticas), lo cual ha promovido, generalmente, la creación de una terminología nueva en lugar de la escolástica. La escuela danesa constituye la expresión más notable de esta tendencia.

22. Cf. HJELMSLEV, *Prolegómenos*.

23. Para la historia del problema sobre estos límites, cf. LLORENTE, “Morfología”. [Para una problemática teórica más actual, cf. SÁNCHEZ DE ZAVALA (ed.), *Semántica y sintaxis*].

1.5. Algunas dicotomías

La lingüística estructural establece una serie de distinciones fundamentales que adoptan, por lo general, la forma de dicotomías²⁴.

1) La primera se presenta entre *sincronía* y *diacronía*, sobre la que hemos ya hablado. A cuanto hemos dicho más arriba conviene añadir tan sólo la observación de que, en los últimos años, aparece con una agudeza creciente el problema del estructuralismo diacrónico, cuyas bases han sido establecidas únicamente para la fonología²⁵. Con ello se renuncia a la idea de Saussure, según la cual el estudio estructural es incompatible con el de la evolución²⁶.

2) La lengua consiste en un sistema de signos. El signo lingüístico ofrece dos aspectos, la envoltura sonora (el “significante” o la “expresión”) y lo que se comunica a través de ella (el “significado” o el “contenido”). A la distinción *significante/significado* (o *expresión/contenido*), hay que añadir otra:

3) Ambos aspectos del signo lingüístico se analizan en “forma” y “sustancia”. La sustancia de la expresión está constituida por los sonidos; a su vez, la forma de la expresión está constituida por fonemas u organización fonológica de la sustancia fónica. Lo que se comunica constituye la sustancia del contenido, mientras que la organización lingüística de la sustancia se denomina forma del contenido. Creemos, en este punto, que un ejemplo aclarará mejor la dicotomía entre forma del contenido y sustancia del contenido, la cual, por lo demás, ha recibido mucha menos atención entre los lingüistas que la dicotomía correspondiente al plano de la expresión²⁷.

Cuando traducimos una frase de una lengua a otra observamos que no sólo se produce la sustitución de una secuencia de sonidos por otro flujo sonoro: entre dos lenguas hay algo más que diferencias de expresión. Así ocurre si comparamos una frase rumana con su equivalente en alemán:

{ *Gheorghe crede că fiica lui a mîncat*
 { *Georg glaubt, dass seine Tochter gegessen hat*²⁸

24. Cf. HJELMSLEV, *Prolegómenos*; SAUSSURE, *Curso* [y LEPSCHY, *Lingüística*].

25. Cf. MARTINET, *Economía*. Aunque en menor medida, para la sintaxis, cf. KING, *Historical Linguistics*.

26. Un trabajo fundamental relativo a esta concepción es el de COSERIU, *Sincronía*.

27. [A estas dicotomías habría que añadir, al menos, la de lengua y habla, pese a que queda implícita en estos apartados sobre el estructuralismo. Cf. LEPSCHY, *Lingüística* y COSERIU, “Sistema”. Y en cuanto al término “forma”, verdaderamente engañoso en la bibliografía usual, cf. LYONS, *Introducción*].

28. [‘Jorge cree que su hija ha comido’. En ocasiones, nos abstenemos de verter los ejemplos al castellano si no hay equivalentes válidos o si, como ahora, se evitan prolijas modificaciones en el texto explicativo que, a cambio, tampoco iban a mejorar la comprensión del original].

En la traducción hemos otorgado a cada palabra rumana una palabra alemana; existe, pues, una correspondencia biunívoca (lo que, en la práctica, resulta bastante raro). No obstante, se ve inmediatamente que la diferencia entre ambos textos no es puramente fonética. En efecto:

a) El artículo enclítico de *fiica* viene exigido por *lui* (el sintagma **fiică lui* sería incorrecto)²⁹, mientras que *seine* como determinante de *Tochter* prohíbe la articulación del sustantivo (siendo, a su vez, incorrecta la sucesión **seine die Tochter*). Las palabras con el mismo significado (en el contexto aducido) *lui* y *seine* se comportan de un modo distinto respecto a las palabras con el mismo significado *fiica* y *Tochter*, frente a las cuales se hallan en la misma relación (de atributo): *lui* requiere artículo, *seine* lo excluye. Pero tampoco el artículo es sólo una realidad fonética (el fonema /a/ rumano frente a la secuencia /dī/ del alemán), sino un instrumento gramatical con significado propio. Si, entonces, el texto rumano contiene un significado gramatical que no se encuentra en el texto alemán, ¿acaso resulta de ello que la traducción no es buena y que el contenido de la comunicación no es idéntico en las dos lenguas? De ninguna manera. La comunicación (la sustancia) es la misma, pero la organización semántica respectiva (la forma) difiere numéricamente (hay un significado más o menos) y estructuralmente (*seine* excluye el artículo, *lui* lo reclama). En el contenido de aquellas dos frases existe identidad de sustancia y diferencia de forma.

b) La palabra *seine* contiene el morfema de femenino -e, mientras que su equivalente rumano *lui* no tiene ni puede recibir morfema de género. A este respecto, se observa que el texto alemán ofrece un elemento más, sin que el mensaje pierda la equivalencia frente a la versión rumana. Insistimos en el hecho de que entre el contenido de ambas frases existe unidad de sustancia y diferencia de forma.

c) Entre los significados de “fiică” y “lui” hay un régimen de determinación. Tanto si se piensa en el significado de “lui” antes o después del significado de “fiică”, en alemán hay obligación de decir *seine Tochter* y, en rumano, *fiica lui* (la expresión *a lui fiică* es posible, pero no normal en el habla familiar)³⁰. La lengua, o sólo un determinado estilo de ella, constriñe al hablante a disponer los significados en una cierta sucesión, cualquiera que sea el orden con que eventualmente quisiera expresarlos. Entre *fiica lui* y *seine Tochter* hay identidad de sustancia y diferencia de forma (en este caso, en el orden de los significados). La misma diferencia se observa también en el predicado de la oración subordinada:

29. [*Fiica lui* sería, literalmente, ‘la hija de él’, siendo ‘él’ = Jorge, aunque no obligatoriamente (cf. en castellano la misma ambigüedad en la versión de la nota anterior). **Fiică lui* equivale a ‘hija de él’, sin artículo en el nominativo regente e igualmente agramatical en español: *‘Hija de él ha comido’].

30. Tiene el mismo significado que *fiica lui*, ‘su hija’, pero con el posesivo situado antes del nombre regente (*a lui* es una variante posicional de *lui*).

la sucesión **hat gegessen* queda excluida de un tal contexto en alemán literario y, en cambio, en rumano el orden natural es el inverso, *a mîncat*.

En resumen, el mismo contenido de unos sintagmas u oraciones se organiza de un modo específico en dos lenguas cualesquiera. Se trata bien del número de elementos semánticos (gramaticales), bien de asociaciones de elementos semánticos obligatorios o exclusos o bien de la disposición obligatoria de los elementos. Todas estas diferencias no pertenecen a la sustancia —ya que la comunicación no se altera—, sino a la forma, entendida como el modo de organizar el contenido.

4) Para terminar, nos referiremos a la dicotomía “eje sintagmático”/“eje paradigmático”.

En cualquier texto o fragmento de habla el lingüista descubre una sucesión de unidades mutuamente vinculadas que se condicionan entre sí como forma sonora y significado. (Se dice, a este respecto, que el lenguaje tiene un carácter lineal.) En diversos niveles de análisis, el texto se descompone en una cadena de fonemas, morfemas, palabras, partes de oración, oraciones y periodos. El eje imaginario sobre el cual se colocan estas unidades en secuencia para realizar una comunicación lingüística se denomina eje *s i n t a g m á t i c o*. El estudio relacional de la lengua tendrá en cuenta, en primer lugar, la interdependencia de los elementos que figuran en el interior de una cadena de fonemas, palabras, etc.

Sin embargo, la comprensión de una unidad en un texto dado se realiza no sólo como consecuencia de su percepción en él, sino por el hecho de relacionarla con una serie entera de otras unidades que no aparecen en el texto pero que existen en la mente del que percibe el texto. El significado de “segunda persona” y “singular” de la forma verbal *entras* (en un contexto como *¿Por qué entras en clase?*) es comprendido por el lector porque *entras* se opone a *entro*, *entra*, *entramos*, *entráis*, *entran*. El significado de “presente” queda igualmente comprendido porque *entras* se opone a las formas *entrabas*, *entraste*, *has entrado*, *habías entrado*, *entrarás*, etc., mientras que el significado de “entrar” es percibido al relacionar *entras* con formas como *hablas*, *comes*, *duermes*, etc. Por consiguiente, la palabra *entras* forma parte de muchas series:

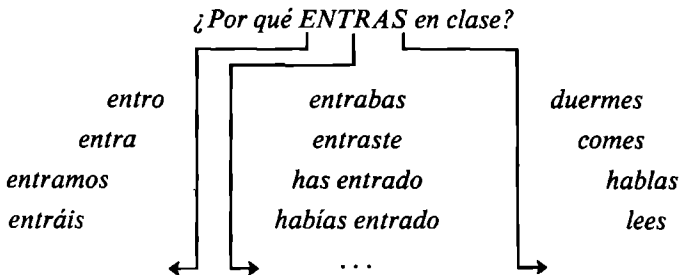


Figura 1

El eje imaginario sobre el cual se colocan las unidades de la lengua en una sucesión del tipo indicado en el esquema se denomina eje *paradigmático*.

Para comunicar una idea, el hablante alinea un cierto número de unidades lingüísticas en el eje sintagmático, pero, por su parte, la colocación de cada elemento sobre este eje está precedida por la elección del mismo y su extracción del eje paradigmático.

1.6. Tipos de relaciones

Entre los miembros de una serie paradigmática surgen diversos tipos de relaciones que constituyen, junto a las relaciones sintagmáticas, el objeto de la lingüística estructural. Pese a que, a primera vista, las relaciones entre unidades a nivel sintagmático difieren considerablemente de las que existen entre los miembros de una serie paradigmática, los dos tipos de relación son *isomorfos*, por cuanto admiten una misma clasificación a base de tres tipos fundamentales, llamados:

$\left\{ \begin{array}{l} a) \text{ interdependencia} \\ b) \text{ determinación} \\ c) \text{ constelación}^{31} \end{array} \right.$	o, en otra terminología,	$\left\{ \begin{array}{l} a) \text{ condicionamiento biunívoco} \\ b) \text{ condicionamiento unívoco} \\ c) \text{ no condicionamiento} \end{array} \right.$
--	-----------------------------	---

a) Si la unidad *A* no puede aparecer sin la unidad *B* y tampoco *B* sin *A*, entonces entre *A* y *B* existe una relación de interdependencia (condicionamiento biunívoco). Por ejemplo, en el eje sintagmático hay interdependencia, en español, entre persona y número verbales (ya que se exigen recíprocamente). En las lenguas con declinación, como latín, rumano, alemán, ruso, etc., también hay interdependencia sintagmática entre número y caso nominales. A su vez, en el eje paradigmático, el número singular presupone el plural y viceversa.

b) Si la unidad *A* no puede aparecer sin la unidad *B*, pero, en cambio, la unidad *B* no implica la presencia de *A*, entonces entre *A* y *B* existe una relación de determinación (condicionamiento unívoco). Así, en el eje sintagmático, una preposición como *contra* requiere un término nominal pero no al contrario. En el eje paradigmático se hallan en relación unívoca los grados de comparación: la existencia del comparativo implica la existencia del positivo, pero no viceversa (existen adjetivos y adverbios que sólo tienen grado positivo).

31. Esta terminología pertenece a la escuela glosemática (cf. HJELMSLEV, *Prolegómenos*). [Para la noción de isomorfismo, cf. 3.25].

c) Si *A* aparece con o sin *B*, y *B* aparece con o sin *A*, entonces entre *A* y *B* existe una relación de constelación (no condicionamiento). Por ejemplo, la conjunción condicional *si* se construye con subjuntivo, pero también con indicativo, y, por su parte, el subjuntivo puede unirse a otras conjunciones distintas de *si*. La autonomía de los paradigmas nominales y verbales constituye asimismo un ejemplo de no condicionamiento en el eje paradigmático.

1.7. La fonética estructural (fonología)

Uno de los problemas fundamentales de la lingüística estructural consiste en la *segmentación*, esto es en el modo de subdividir un enunciado en unidades de distintos niveles (fonemas, morfemas, etc.). Estrechamente ligado con ello se halla el problema de establecer los tipos de unidades lingüísticas.

El primero de estos tipos descubierto y estudiado por los estructuralistas es el *fonema*. La investigación de los sonidos del habla con ayuda de aparatos ha demostrado que lo que parece ser “el mismo sonido” resulta ser en rigor sonidos diferentes, con particularidades acústicas más o menos específicas. Pronunciando diez veces seguidas la palabra *mar*, tanto nosotros como nuestros interlocutores tenemos la impresión de que al final se ha pronunciado siempre [r]. El análisis instrumental con aparatos, sin embargo, prueba inmediatamente que este sonido ha sido pronunciado de diez maneras (r_1, r_2, r_3, \dots , etc.). Pronunciando luego varias veces la palabra *mal*, emitiremos también una serie de sonidos finales distintos (l_1, l_2, l_3, \dots , etc.) y, no obstante, continuamos con la impresión de que se ha oído constantemente [l], el mismo sonido, diferente de [r]. Así, pues, los sonidos distintos r_1, r_2, r_3 , etc. se han percibido como si fueran idénticos, mientras que los sonidos igualmente distintos r_1 y l_1 han sido percibidos como distintos. Con ello se plantea el problema de averiguar qué es lo que permite la *identificación* de los sonidos r_1, r_2 , y lo que determina la distinción de los sonidos r_1 y l_1 . La fonología ofrece a esta cuestión la siguiente respuesta: se perciben como idénticos los sonidos diferentes que, en caso de permutarse uno por otro en la misma palabra, no producen en ella cambio de significado; se perciben, por el contrario, como distintos los sonidos que, al permutarse dentro de una palabra, cambian su significado. De ahí que al cambiar, en *mal*, l_2 por l_3 , pongamos por caso, el significado se conserva: l_2 y l_3 constituyen, pues, *variantes* del mismo fonema. Por su parte, el cambio de l_2 por r_1 en *mal* se traduce en un cambio de significado: *mal* significa algo distinto de *mar*, y, por tanto, l_2 y r_1 pertenecen a fonemas distintos.

La relación entre dos sonidos que no determinan cambio de significado se denomina relación de *variación*; la que se entabla entre dos sonidos que sí determinan cambio se llama relación de *comutación*. Este criterio para el

análisis de los sonidos constituye en realidad su función en el proceso comunicativo. Dos sonidos distintos se hallan en relación de conmutación cuando sirven para comunicar mensajes distintos, y en relación de variación si no tienen esta prerrogativa. En este último caso, presentan, desde el punto de vista comunicativo, la misma función, de donde resulta que sean percibidos como sonidos idénticos.

En el estudio relacional de los fonemas, la atención de los estructuralistas se concentra sobre las relaciones que se establecen en el eje paradigmático, mostrando, en especial, cómo se integran los fonemas de una lengua en forma de sistema y se vinculan entre sí por medio de tipos variados de oposición³².

1.8. La morfología estructural

Como hemos visto, los fonemas son unidades susceptibles de producir diferencias de significado si se permutan mutuamente. Pero los fonemas, aunque promueven la diferenciación de significados, no poseen por sí mismos significado propio: *l* es un fonema distinto de *r* por cuanto *mal* significa algo diferente de *mar*, pero sobre *l* y *r* no cabe decir que signifiquen algo. La más pequeña unidad portadora de significado es el m o r f e m a. El registro de morfemas en la lengua se efectúa, en lingüística estructural, por medio de procedimientos bastante diversos de una escuela a otra. Sin embargo, éstos pueden reducirse a dos tipos principales en tanto que se valgan o no del punto de vista semántico:

Sobre la base de un criterio semántico, una forma como el rumano *paturile*, ‘las camas’, admite un análisis en tres morfemas: *pat*, ‘cama’ (que significa una cierta clase de mueble) + *uri* (con significado de “plural”) + *le* (con el significado de artículo determinado). Por otra parte, en el rumano *lucrătorului*, ‘del/al trabajador’, distinguimos el morfema *lucr* [equivalente al esp. ‘trabaj’], seguido de *ător* [esp. ‘ador’] (“persona que realiza una acción”) + *ul* (con el significado de artículo determinado) + *ui* (con el sentido de genitivo o dativo). Evidentemente, ni *pat* ni *uri*, *le*, *lucr*, *ător*, *ul* o *ui* admiten aquí una descomposición ulterior en unidades más pequeñas que posean un significado determinable³³.

En la escuela descriptivista, un cierto número de estudiosos preconiza la segmentación de los enunciados sin tener en cuenta sus particularidades semánticas³⁴. Para ello, se recurre al concepto de distribución, definido como la totalidad de los contextos en que puede aparecer una unidad dada de la lengua.

32. Cf. TESNIÈRE, *Éléments* y MARTINET, *Economía*. Para los métodos fonológicos aplicados al español, cf. HARRIS, *Spanish Phonology*.

33. [Para este ejemplo y los que siguen inmediatamente, cf. nota 28].

34. Cf. HARRIS, *Structural Linguistics*.

Por ejemplo, el rumano *atom*, ‘átomo’, puede figurar [como masculino] en los contextos:

<i>atom</i>	{	- + <i>ul</i> = <i>atomul</i> , ‘el átomo’
		- + <i>ului</i> = <i>atomului</i> , ‘del/al átomo’
		- + <i>i</i> = <i>atomi</i> , ‘átomos’
		- + <i>ii</i> = <i>atomii</i> , ‘los átomos’
		- + <i>ilor</i> = <i>atomilor</i> , ‘de/a los átomos’

mientras que, a su vez, *cas-* [esp. ‘cas-’ de ‘casa’] puede aparecer [como femenino] en los contextos:

<i>cas-</i>	{	- + <i>a</i> = <i>casa</i> , ‘la casa’
		- + <i>ei</i> = <i>casei</i> , ‘de/a la casa’
		- + <i>e</i> = <i>case</i> , ‘casas’
		- + <i>ele</i> = <i>casele</i> , ‘las casas’
		- + <i>elor</i> = <i>caselor</i> , ‘de/a las casas’

[siendo *casă*, ‘casa’, la palabra base]³⁵.

Pronunciando varias veces la palabra *atom*, obtenemos realizaciones fonéticas distintas. El empleo del criterio semántico reconocerá aquí una serie de variantes respecto al mismo morfema, ya que el significado ha permanecido inalterado. Si se renuncia a este criterio y se le sustituye por el distribucional, habrá que decir que las diferentes realizaciones fonéticas de *atom* pertenecen al mismo morfema porque tienen una distribución idéntica: todas pueden figurar, en efecto, en los contextos *ul*, *ului*, *i*, *ii*, *ilor*. Sin embargo, a veces surgen situaciones más complejas.

Comparando los segmentos sonoros del rumano *învăţ*, *înceţ* y *învaţ* [distintos elementos radicales del verbo *a învăţa*, ‘aprender, enseñar’], se comprueba que tienen una distribución específica: *învăţ* puede aparecer en los contextos *ăm*, *aţi*, *am*, *asem*, etc. [*învăţăm*, ‘aprendemos’; *învăţaţi*, ‘aprendéis’, etc.]; se trata, pues, de formas conjugacionales, donde no pueden aparecer las raíces *înceţ* y *învaţ*. Por su parte, *înceţ* figura en el contexto *i* [*înceţi*, ‘aprendes’], en el que se excluyen las otras alternativas aducidas, mientras que *învaţ* tiene, como contexto privativo, *ă* [*învaţă*, ‘aprende’]. Dos o más segmentos sonoros que aparezcan cada uno en contextos donde el otro (o los otros) están excluidos se

35. [En rumano, los nombres femeninos o neutros terminados en *-ă* (*casă*, *iarbă*, ‘hierba’, *cerneală*, ‘tinta’, etc.) toman el artículo determinado modificando *-ă* por *-a*, con lo que se forma una oposición *casă* / *casa* [kásə / kása], etcétera. Los masculinos y neutros terminados en consonante (*atom*, *lup*, ‘lobo’, *tren*, ‘tren’, etc.) añaden el segmento *-ul* para el mismo cometido. De ahí que *atom* exista aisladamente, pero no *cas-*, según el texto].

hallan en distribución complementaria. Las tres formas indicadas, extraídas de la conjugación del verbo *a învăța* [pron. a învətsá], son, pues, distribucionalmente complementarias.

¿Cómo llegar, entonces, a la conclusión de que estas tres formas pertenecen al mismo morfema (son variantes de él) sin referirnos a su significado, que es indudablemente idéntico? Para ello, hay que verificar si las formas *învăți*, *înveți* y *învat* cumplen con la siguiente condición: la de que exista en la lengua rumana alguna otra forma cuya distribución sea equivalente a la suma de las distribuciones de los casos examinados. En efecto, la condición se satisface: el segmento fónico *plimb* [raíz única del verbo *a (se) plimba*, ‘pasar(se)’] aparece en los contextos de *învăți* (*ăm, ați, asem*, etc.), en el de *înveți* (*i*) y en el de *învat* (*ă*). Tal condición excluye, como se ve, la intervención del criterio semántico. [Nótese que, en español, hay ejemplos análogos en abundancia. Así, la distribución de la forma *tem-* (de *temer*) cubre todas las distribuciones particulares de las variantes *ten-*, *tien-*, *teng-* y *tuv-* (de *tener*): *tenía - temía, tiene - teme, tengo - temo, tuviera - temiera*, etc.].

La renuncia al criterio semántico se debe sobre todo al hecho de que éste da lugar a grandes dificultades. Por ejemplo, en las combinaciones:

*despacho*₁ *de roble*
*colegas de despacho*₂
*se ha ido al despacho*₃
*despacho*₄ *oficial*

aparece constantemente la secuencia *despacho*, considerada como morfema desde el momento en que tiene significado y no admite la subdivisión en unidades más pequeñas dotadas de sentido en los contextos indicados. Así, las unidades *despacho*₁, *despacho*₂, etc. plantean la cuestión de si son variantes del mismo morfema o bien morfemas diferentes. Como el significado no es idéntico, deberíamos reconocer aquí cuatro morfemas. Ahora bien, como también la mayoría de las palabras tienen más de un significado, más o menos distinto (en función de las proposiciones en que se emplean), habría que concluir que el número de morfemas, como el de sonidos, tiende al infinito.

La lingüística tradicional considera que *despacho*₁, *despacho*₂, etc. representan una sola unidad, y en cambio que, p. ej., *estepa*₁ del contexto *la inmensa estepa* constituye unidad diferente respecto de *estepa*₂ del contexto *un haz de estepa*. ¿A qué se debe, entonces, que la diferencia entre los significados “erial extenso” y “mata de cistinea” sea relevante, mientras que la que opone “un determinado mueble” a “una comunicación estatal” no lo sea?³⁶

36. [El *Diccionario* de la R. A. E. (ed. 1970), basado en razones etimológicas, incluye bajo una misma rúbrica, y junto con otros, los cuatro sentidos de *despacho* y separa, como entradas diferentes, los dos de *estepa*].

La crítica al análisis distribucional ha demostrado que el criterio contextual puede sustituir al semántico con una cierta utilidad cuando se trata de la agrupación de variantes morfélicas (denominadas, a veces, a l o m o r f o s) en morfemas (cf. el ejemplo anterior). Pero entonces se plantea la cuestión de cómo se han de descubrir las variantes. De hecho, la aplicación del método distribucional exige una labor gigantesca, prácticamente irrealizable. En efecto, hay que obtener las variantes de los morfemas mediante la segmentación de los enunciados lingüísticos y, si no se atiende al significado, es necesario buscar todas las divisiones posibles e investigar la distribución de todos los segmentos que resultan de la inmensa multitud de divisiones “de prueba”³⁷.

La aplicación del estructuralismo en morfología aún tiene por delante graves dificultades. Hasta el momento, sus resultados, desde luego apreciables, no han alcanzado todavía la precisión que se ha conseguido en fonología.

1.9. La palabra y las unidades superiores a la palabra

La unidad inmediatamente superior consiste en una combinación de morfemas denominada por algunos estructuralistas, de acuerdo con la tradición, p a l a b r a o, con terminología más reciente, s i n t a g m a m í n i m o. El hecho de que el morfema carezca aún de una precisa definición, lo mismo que sus características combinatorias en cada una de las lenguas, explica por qué no se ha formulado todavía una teoría estructuralista unitaria para este tipo de elemento³⁸. La palabra (o sintagma mínimo) aparece como una entidad distinta del morfema debido a su relativa autonomía en el proceso de comunicación. Esta cualidad se refleja en:

a) la posibilidad de modificar el orden de las palabras: *Jorge admira a las chicas, A las chicas admira Jorge*, etc. (mientras que el orden morfemático de *chic-a-s*, p. ej., es único). Este fenómeno resulta aún más evidente en latín, donde oraciones como *Magister discipulis dixit*, ‘El profesor dijo a los alumnos’, admiten cualquier orden entre sus palabras constitutivas³⁹;

b) el hecho de que una palabra pueda formar por sí sola una comunicación (*Duerme. ¡Nieve! ¡Rápido! Carlos*, esta última como respuesta a preguntas del tipo “¿Quién te lo dijo?”). En cambio, ninguno de los morfemas que componen

37. Para la crítica del análisis distribucional, cf. también MALMBERG, *Nuevos caminos*.

38. Cf. KRÁMSKÝ, *The Word*.

39. [La capacidad transpositiva de las palabras dentro de la oración fluctúa de una lengua a otra y, en general, se halla ligada al factor concordancia (cf. 8.14 y ss.). Las lenguas llamadas polisintéticas tienen reducida al mínimo esta característica; las de tipo flexivo, en cambio, suelen tenerla potenciada (cf. HJELMSLEV, *Lenguaje* y SAPIR, *Lenguaje*). Por otra parte, aunque el orden oracional de las palabras sea relativamente libre, hay siempre tendencias normativas en favor de una u otra disposición determinada. Cf. COSERIU, “Sistema”].

palabras como *desarme, bonito, cantando*, puede aparecer contextualmente aislado (*des, arm, e, bon, it, o, cant, ando*).

Tales apreciaciones han inducido a ciertos estructuralistas a definir la palabra bien como la más pequeña unidad con significado y desplazable en el interior de un enunciado, bien como “forma mínima libre”⁴⁰.

La lingüística tradicional reconoce como unidades superiores a la palabra la parte de oración (*Esta novela de aventuras es muy fascinante* contiene siete palabras, pero sólo cuatro partes de oración: sujeto, dos atributos y predicado nominal), la oración y el período. Por su parte, los estructuralistas admiten, en general, una sola unidad superior a la palabra, el *e n u n c i a d o*, definido como “segmento fónico capaz de aparecer entre dos pausas”, lo que, en terminología corriente, corresponde a una parte de oración, a una oración entera e incluso a un período muy largo y complejo. Las partes de un enunciado se combinan entre sí en grupos diferentes por su extensión y estructura. Para tales grupos, los estructuralistas emplean la denominación de *sintagma o construcción*⁴¹. Se concibe el sintagma, en general, como unidad binaria, en el sentido de que se descompone en dos miembros, entre los cuales existe una relación de subordinación. Por ejemplo, en la oración *He leído una nueva novela de aventuras*, resultan ser sintagmas (construcciones):

una nueva novela de aventuras
una nueva novela
de aventuras
una novela.

Por lo demás, toda la oración puede ser considerada sintagma (o construcción) en la frase *He leído una nueva novela de aventuras que compré ayer*. Sobre este problema, que concierne al tratamiento estructural de aquél compartimento lingüístico que tradicionalmente se denomina sintaxis, se tratará ampliamente en los capítulos siguientes.

El paso de un nivel de segmentación a otro, de un tipo a otro de unidad, se realiza con arreglo a un principio conocido por el nombre de estratificación: una unidad lingüística de un nivel determinado está compuesta por una serie de unidades del rango inmediatamente inferior, y eventualmente por una sola unidad inferior. De este principio de jerarquización deriva una consecuencia natural:

40. La primera corresponde a la definición preferida por la Escuela de Praga. La segunda se encuentra en la obra de descriptivistas como BLOCH-TRAGER, *Outline*; BLOOMFIELD, *Lenguaje* y GLEASON, *Introducción*, entre otros. La palabra, como unidad sintáctica y sus diversas definiciones gramaticales, se trata también en STATI, *Teoria*.

41. En lingüística estructural, el término “sintagma” no siempre tiene un significado unitario. Para algunos, indica cualquier agrupamiento de dos elementos, uno regente y otro subordinado; para otros, se refiere a una unidad formada por dos partes de oración con relaciones de subordinación en el ámbito de otra oración. [Cf. LAZARO, *Diccionario*].

el número de unidades de rango superior es mayor que el número de unidades de nivel inferior. Por ejemplo, los morfemas se componen de fonemas (según el principio de la estratificación), pero el número de morfemas supera con mucho al de fonemas (consecuencia del principio de la estratificación). Del mismo modo, los enunciados están formados por cadenas de palabras, pero el número de enunciados es mucho mayor que el de palabras.

1.10. La sintaxis tradicional

En la sintaxis tradicional, las palabras de una oración se agrupan, de acuerdo con sus relaciones gramaticales, en diversos niveles. Esta agrupación suele representarse con ayuda de un esquema del tipo:

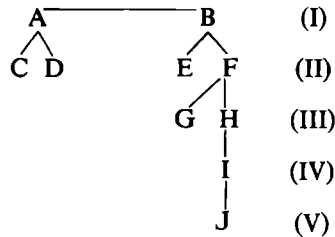


Figura 2

en el cual las letras (A-J) simbolizan las partes de la oración, cada segmento no horizontal une cada miembro subordinado a su regente y las cifras I-V indican los cinco niveles de oración representados en el esquema.

Así se manifiesta el orden estructural de oraciones como: *Mi hermano Juan se ha comprado un aparato de radio de muy buena calidad* o bien *El tercer volumen de versos cierra así un largo período de creación inspirada en el folklore*. El esquema es válido para ambos ejemplos, aunque éstos difieran no sólo con relación a su contenido y a las nociones que evocan, sino también desde el punto de vista gramatical (partes del discurso entre las que se establecen las relaciones sintácticas, tipo de determinantes, tónica). Las diez partes de oración de los textos ejemplificados obedecen a cuatro clases o categorías: sujeto, atributo, predicado y complemento. Así, pues, el análisis sintáctico de una oración consta de:

- a) su descomposición en partes de oración;
- b) la asignación de cada una de estas partes a una de las cuatro categorías reseñadas antes; y

c) la indicación de cómo se hallan unidas entre sí las partes de oración (p. ej., x es atributo de y , el cual es, a su vez, complemento de z ; eventualmente, y está al mismo tiempo coordinado con w , también complemento de z).

Si se le pide a la gramática tradicional que analice el enunciado en unidades más pequeñas, observaremos que algunas partes de oración vienen expresadas por una sola palabra y otras por dos o más (p. ej., *de calidad, en el folklore, un aparato de radio, etc.*).

A los procedimientos analíticos tradicionales se les puede reprochar ciertos defectos:

1) Carecen de unidad. Realmente, el análisis del período mantiene un paralelismo con el de la oración (la unidad se refleja, p. ej., en el principio de correspondencia entre oraciones subordinadas y partes de oración)⁴², pero ambos análisis nada parecen tener en común con el análisis morfológico. ¿No existen, también aquí, paralelismos?

2) No está claro en absoluto el número y el orden jerárquico de las unidades sintácticas de la lengua. ¿Qué relación hay entre “palabra” y “parte de oración”? ¿Qué lugar ocupan las locuciones y las palabras compuestas en la jerarquía de niveles? ¿Acaso no vale la pena reconocer también unidades sintácticas más grandes que la parte de oración y más pequeñas que la oración misma? (algunos emplean, a este propósito, el término de “grupo del sujeto”).

3) El problema de los tipos de relación entre unidades del mismo nivel tampoco está resuelto de un modo satisfactorio. En sintaxis se habla de coordinación y subordinación, pero esta distinción no se corresponde con la estructura de la palabra. Más aún, los dos tipos de relaciones sintácticas mencionadas no lo comprenden, evidentemente, todo. ¿Cómo debe calificarse la relación entre cópula y nombre predicativo o entre una oración y un nombre? La aposición y la oración apositiva, ¿contraen verdaderamente una relación de subordinación?

1.11. La sintaxis estructural

Entre los ensayos de sintaxis estructural que han encontrado mayor número de adeptos y que han resistido mejor el paso del tiempo se halla el método de análisis en constituyentes inmediatos, aparecido en el marco de la escuela descriptivista⁴³. Como creador del método puede considerarse al lingüista americano K. L. Pike, el cual lo delineó en un estudio aparecido en 1943, en la revista

42. Sujeto = oración subjetiva, objeto directo = oración subordinada directa (objetiva), etc.

43. Cf. HARRIS, *String Analysis*; MARTINET, *El lenguaje*; MUJIN, *Funktsionalnyi analiz*; NIDA, *Synopsis* y TESNIÈRE, *Éléments*.

Language, e inspirado en ciertas ideas de L. Bloomfield, fundador de la escuela descriptivista⁴⁴. Este método ha entrado en una fase nueva con la creación de la teoría de las gramáticas generativas. Decenas de lingüistas de distintos países europeos trabajan hoy de un modo habitual con la noción de constituyentes inmediatos⁴⁵.

Se consideran datos todos los enunciados posibles en una lengua cualquiera. Así, pues, el objeto sometido al análisis en constituyentes inmediatos forma un conjunto infinito de enunciados (definidos como “comunicación lingüística que puede ir precedida y seguida de pausa”). Las partes en que puede dividirse el enunciado se denominan constituyentes. Por regla general, el enunciado se descompone al principio en dos segmentos, cada uno de ellos se descompone de nuevo en otros dos y así sucesivamente hasta llegar a los segmentos que no admiten nuevas divisiones porque producirían complejos fónicos sin significado propio. La segmentación se detiene, por tanto, al nivel de los morfemas en tanto que incluyen toda la estructura gramatical. Los elementos que resultan de la primera división de un enunciado o de una parte de enunciado se llaman *constituyentes inmediatos* del enunciado (y, respectivamente, del fragmento del enunciado). Los constituyentes inmediatos de la primera de las oraciones indicadas en el apartado anterior serán, entonces:

*Mi hermano Juan
se ha comprado un aparato de radio de muy buena calidad.*

Mi hermano Juan se divide, a su vez, en dos nuevos constituyentes inmediatos:

*Mi hermano
Juan*

y así sucesivamente.

Todo grupo de dos constituyentes inmediatos determina una *construcción*.

A nivel de palabra, se ejecuta también una división similar a ésta, a fin de descomponerla en morfemas. Por ejemplo, *intratable* se descompone en *in + tratable*, y *tratable* en *trat + able*⁴⁶.

Después de efectuar todas las segmentaciones posibles, el enunciado se ofrece como una secuencia de morfemas, denominados *constituyentes últimos*.

44. Cf. PIKE, “Taxemes” y también WELLS, “Immediate constituents”.

45. Para la historia de las teorías y métodos de la sintaxis estructural, cf. AJMANOVA-MIKAEELAN, *Theory, y STATI. Teoria*.

46. [En rigor, en *able* habría que distinguir *a* y *ble*, ya que *a* puede alternar con *i* (*indecible*, etc.). Cf. la nota siguiente].

mos del enunciado. Volviendo a nuestro ejemplo, cabe representar así los constituyentes últimos que lo forman:

*mi herman-o Juan se ha compr-a-do un aparat-o de radio de muy buen-a calidad*⁴⁷.

Hasta ahora, la finalidad de este análisis parece consistir en el descubrimiento de las partes componentes de un enunciado. Deteniéndonos un momento aquí, comprobamos tres diferencias evidentes respecto al análisis tradicional:

1) se opera con divisiones (dicotomías) sucesivas;

2) el resultado del análisis en todas sus fases produce unos segmentos denominados de la misma manera ("constituyentes inmediatos"), de forma que los constituyentes últimos también son, en último término, constituyentes inmediatos. Ocurre, pues, que el problema de determinar el número de niveles en funcionamiento y la definición de cada nivel no interesa por el momento (sólo se trata con enunciados y constituyentes inmediatos);

3) los límites sintácticos entre período, oración y morfología desaparecen.

Yendo más allá, el análisis se propone dar una representación esquemática sobre la estructura del enunciado descompuesto y, con ello, descubrir su modo de lo. Para ilustrar este punto, utilizaremos una oración como *Él compra un sombrero nuevo*

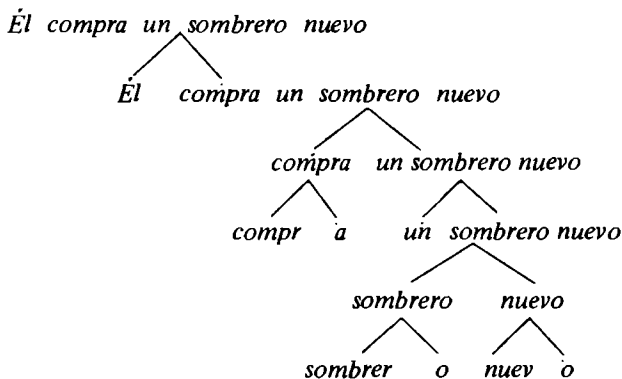


Figura 3

47. Naturalmente, habría un buen número de objeciones contra el análisis propuesto aquí. Confiamos, sin embargo, en que el lector reconocerá en él un simple medio para comprender el método de análisis en componentes inmediatos.

sombrero nuevo, cuyos constituyentes sean, a este propósito, *él, compr, a, un, sombrer, o, nuev y o*. La estructura oracional se representa en la figura 3.⁴⁸

Cientos de proposiciones en español (p. ej., *Pedro trae unos libros prohibidos, Tú lees un poema escabroso*, etc.) se incluyen en el mismo esquema organizativo de constituyentes y, por tanto, en el mismo modelo. Haciendo abstracción del contenido concreto de cada una de estas proposiciones, obtenemos la representación ideal del modelo:

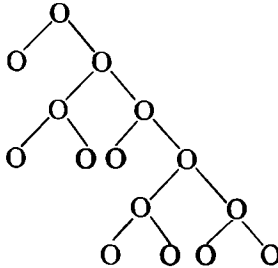
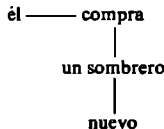


Figura 4.

Sobre la base de un conjunto de decenas o cientos de miles de proposiciones en español podríamos descubrir los distintos modelos que las conforman. El número de estos modelos resulta, desde luego, mucho más pequeño que el de las proposiciones posibles, pero todavía es muy grande. A través del método analítico en constituyentes se persigue una reducción considerable de este número, partiendo de la hipótesis de que los modelos identificados están contruidos, en realidad, sobre un inventario mucho más reducido de modelos más simples. Se buscan, pues, modelos con una capacidad superior de generalización. Más adelante trataremos sobre el procedimiento utilizado con esta finalidad.

Ocurre, en consecuencia, que el segundo objetivo del método en discusión consiste en la reducción del número infinito de enunciados a un número finito (y lo más pequeño posible) de tipos estructurales o, dicho en otras palabras, en la reducción de las variantes a invariantes⁴⁹. El conjunto de enunciados posibles

48. Compárese con la representación esquemática tradicional:



49. Se procede de un modo análogo al fonológico, que se propone reducir el número infinito de los sonidos a un número finito (del orden de las decenas) de invariantes, llamadas fonemas (cf. 1.7).

en una lengua dada no es más que un número de realizaciones variadas de algunos modelos. Más concisamente: esta segunda finalidad del análisis en constituyentes inmediatos reside en la *m o d e l a c i ó n* gramatical de los enunciados.

1.12. Técnica de segmentación en constituyentes inmediatos

En los ejemplos aducidos más arriba, hemos indicado cómo se segmenta el enunciado en constituyentes inmediatos sin *j u s t i f i c a r* las sucesivas divisiones. ¿Cuáles son los criterios de segmentación? ¿Por qué hemos cortado cada una de las porciones del texto en el punto *x* y no en el punto *y* o *z*? ¿Por qué distribuimos *Él compra un sombrero nuevo* a base de *Él + compra un sombrero nuevo* y no, p. ej., a base de *Él compra + un sombrero nuevo*? Este problema, en rigor el más difícil, alude directamente a la llamada *t é c n i c a d e s e g m e n t a c i ó n* en constituyentes inmediatos.

El descriptivista Nida concede una preferencia evidente al criterio semántico: nosotros “sentimos” que en un ejemplo como *Él compra un sombrero nuevo, nuevo* va con *sombrero* y que el grupo entero *un sombrero nuevo* está subordinado al verbo *compra*⁵⁰. La función de los criterios formales, estructurales, consiste en confirmar y a veces en corregir los datos obtenidos por este medio. Tales criterios preponderan, naturalmente, allí donde el significado no nos ayuda a elegir entre dos o más alternativas posibles. Pero, ¿cuáles son los criterios estructurales? Conviene precisar, a esta sazón, que, aparte de Nida, los principales representantes del descriptivismo fundan exclusivamente sobre estos criterios la técnica de segmentación en constituyentes.

El criterio básico es la *s u s t i t u c i ó n*. La división de los enunciados en segmentos se justifica por el hecho de que hay porciones en el enunciado sustituibles (equivalentes, pues) por porciones de otros enunciados. Dado un enunciado de una extensión cualquiera, se intenta sustituir una parte del mismo por otro material lingüístico. Se consideran válidas sólo aquellas sustituciones que dan lugar a enunciados *i d e n t i f i c a b l e s*, es decir admitidos por la lengua respectiva. Si en *Él compra un sombrero* reemplazáramos la porción *Él compra un* por *amarillas*, se obtendría el texto inidentificable **amarillas sombrero*. En cambio, si realizáramos la sustitución por *este* o por *Sofía pone un lacito en el*, el resultado determinaría dos textos identificables: *este sombrero*; *Sofía pone un lacito en el sombrero*. Se verifica luego si la segunda porción del enunciado inicial (*sombrero*) es sustituible, cosa, evidentemente, muy simple (obtenemos, por ejemplo, *Él compra un calcetín*). Por consiguiente, hemos dividido el enunciado *Él compra un sombrero* en dos porciones (estableciendo el límite después

50. Cf. NIDA, *Morphology* y “Taxemes”.

de *un*) y hemos comprobado que ambas son sustituibles por otro material lingüístico.

En este caso decimos que el enunciado aducido *p u e d e tener* como constituyentes inmediatos las porciones *Él compra un* y *sombrero*. Sin embargo, se ve fácilmente que la dicotomía elegida aquí no es la única posible, ya que si la segmentación se efectúa después de *él*, observamos que *él* es sustituible, p. ej., por *Juan*, y *compra un sombrero* por *come*, de modo que *Él + compra un sombrero p u e d e n* ser considerados también constituyentes inmediatos⁵¹.

La técnica de segmentación ha de indicarnos incluso cómo se elige, entre diversas dicotomías posibles, la mejor. Pero, por el momento, es más oportuno introducir la noción de clase de secuencias. Se considera que, cualquiera que sea la forma de segmentar un enunciado, el resultado ha de producir una pareja de secuencias o series de morfemas. Una serie tal tendrá una extensión mínima de un morfema. Si segmentamos la proposición, p. ej., en los constituyentes inmediatos *Él* y *compra un sombrero*, el primero consta de la secuencia monomorfemática *él*, y el segundo, de la secuencia *compr + a + un + sombrer + o*. Todas las secuencias monomorfemáticas que pueden aparecer en lugar de *él* (*éste*, *ella*, *Juan*, etc.) forman, junto con ella, una clase de secuencias⁵². Generalizando, una secuencia monomorfemática *S* pertenece a la clase de secuencias determinadas por el contexto *C* si *SC* compone un texto admisible por la lengua, es decir un texto identificable⁵³.

Si la secuencia contiene más morfemas (p. ej., *compra un sombrero*), entonces pertenece a una clase de secuencias que comprende a todas aquellas que se caracterizan por el hecho de:

a) contener el mismo número de morfemas;

b) que su primer morfema pertenezca a la misma clase de secuencias monomorfemáticas, y que el segundo, el tercero, etc., también. Así, pongamos por caso, *compr* entra en la misma clase con *busc*, *mir*, etc.; *sombrer* entra en la misma clase con *libr*, *zapat*, etc.; *a*, en la misma clase con *aba*, *ó*, *ará*, etc., y así sucesivamente. La secuencia *compra un sombrero* se halla, pues, en la misma clase de secuencias que *buscaba un libro*, *miró un zapato*, etcétera.

Como consecuencia de ello, toda segmentación de un enunciado en dos constituyentes comporta su descomposición en dos miembros pertenecientes a dos clases de secuencias, de modo que cada constituyente representa a una clase.

51. El lector puede continuar por sí mismo la prueba haciendo nuevas segmentaciones tanto en otros puntos del enunciado como en el interior de una palabra; p. ej., *él compr + a un sombrero*, etc.

52. En la misma clase de *él* no entran palabras como *la mujer*, *este cliente*, *el profesor de dibujo*, ..., porque son secuencias con más de un morfema.

53. Así, pues, los morfemas se agrupan en clases según los contextos en que pueden aparecer.

Ahora, para acercarnos mejor a la cuestión sobre cómo elegir la mejor entre diversas dicotomías posibles, convendrá presentar las nociones de:

— expansión;

— foco;

— clase de focos (también llamada clase de formas o clase de distribución)⁵⁴.

Supongamos que una porción de un enunciado está formada por la secuencia S_1 y que S_1 es reemplazable por S_2 . Si S_1 es más largo que S_2 (contiene más morfemas) o es, como máximo, igual a S_2 (S_1 y S_2 tienen el mismo número de morfemas) y S_1 es distinto estructuralmente de S_2 (ambos no pertenecen a la misma clase de secuencias), entonces S_1 se denomina expansión de S_2 , y S_2 es el modelo de S_1 .

En el enunciado *Él compra un sombrero*, la secuencia *él* es sustituible por cualquiera de las secuencias *el cliente*, *mi sobrina*, *el vecino de tu médico de cabecera*, etcétera. (Evidentemente, ninguna de estas secuencias pertenece a la misma clase de *él*). La secuencia *él* constituye su modelo y, a su vez, éstas son algunas de las expansiones posibles de *él*.

Cada una de las secuencias:

él

el cliente

mi sobrina

el vecino de tu médico de cabecera

forman parte de clases diferentes de secuencias (se distinguen, en primer lugar, por el número de morfemas), pero presentan, por lo menos, una propiedad común: la de formar, junto con *compra un sombrero*, un enunciado identificable.

La parte de un enunciado que permanece inalterada a lo largo de las operaciones de sustitución se llama *entorno o contexto*, y la porción sobre la que se realizan las pruebas, *foco*⁵⁵. En nuestro ejemplo, *él* es el foco del enunciado y *compra un sombrero*, el entorno o contexto del foco. Bien entendido que, si hubiéramos probado las sustituciones posibles de la secuencia *compra un sombrero*, ésta se hubiera convertido en foco, y *él* —, en contexto.

Todas las clases de secuencias que pueden aparecer en un determinado contexto llevan el nombre de clase de *focos* (o clase de distribución) relativa a aquel contexto. La clase de distribución o clase distribucional es un concepto supraordenado a la clase de secuencias y es, por tanto, más general que el de ésta.

54. Cf. WELLS, "Immediate constituents".

55. Esta denominación sugiere el hecho de que sobre esta parte del enunciado se concentra la atención o, como si dijéramos, la luz del análisis.

Las secuencias de estructura diversa, citadas más arriba, que pueden aparecer en lugar de *él* en el contexto *compra un sombrero* forman tan sólo una mínima parte de los miembros de esta clase distribucional. Si hubiéramos segmentado la proposición de otro modo, p. ej., a base de *él compra + un sombrero*, y hubiéramos tanteado las posibilidades sustitutivas de *él compra*, hubiesen aparecido igualmente numerosos miembros de esta nueva clase distribucional (v. gr., *Teresa escoge; Dame, por favor; No me regales precisamente*, etc.). Cabe estudiar también la clase distribucional relativa en el contexto *sombrero*, es decir la clase de secuencias que pueden aparecer en lugar de *Él compra un*. ¿Hasta qué punto nos ayuda la noción de clase distribucional a elegir la segmentación más adecuada en dos constituyentes inmediatos?

La respuesta de los partidarios de este método puede resumirse así: toda posible segmentación da lugar a una clase distribucional. Se elige, entonces, aquella que produzca una clase distribucional más rica en:

- a) el número de contextos en que pueda aparecer, y
- b) el número de las clases de secuencias que engloba.

Si se comprueba, p. ej., que la clase distribucional de *él* resulta más rica que la clase de *él compra*, la segmentación de nuestra proposición se realizará después de *él*, y este elemento se convertirá, así, en el primer constituyente inmediato del enunciado.

De los dos criterios según los cuales se valora la riqueza de una clase distribucional, el segundo (basado en la variedad estructural) no requiere ningún comentario; en cambio, el primero implica una operación que todavía permanece inédita: se examina la distribución o número de contextos en que puede aparecer una clase distribucional o de focos. En otras palabras, si al principio del análisis el contexto era mantenido constantemente, ahora se estudian también sus variaciones.

La clase distribucional que admite miembros de las clases de secuencias de tipo:

él
el cliente
 etc.

puede aparecer en contextos como:

paga
compra un sombrero
elige unas corbatas horrendas
 etc.

Estos contextos son, en realidad, miembros de clases diferentes de secuen-

cias (las diferencias radican, en primer lugar, en el número de morfemas), lo mismo que los elementos lingüísticos que pueden sustituir al foco. El conjunto de clases de secuencias que son capaces de desempeñar el papel de contexto respecto a la clase distribucional *él* constituye también una clase con un gran carácter de generalidad y podría ser llamada *clase de contextos*. Las dos clases de formas están constituidas de tal modo que, si un miembro de la primera se puede combinar con un miembro de la segunda, entonces cualquier miembro de la primera clase se puede combinar con cualquier miembro de la segunda. Esquemáticamente:

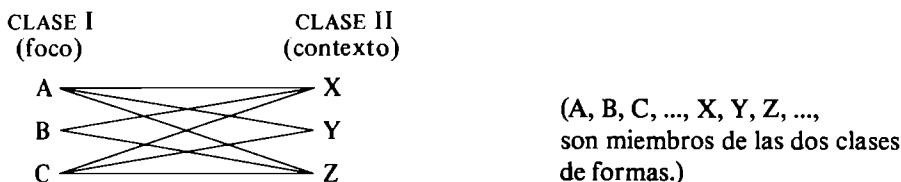


Figura 5.

Recapitulando en el orden de las operaciones:

a) el enunciado *Él compra un sombrero* admite diversas divisiones en constituyentes inmediatos. La validez de cada una viene demostrada por el hecho de que produce dos porciones, ambas sustituibles por otro material lingüístico. Se plantea entonces el problema de qué dicotomía es preferible;

b) se examinan por orden cada una de las dicotomías posibles, eligiendo un foco y un contexto. Se valora la riqueza de cada clase distribucional en función del número de clases de secuencias que comprende;

c) se investiga, en fin, la variación del contexto que admite cada clase distribucional.

Una vez efectuada esta sucesión de operaciones, parece que podríamos hallarnos, por fin, en condiciones de responder a la pregunta del principio (“¿Cómo se elige, entre diversas dicotomías posibles, la mejor?”): la dicotomía que produzca clases más ricas de formas, de modo que las dos partes resultantes de esta segmentación serán los constituyentes inmediatos del enunciado.

El criterio basado en el cálculo de la riqueza en las clases distribucionales no tiene, sin embargo, una fácil aplicación. En rigor, con cualquier tipo de segmentación se obtienen habitualmente secuencias con enormes posibilidades de sustitución. El inventario de todas las clases de secuencias susceptibles de aparecer en el foco o en el contexto exige un esfuerzo extremadamente laborioso y a menudo irrealizable. Los autores del método se han dado cuenta de este inconveniente y han propuesto una serie de simplificaciones.

En la práctica, no se realiza un cálculo completo, ya que se descartan ciertas segmentaciones posibles a causa de la pobreza relativa que manifiestan sus clases de formas resultantes. Por ejemplo, si hubiera que segmentar la proposición *Veo una revista ilustrada*, la dicotomía posible *Veo una revista + ilustrada* sería prontamente abandonada en favor de la dicotomía *Veo + una revista ilustrada*, puesto que fácilmente se intuye que los determinantes susceptibles de aparecer en lugar de *ilustrada* son mucho menos numerosos que los determinantes virtuales de *veo* (¡existen menos tipos de revista que objetos visibles!). De la misma manera, la segmentación *Veo + una revista ilustrada* es preferible a la segmentación *Veo una + revista ilustrada*, porque después de *veo una* debe seguir un nombre femenino, mientras que después de *veo* el género del nombre no está prefijado. En cambio, es muy difícil determinar, en el caso de una proposición como *Ha comprado libros y cuadernos*, cuál es la segmentación preferible entre:

ha comprado + libros y cuadernos o bien
ha comprado libros + y cuadernos.

Comprobamos, pues, que en nuestro análisis interviene una suerte de aproximación y que, de hecho, no es forzoso calcular la riqueza de las sustituciones para todas las segmentaciones posibles.

Con el mismo objeto de simplificar el procedimiento se introduce todavía un nuevo criterio: en un enunciado que comprenda por lo menos dos palabras se examinan tan sólo las segmentaciones que no impliquen un corte en el interior de una palabra. Por ejemplo, en *Veo una revista ilustrada* no se toma en consideración la posibilidad de dividir *Veo una revist + a ilustrada*. Respetando así la unidad de la palabra, no se plantea entonces el problema de la segmentación en constituyentes inmediatos ante todo enunciado compuesto por dos palabras.

Por fin, interviene asimismo el criterio “del sistema”: puesto que un análisis en constituyentes inmediatos implica otros análisis, su coherencia no se considera verificada mientras no se han explorado sus más remotos efectos sobre el sistema. Lo que se acepta o se rechaza no es, en realidad, el análisis de una sola proposición, sino el sistema de constituyentes inmediatos de una lengua, es decir una serie completa o un **s i s t e m a** de análisis para todos los enunciados de la lengua.

1.13. Las relaciones gramaticales

El método de análisis en constituyentes también se propone examinar los tipos de relaciones que contraen los constituyentes inmediatos entre sí o con la totalidad (la “construcción”).

Para encontrar las relaciones que existen entre las partes de un enunciado

recurrir de nuevo al criterio de reemplazar uno de los constituyentes inmediatos, sólo que, esta vez, la sustitución se realiza con la ausencia de constituyente (es decir, se suprime el constituyente). Si después de la supresión se obtiene un enunciado identificable, admitido por la lengua, se considera que la porción omitida está subordinada al otro constituyente. En una proposición como *El ciudadano inteligente desconfía*, la sustitución por cero de *inteligente* produce el enunciado admisible *El ciudadano desconfía* y de ahí se postula el carácter subordinado de *inteligente* con respecto a *ciudadano*. Si no se obtiene un enunciado identificable, se considera entonces que la parte omitida es superordinada, regente⁵⁶. Tal ocurre en la sustitución por cero de la palabra *ciudadano*, con la que se produce el enunciado inaceptable **inteligente desconfía*.

En función de las relaciones entre los constituyentes y la totalidad, los enunciados se clasifican en **endocéntricos** y **exocéntricos**⁵⁷.

a) Si el enunciado (totalidad o construcción) puede aparecer dondequiera que aparezca uno de sus constituyentes, o ambos, se denomina **endocéntrico**.

α El enunciado *Compra flores* tiene contextos comunes con *compra* (se admiten tanto la totalidad como la parte en contextos como *regalos caros para su suegra*). La construcción endocéntrica corresponde al tipo **subordinativo**.

β El enunciado *Escribe, lee* tiene contextos comunes con cada uno de sus constituyentes inmediatos. Puede decirse:

escribe, lee mucho todos los días
escribe mucho todos los días
lee mucho todos los días

• La construcción endocéntrica corresponde aquí al tipo **coordinativo**.

b) Si la totalidad no tiene contextos comunes con ninguno de sus constituyentes, al enunciado se le llama **exocéntrico**. Así, p. ej., *si regresará pronto* (analizado en constituyentes a base de *si + regresará pronto*) aparece en el contexto *Le he preguntado*, pero éste no admite ni a *si* ni a *regresará pronto*.

Decimos:

Le he preguntado si regresará pronto, pero no
**Le he preguntado si*, o
**Le he preguntado regresará pronto*⁵⁸.

56. En lugar de “subordinado” y “regente”, en la lingüística americana se prefieren los pares terminológicos “satélite - núcleo” o “modifier - head”, que podría traducirse por “modificador - centro”. E. A. Nida distingue, dentro de los constituyentes inmediatos, los nucleares de los periféricos.

57. Cf. BLOCH-TRAGER, *Outline* y GLEASON, *Introducción*.

58. Este ejemplo se vuelve correcto introduciendo una pausa después de *preguntado* (cf. *Le he preguntado. Regresará pronto*), pero entonces obtenemos dos enunciados y no uno solo (correcto), tal como se cuestiona.

Desde luego, el método del análisis en constituyentes inmediatos no resuelve todos los problemas con que tropieza la gramática tradicional, pero logra en cambio ofrecer una presentación unitaria de la estructura de un enunciado descomponiéndolo sucesivamente desde el nivel más alto hasta los morfemas. Siempre se obtienen unidades del mismo nombre (“constituyentes inmediatos”) y además tanto sus relaciones mutuas como sus relaciones con la construcción entera son del mismo tipo.

Otra ventaja práctica del método consiste en que supera o, en todo caso, evita el problema de definir las unidades de diferentes niveles (palabra, parte de oración, oración) y de especificar las relaciones entre morfología y sintaxis.

En tercer lugar, agrupa las unidades en clases, que define a partir de los contextos en que pueden aparecer. La introducción del punto de vista distribucional incluye ciertas ventajas sobre los criterios semánticos de clasificación. Ya no diremos, pues, que los adjetivos son nombres de cualidades o propiedades, sino una clase de palabras caracterizada por aparecer en una serie de contextos (una “clase de contextos”).

En fin, este procedimiento de análisis formal induce, además, al descubrimiento de los tipos de construcción a que se reducen los infinitos enunciados posibles de una lengua dada. Se obtiene con él infinitos enunciados posibles de una lengua dada. Se obtiene con él una modelación más detallada y sistemática que la presentada en gramáticas de tipo tradicional.

1.14. Gramáticas generativas

El método de análisis en constituyentes inmediatos también conoce desde hace años una gran acogida fuera de los Estados Unidos en el marco teórico de las gramáticas generativas. Uno de los tipos de gramáticas generativas se llama precisamente “gramática de constituyentes inmediatos”.

Una gramática generativa aplicada a una lengua se propone producir (y de ahí el nombre de generativa) los enunciados correctos de la misma a partir de un cierto número de símbolos para las palabras (el conjunto de estos símbolos se denomina “vocabulario” o “léxico”) y de ciertos tipos de unidades gramaticales (sintagmas, partes del discurso, morfemas), utilizados ambos a tenor de determinadas reglas. Bajo esta concepción, una gramática generativa constituye un modelo del mecanismo con que operan los hablantes cuando formulan enunciados de la lengua en cuestión. El tipo más frecuente de regla es “ $X \rightarrow Y$ ”, que se lee: “sustituye (o describe) el símbolo X por el símbolo Y ”. Las gramáticas generativas operan con un cierto número de sustituciones simbólicas en conformidad con las reglas formuladas, aplicadas en un determinado orden. He aquí un sencillo ejemplo de gramática generativa basada en constituyentes

inmediatos, que produce un tipo de proposiciones muy comunes en español.

Dados los símbolos de las siguientes palabras españolas: *árbol*, *nube*, *madre*, *pan*, *cantar*, *mirar*, *traer*, *ir*, etc. (estas notaciones tienen un carácter puramente simbólico con relación a las respectivas palabras, cualquiera que sea su forma morfológica; de modo que, p. ej., *árbol* no expresa singular, *cantar* no es el infinitivo presente activo, y así sucesivamente). Se escogen, asimismo, símbolos tales como: *O* = oración, *N* = nombre, *A* = artículo, *FN* = frase nominal (es decir, un nombre acompañado de uno o más determinantes, etc.).

Se dan luego una serie de reglas de generación, como:

(I) $O \rightarrow FN + FV$ (que se lee: ‘recribir el símbolo “oración” por la secuencia de símbolos “frase nominal” + “frase verbal”’);

(II) $FN \rightarrow A + N$ (*A* = artículo, *N* = nombre);

(III) $FV \rightarrow V + FN$ (*V* = verbo);

(IV) $N \rightarrow \text{árbol, nube, madre, pan, ...}$;

(V) $A \rightarrow \text{art}_k$ (art_k = forma de artículo determinado de la categoría morfológica y de la función sintáctica de *N*; el índice *k* puede tomar, así, los valores de “masculino, singular”, “femenino, plural”, etc.);

(VI) $V \rightarrow \text{cantar, mirar, traer, ir, ...}$

Mediante la aplicación sucesiva, en un orden determinado, de las reglas I-VI, se obtiene una oración en español. Aplicando (I), se produce la secuencia *FN + FV*, a la cual se le aplica (II) y se transforma en *A + N + FV*. Al símbolo *FV* se le aplica la regla (III), de modo que la secuencia se convierte en *A + N + V + FN*. Entonces, si a este segundo *FN* se le aplica de nuevo la regla (II), se llega al resultado *A + N + V + A + N*. Sustituimos ahora el primer símbolo *N* por uno de los símbolos de las palabras comprendidas en la regla (IV); suponemos que se elige *madre*. La secuencia adopta así la forma *A + madre + V + A + N*, con los demás símbolos sustituibles de un modo análogo. El proceso entero, pues, se desarrolla en una serie de etapas, cuya representación esquemática sería:

O

FN + FV (según I)

A + N + FV (según II)

A + N + V + FN (según III)

A + N + V + A + N (según II)

A + madre + V + A + N (según IV)

Art_k + madre + V + A + N (según V)

Art_k + madre + traer + A + N (según VI)

$Art_k + madre + traer + A + pan$ (según IV)

$Art_k + madre + traer + art_k + pan$ (según V)

Esta sucesión de operaciones lleva por nombre *derivación*, y se considera cerrada (terminada) si la última secuencia producida no contiene ningún símbolo susceptible de ser rescrito por alguna de las reglas formuladas (en nuestro caso, sólo seis). La última secuencia de una derivación cerrada (como la de más arriba) se denomina *secuencia terminal*⁵⁹.

Así, la secuencia $art_k + madre + traer + art_k + pan$ es un símbolo de la oración *La madre trae el pan*. Para pasar de la representación simbólica a la representación fonemática de esta oración hay que introducir una nueva serie de reglas del tipo:

(I) $Art_k + madre \rightarrow /la + madre/$ (entre barras oblicuas se indica la secuencia de fonemas que componen la forma resultante);

(II) $traer \rightarrow /trae/;$

(III) $Art_k + pan \rightarrow /el + pan/.$

Tales reglas reciben el nombre de *morfofonémicas*. Sólo después de su aplicación cabe decir que se ha formado una oración correcta en español.

Otra clase de gramáticas generativas, las *transformacionales*, utilizan además un tipo de reglas denominadas de transformación, que producen la derivación de oraciones o frases complejas partiendo de oraciones muy simples (llamadas oraciones *nucleares*)⁶⁰. En rigor, utilizan la concepción según la cual todos los enunciados posibles de una lengua no son más que repeticiones variadas en forma y más o menos desarrolladas de un número reducido de tipos o modelos estructurales.

Existen reglas de transformación que sustituyen un elemento de la oración por una oración entera (subordinada). Así se generan períodos a base de oraciones relacionadas en términos de subordinación. Otras reglas consisten en la transformación de un verbo en forma personal a un verbo en forma no personal o a un nombre de la misma familia de palabras (p. ej., *El "Titanic" se hunde* se puede transformar de este modo en *El hundimiento del "Titanic"*). Tales reglas cambian una construcción activa a otra pasiva (p. ej., *La radio dio la noticia* se convierte en *La noticia fue dada por la radio*).

Es significativo el hecho de que se opera mucho más con símbolos de clases de unidades bien conocidas desde la gramática tradicional, en primer lugar con

59. [Para evitar construcciones de tipo *La nube canta el pan*, posible según las reglas aquí esbozadas, y, en general, para todo lo relativo a la noción de "gramaticalidad", cf. CHOMSKY, *Estructuras y Aspectos*, en cuyas reediciones se contienen referencias más recientes].

60. Ésta es la concepción transformativa en su primera variante. Cf. CHOMSKY, *Estructuras*.

oraciones y palabras. Incluso encontramos de nuevo en estudios de gramática generativa las tan criticadas “partes del discurso” de los tradicionalistas.

En las más recientes formulaciones de la teoría transformacionalista, se propone la distinción entre la estructura superficial y la estructura profunda de los enunciados⁶¹. Los sintagmas *amor Dei* (= Dios ama) y *amor Dei* (= Dios es amado) presentan la misma estructura superficial, pero difieren en su estructura profunda. Esto, por consiguiente, contiene en germen las relaciones gramaticales fundamentales mediante la interpretación semántica de los enunciados. Por ejemplo, la estructura profunda del enunciado *El cuaderno rojo falta es: el cuaderno₁ es rojo + relativización + el cuaderno₂ falta* (siendo *el cuaderno₁ = el cuaderno₂*). La conversión de las estructuras superficiales en estructuras profundas se realiza por medio de operaciones llamadas de transformación. Como se ve, en la versión de la teoría que ofrece Chomsky en *Estructuras*, las transformaciones se aplicaban a las oraciones nucleares, mientras que en la versión de *Aspectos* se aplican a las estructuras profundas.

En los últimos años, los teóricos de la concepción transformacional se muestran interesados en la posibilidad de integrar la semántica lexical en el modelo generativo⁶².

1.15. Semántica estructural

Las preocupaciones semánticas se manifiestan ya en la antigüedad: la homonimia, la sinonimia, los valores estilísticos y expresivos de las palabras, los usos figurados de las mismas, la metáfora, etcétera⁶³. Como rama de la lingüística, la semántica no se constituye hasta finales del s. XIX, adoptando como objeto de estudio tan sólo las mutaciones de significado en las palabras, los cambios significativos: reducción, ampliación, transferencia, degradación, ennoblecimiento del sentido. El fundador de la semántica —asimismo creador del término mismo— fue M. Bréal, con el cual se inicia el primer periodo de la semántica, que comprende aproximadamente la etapa que va de 1880 a 1931, momento en que se publica el primer trabajo de semántica estructural, elaborado por J. Trier⁶⁴.

Las concepciones básicas de la semántica estructural son:

1) El vocabulario cubre un sistema de significados; el léxico de una lengua es una *macroestructura* semántica. Ello se demuestra comprobando que, en toda lengua, las palabras se agrupan en “campos léxicos”⁶⁵.

61. Cf. CHOMSKY, *Aspectos*; LEPSCHY, *Lingüística*; GOLOPENȚIA-VASILIU, *Sintaxa* y RUWET, *Introducción*.

62. Cf. CHOMSKY, *Aspectos*; KATZ, *Semantic Theory* y KATZ-POSTAL, *Integrated Theory*.

63. Cf. ROBINS, *Breve historia*.

64. Cf. TRIER, *Wortschatz*.

65. Cf. TRIER, *op. cit.*

2) El contenido de toda palabra constituye una *microestructura* semántica: esto se refiere tanto a la existencia de la polisemia (los significados de una palabra forman un pequeño sistema) como al hecho —particularmente estudiado en nuestros días— de que cada significado es el resultado de la combinación de diferentes *semas* (rasgos mínimos del significado, análogos a los rasgos distintivos que componen los fonemas).

Entre los trabajos más importantes, realizados a la luz de estos criterios, cabe mencionar los de J. Lyons y A. J. Greimas, junto a los planteamientos de B. Pottier, E. Coseriu, G. Mounin, etcétera⁶⁶. Más recientemente, ha aparecido una variante especial de la semántica estructural en el marco de la teoría de las gramáticas generativas⁶⁷.

La idea de que la totalidad de las palabras no es una masa desorganizada era conocida y aceptada incluso antes de la creación de la semántica estructural. Desde el punto de vista de su comportamiento gramatical, los elementos del léxico se agrupan en las llamadas “partes del discurso” y, por otra parte, existe también una estratificación de las palabras en función del estilo (poético, familiar, argot, técnico, etc.). Se sabía igualmente que las palabras se reúnen en las llamadas series asociativas o “constelaciones”⁶⁸. Este concepto incluye al de “familias de palabras”, conocido ya por la lingüística tradicional, y al de “serie sinonímica”. Jost Trier concibe el campo semántico como un conjunto de relaciones entre palabras, las cuales adquieren un significado justamente en virtud de estas relaciones. Un significado sólo existe, por tanto, en el interior de un campo. Cada campo forma, junto con otros, un campo más amplio, hasta que en reuniones sucesivas se llega al léxico de la lengua, el campo semántico más vasto. Este último tiene el aspecto de un mosaico, sin huecos ni superposiciones.

Más tarde, los defensores de la semántica estructural combatieron esta concepción, basándose en que el análisis minucioso sobre el vocabulario de diversas lenguas demuestra que no existe una organización tan armoniosa y perfecta. Particularmente importante es el hecho de que la misma palabra, e incluso el mismo significado de una palabra, puede participar en más de un campo. De una organización sistemática y de un vocabulario como macroestructura semántica sólo podremos hablar si recurrimos al análisis de las palabras en rasgos distintivos (“semas”). Este tipo de análisis, denominado *componencial*, se propone demostrar que el enorme conjunto de significados expresados por la totalidad del léxico puede reducirse a diversas combinaciones de un número más reducido de entidades semánticas mínimas (“átomos semánticos” o “se-

66. Cf. LYONS, *Structural Semantics*; GREIMAS, *Semántica*; POTTIER, “Recherches”; COSERIU, “Sémantique”; MOUNIN, *Traducción*; [FERNÁNDEZ, *Introducción*].

67. Cf. KATZ-POSTAL, *Integrated Theory*; KATZ, *Semantic Theory*; VASILIU-GOLOPENȚIA, *Transformational Syntax*, y, para la crítica, COSERIU, *Einführung* y UHLENBECK, *Critical Comments*: [LYONS, *Semantics*].

68. Cf. SAUSSURE, *Curso*.

mas”)⁶⁹. Hasta el presente, disponemos de tales análisis en semas para zonas relativamente reducidas del léxico: términos de parentesco (donde los semas son “sexo”, “grado”, “línea” —directa o colateral—, “tipo” —parentesco de sangre o político—, “dirección” —ascendente o descendente—), términos de vivienda, animales domésticos, edad, colores, etcétera. Bernard Pottier ha propuesto, en un trabajo sobre los nombres franceses que se refieren a asientos de diversos tipos, los siguientes semas: “con/sin respaldo”, “con/sin patas”, “para una persona/para más de una persona”, “con/sin brazos”, “de material rígido/blando”, junto con el sema, común a todos aquellos términos, “para sentarse”⁷⁰. Para este grupo y para otros similares, las lenguas disponen de una palabra que incluye a las demás: en español *asiento* (fr. *siège*) para los muebles que sirven para sentarse; o *pariente* (rum. *rudă*, it. *parente*, etc.) para los grados de parentesco, etcétera. Tales palabras se denominan *archilexemas*.

69. Cf. GREIMAS, *Semántica*.

70. Cf. POTTIER, “Recherches”.

2. Sinopsis general de la lingüística matemática

2.1. Objeto y método

La lingüística matemática estudia los fenómenos de lengua mediante procedimientos matemáticos. Así, pues, su objeto es la lengua y su método es matemático.

Teniendo en cuenta, por una parte, que los métodos matemáticos sólo se aplican a determinados aspectos de los fenómenos, precisamente a los cuantitativos y formales, y, por otra parte, que la única ciencia que ha elaborado métodos perfeccionados para el estudio de tales aspectos es la matemática, cabe concluir que la lingüística matemática se puede definir también como el estudio de los aspectos cuantitativos y formales de los fenómenos de la lengua.

Y aún más, considerando los objetivos inmediatos, prácticos, de la lingüística matemática, así como los problemas prácticos que estimulan su desarrollo, cabe definir asimismo la lingüística matemática como el estudio de la lengua en tanto que objeto con el que trabaja la técnica. En virtud de esta acepción, la lingüística matemática también recibe el nombre de lingüística aplicada.

2.2. Cómo y cuándo ha nacido la nueva disciplina

Para comprender la aparición, el desarrollo y las perspectivas de la lingüística matemática es preciso relacionarla con determinados acontecimientos y procesos de la historia científica y técnica y con su evolución general.

Como disciplina constituida, la lingüística matemática aparece muy tarde, apenas hace más de veinte años, casi simultáneamente en la U.R.S.S. y en los EE-UU, y después en otros países, como consecuencia del desarrollo vertiginoso de los ordenadores electrónicos. Entre otras utilidades no aritméticas de las máquinas electrónicas de cálculo, una de las más interesantes ha consistido en la

traducción de una lengua a otra. El nacimiento, en estas condiciones, de la lingüística matemática viene a confirmar el papel especialmente destacado que desempeña la técnica en el desarrollo de la ciencia.

La construcción de ordenadores electrónicos de gran velocidad ha puesto de manifiesto ciertas tendencias que desde mucho tiempo atrás se abrían camino en la lingüística, en la matemática, en la técnica y en la actividad práctica de los hombres, y que expresaban exigencias sociales y científicas sumamente candentes. Pensamos en la conciencia cada vez más clara de los lingüistas acerca de la importancia que tiene la estructura en la lengua, así como en la estimación cada vez mayor que han empezado a tener en matemática, especialmente desde finales del siglo pasado, las lenguas simbólicas, formalizadas. Pensamos especialmente en la necesidad cada vez más imperiosa de que se perfeccionen los métodos de transmisión de la lengua por diferentes canales de comunicación, los métodos de escritura rápida y de descifrado de esta escritura (los sistemas de estenografía), los dispositivos automáticos de lectura, la preparación de la información en las máquinas informativas, la recensión automática de la bibliografía científica, el mando oral de las máquinas o la elaboración de sistemas de escritura para lenguas solamente habladas; pensamos, en fin, en el problema de reconstituir, sobre una base científica, los textos mutilados, en la medición de la redundancia contenida en un texto (cf. 8.14) o en el perfeccionamiento de los métodos para enseñar las lenguas a partir de la estructura estadística del léxico y de la gramática.

2.3. El origen matemático de la lingüística matemática

Desde hace siglos y junto al desarrollo del álgebra, la matemática ha empezado a utilizar determinados grupos de signos y de procedimientos para operar con ellos. A mediados del siglo pasado, George Boole sienta las bases del álgebra lógica al observar que ciertas inferencias lógicas resultan fácilmente estudiables con ayuda de un cálculo algebraico. De este modo se extiende también en la lógica el hábito de trabajar, mediante reglas bien determinadas, con ciertos sistemas simbólicos.

A finales del siglo pasado, como consecuencia del desarrollo impetuoso de la matemática y en especial de la teoría de los conjuntos, aparecen ciertas paradojas que ponen en duda la legitimidad de las demostraciones matemáticas y de los conceptos con que se había operado hasta entonces. Nociones tales como “número”, “conjunto” o “infinito” abocaban a graves callejones sin salida. Así nació la necesidad de someter las demostraciones matemáticas a un severo examen. Había que elaborar, en suma, una verdadera ciencia de las demostraciones matemáticas, una ciencia en que los objetos de estudio fuesen precisamente las teorías matemáticas en sí mismas.

La aritmética fue la primera rama matemática sometida a este tipo de control. Se descompuso, analizó y explicitó cada demostración en pasos lógicos elementales, a fin de comprender con exactitud las premisas que intervienen en el razonamiento y los procedimientos lógicos empleados. Con ello, la aritmética, la teoría de los conjuntos y la geometría han sido colocadas sobre una base axiomático-deductiva, es decir construidas como sistemas lógico-matemáticos formales en los que se hace abstracción de cualquier significado intrínseco de los símbolos, y en los que, en cambio, partiendo de un número bien determinado de relaciones simbólicas, se obtienen las demás proposiciones por medio de unas reglas establecidas.

Esta forma de abordar los problemas relativos a los fundamentos de la matemática se ha convertido en una preocupación decisiva para la lógica matemática y, en ciertos aspectos, en su propio modo de existir. Mediante el empleo consecuente de los sistemas lógico-matemáticos formales —sistemas que reciben el nombre de “lenguas artificiales” o “lenguas formalizadas”— se definen cada vez con mayor precisión las relaciones de la matemática con la lógica-matemática, por una parte, y con la lingüística, por otra. El carácter de “lengua” de estos sistemas formales se vuelve más pronunciado con el desarrollo de la semántica lógica, que estudia los vínculos existentes entre los sistemas formales y la realidad que éstos reflejan.

Huelga insistir sobre las diferencias que hay entre las lenguas naturales y las artificiales. Las lenguas naturales alcanzan un grado de complejidad incomparablemente mayor; como contrapartida, en una lengua artificial no cabe la ambigüedad en el empleo de los símbolos.

2.4. El origen técnico de la lingüística matemática

Paralelamente al proceso descrito en el apartado anterior, tiene lugar otro en el marco de la técnica. Con el desarrollo de los medios de transmisión informativa, adquiere una importancia cada vez mayor el problema de codificar las lenguas naturales. Las letras del alfabeto y los sonidos del habla presentan una forma demasiado complicada para ser transmitidos como tales. Con las máquinas hay que proceder de un modo más simple.

De esta manera se acrecienta el interés por la teoría de los códigos, la cual, hasta cierto punto, pertenece a la matemática. Tales códigos constituyen también lenguas formalizadas, sistemas de símbolos con que se opera por medio de reglas determinadas. La teoría de los códigos ha experimentado en los últimos veinte años un rápido crecimiento junto con la amplitud que han alcanzado asimismo algunas ramas de la cibernética. Para los lingüistas, dos de estas ramas ofrecen una especial importancia: la teoría de la información y la traducción

electrónica. Esta última ha requerido la construcción de lenguas artificiales, formalizadas, o modelos formalizados de lengua, como algunos todavía dicen. Así, en este punto de relación con las lenguas formalizadas, se han unido y valorificado las antiguas preocupaciones y los hábitos específicos de la matemática y de la lógica. En estos métodos matemáticos de lengua, se fijan rigurosamente las reglas para la utilización de las expresiones y del significado de las mismas, de modo que no pueda influir ningún factor psicológico en su construcción. Un problema más delicado, que aquí soslayamos, consiste en saber en qué medida constituyen estos modelos una aproximación a las lenguas naturales o un medio de profundizar en su estudio.

Las perspectivas excepcionales que existen en la actualidad en lo que se refiere a la resolución de ciertos problemas de la lingüística aplicada están estrechamente relacionadas con el reciente desarrollo de una nueva rama científica, la teoría de la información (cf. 8.7). Esta disciplina considera la información no desde el punto de vista de su contenido mismo, sino de su aspecto cuantitativo; estudia, al mismo tiempo, los procedimientos óptimos para codificar la información con el propósito de transmitirla a través de distintos canales comunicativos. La importancia de esta disciplina en el estudio de la función comunicativa de la lengua se pone de manifiesto por la misma definición de su objeto. De este modo, la noción de *entropía* (cf. 8.9), fundamental en la teoría de la información, se ha mostrado útil en el estudio de la lengua como proceso probabilístico. La entropía de la lengua mide el estado de indeterminación, lo arbitrario que existe en ella. Especialmente significativas son la entropía gramatical, que mide el grado de arbitrariedad que existe en la estructura gramatical de la lengua, y la entropía fonológica o medición de lo arbitrario en la estructura fonológica. La evaluación de la entropía en las lenguas naturales requiere una cantidad tan grande de operaciones aritméticas, que la ayuda de las máquinas resulta, en este terreno, esencial. (Para la entropía de la lengua rumana no existen, hasta el momento, más que conclusiones parciales. La fijación de la entropía léxica y gramatical del rumano, en sus distintas fases de desarrollo, constituiría un auxiliar inestimable para comprender las leyes objetivas que gobiernan su evolución (cf. 8.15).)

Entre los problemas prácticos que estimulan el desarrollo de la lingüística matemática, el que más interesa en la actualidad a los círculos lingüísticos es el de la traducción automática. ¿A qué se debe este fenómeno? Las principales dificultades que plantea la traducción automática no sólo conciernen a la ingeniería, como podría creerse, sino también a la lingüística. Se trata, en primer lugar, de un análisis de los hechos de lengua que exige un espíritu nuevo, un análisis que valorifica, al propio tiempo, las mejores tradiciones en el estudio de las estructuras lingüísticas. Una cuestión fundamental consiste en establecer el modo o el tipo de lengua a través de la cual el hombre se ponga en comunicación con la máquina. Una preparación de esta naturaleza sobre las lenguas naturales

se halla en la base de la teoría de los modelos matemáticos de lengua. Tales modelos son aproximaciones axiomático-deductivas de las lenguas naturales y se apoyan sobre algunos conceptos y nociones habituales en la teoría de los conjuntos y en el cálculo de probabilidades. En el marco de un modelo como éste, se tienen en cuenta las relaciones formales y posicionales entre los elementos de la lengua, así como su distribución (cf. 1.12). Y se excluye cualquier relación que deba recurrir al significado.

2.5. La traducción de una lengua a otra con máquinas electrónicas de cálculo

Una de las aplicaciones más interesantes de las máquinas electrónicas consiste en su empleo para la traducción de una lengua a otra. Ya en 1933, el sabio soviético Smirnov-Troianski conjeturó la posibilidad de realizar traducciones mecánicas, pero sólo con la aparición de las máquinas electrónicas de cálculo se facilitaron las condiciones materiales para poner en práctica sus ideas. Durante los últimos veinte años se ha pasado a la experimentación de las traducciones automáticas, primero del inglés al ruso (en la Unión Soviética) y del ruso al inglés (en los Estados Unidos).

La traducción automática de una lengua a otra se hace con ayuda de las máquinas electrónicas de cálculo. Para el futuro se espera construir unas “máquinas traductoras” electrónicas especialmente dispuestas para la traducción. Para ello, deben sustituirse las letras por impulsos eléctricos. Se introduce, entonces, en la máquina un “diccionario electrónico” correspondiente a las dos lenguas y dotado de un número no excesivamente grande de palabras, aproximadamente varios miles. Estas palabras son elegidas tanto de la lengua común como de la esfera lingüística a que pertenecen los textos sometidos a traducción. Conviene realizar, pues, un estudio previo sobre la frecuencia de las palabras en las lenguas respectivas, ya que sólo un estudio de esta naturaleza permite introducir en la máquina un diccionario científicamente confeccionado.

2.6. El principio regulador de las máquinas electrónicas

El principio funcional de las máquinas electrónicas de cálculo consta de una escritura especial de los números con el auxilio de válvulas electrónicas. Las cifras 0 y 1 se ponen en correlación con dos estados posibles de una válvula electrónica. La cifra 0 corresponde al momento en que la válvula interrumpe la corriente y, a su vez, la cifra 1 corresponde al momento conductor. Todo número admite una representación por medio de una determinada secuencia de cifras 0 y 1, repetidas de un modo conveniente, y, por tanto, de una secuencia de

estados de interrupción y conexión en la corriente de una válvula electrónica o de una secuencia de válvulas cerradas y abiertas. Éste es el sistema de numeración binaria o de base 2. En él, el número 23, p. ej., se representa mediante la notación 10111, ya que $23 = 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$; el número 15 se representa 1111, ya que $15 = 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$, y así sucesivamente¹. Así, pues, para la escritura del número 23, la máquina emplea cinco válvulas, de las cuales cuatro están abiertas y una cerrada; mientras que para escribir el número 15, la máquina se sirve de cuatro válvulas abiertas. Aunque la escritura de un número con base 2 es más larga que con base 10, la elevada velocidad con que trabaja una máquina electrónica de cálculo elimina este inconveniente.

2.7. Cómo se buscan las palabras en el diccionario electrónico

Para la traducción con máquinas electrónicas de cálculo, se escriben las letras del alfabeto en forma de códigos cifrados, como los que se utilizan en telegrafía. Por lo tanto, cada palabra se representa por medio de una cadena de cifras, que la máquina registrará a base de una determinada secuencia de válvulas abiertas y cerradas.

La primera operación que efectúa la máquina consiste en “leer”, en trabar conocimiento con la frase que debe traducir. Esta operación puede automatizarse fácilmente si se escribe la frase mediante unas perforaciones como en las

1. [En general, toda expresión numérica de cualquier base se interpreta como una suma indicada de tipo:

$$\Sigma = a_n \cdot x^{n-1} + \dots + a_3 \cdot x^2 + a_2 \cdot x^1 + a_1 \cdot x^0$$

donde Σ es la expresión resultante, a un dígito del paradigma, n un subíndice que alude a la posición del dígito en la expresión resultante (a partir de la derecha, en las representaciones usuales) y x la base de numeración. Entonces, 23, en numeración decimal, equivale a $a_2 \cdot 10^1 + a_1 \cdot 10^0$, es decir:

$$23 = 2 \cdot 10^1 + 3 \cdot 10^0;$$

en numeración binaria, tal como se indica en el texto:

$$10111 = 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0;$$

en numeración ternaria (con un paradigma “0, 1, 2”):

$$212 = 2 \cdot 3^2 + 1 \cdot 3^1 + 2 \cdot 3^0,$$

y así sucesivamente. Obsérvese, sin embargo, que en estas dos últimas comprobaciones se ha expresado x y sus exponentes en numeración decimal y con ello 10111 y 212 han quedado transferidos a la base 10. Para mantener las expresiones en sus propias bases, hay que indicar x y sus exponentes en su numeración respectiva:

$$10111 = 1 \cdot 10^{100} + 0 \cdot 10^{11} + 1 \cdot 10^{10} + 1 \cdot 10^1 + 1 \cdot 10^0,$$

$$212 = 2 \cdot 10^2 + 1 \cdot 10^1 + 2 \cdot 10^0,$$

de donde se deduce que x tiene siempre el aspecto de 10, aunque con distinto valor, si se utilizan paradigmas ordenados en forma de 0, 1, 2, ... Obsérvese también que el número de dígitos del paradigma es inversamente proporcional al número de posiciones necesarias (n) para una expresión dada].

bandas telegráficas. Sigue, luego, la búsqueda de cada palabra de la frase en el diccionario incorporado de antemano a la máquina, en el cual las palabras se hallan inscritas, como se ha indicado antes, con ayuda de válvulas abiertas y cerradas. Para ello se compara cada palabra de la frase con cada palabra del diccionario de la máquina a base de ir restando sucesivamente el número cifrado de la palabra en cuestión a cada uno de los números que componen las palabras inventariadas en el diccionario. Si en éste existe un número tal cuya resta respecto del número introducido da 0, significa que la palabra ha sido hallada en el diccionario. Si, por el contrario, el resultado de la resta no coincide nunca con 0, significa que la palabra respectiva no figura en el diccionario de la máquina. Esta operación se verifica con extrema rapidez. Nótese que, si la máquina efectúa, p. ej., 10 000 restas por segundo, la búsqueda de una palabra dura, como máximo, una décima de segundo.

2.8. La máquina también “aprende” gramática

Las primeras experiencias de traducción automática, anteriores al año 1955, se han dedicado al léxico, soslayando así los problemas de gramática. Era el período considerado de traducción “palabra por palabra”, en el cual se atendía exclusivamente a la situación individual de las palabras y se ignoraba cualquier otro tipo de análisis lingüístico. Así, en el diccionario de la máquina sólo se registraba el significado genérico de las palabras y se hacía abstracción de las modificaciones flexivas. Pero semejante procedimiento se mostró insuficiente. A los lectores que deseen convencerse de ello les recomendamos la siguiente experiencia: se toma un texto cualquiera y se conserva la raíz de cada palabra, eliminando los elementos desinenciales y sintácticos. Por ejemplo, del rumano *caselor* (“de/a las casas”) se retiene *cas*, de *vorbesc* (“hablo/hablan”), *vorb*, y así sucesivamente. Se comprobará que desaparece la comprensión del texto resultante, pese a que se conservan indicios vagos del tema tratado. De este modo, la atención de los investigadores se ha orientado cada vez más hacia el análisis concreto de los hechos lingüísticos: el análisis de la frase en la lengua de partida y la síntesis de la frase en la lengua destinataria.

A diferencia del diccionario, cuyo contenido depende del dominio al que pertenece el texto traducido, la programación relativa al establecimiento de una correspondencia de estructura gramatical entre dos lenguas —la lengua de que se traduce y la lengua en que se traduce— es más bien independiente de la esfera temática del texto, la cual puede pertenecer a la matemática, a la física, a la literatura o a cualquier otra rama de actividad. El carácter universal de estas programaciones gramaticales ha sido puesto de relieve por la lingüista soviética I. K. Belskaia. Ella demostró que la misma programación gramatical para la

traducción del inglés al ruso resulta igualmente válida al aplicarse a un texto de matemáticas que a una narración de Dickens.

Es sabido que la mayoría de palabras de la lengua rumana presentan diversas formas a causa de la flexión. Un verbo, p. ej., admite, como en español, distintas formas correspondientes a modificaciones de modo, tiempo, número y persona. Como el diccionario de la máquina no puede contener una cantidad excesivamente grande de palabras, no cabe considerar cada una de las formas flexivas de un verbo o de un nombre como si se tratara de una palabra diferente. Por ello, el diccionario de la máquina no es tanto un diccionario de palabras como un diccionario de raíces léxicas. Así, pongamos por caso, la palabra *omului* ('del/al hombre') no figura en el diccionario; en su lugar aparece *om* ('hombre'). La máquina, entonces, "aprende" a separar la terminación *ului* a fin de identificar el nombre *om*. La máquina aprende asimismo a reconocer el hecho de que la terminación *ului* corresponde al genitivo o al dativo singular. Estas operaciones se introducen también sobre una base binaria. Por ejemplo, puede convenirse que la cifra 1 corresponde al singular y 0 al plural, que 1 corresponde a los casos nominativo y acusativo, y 0 al genitivo y dativo, que 1 corresponde al género masculino y 0 al femenino (el neutro se reduce al masculino en singular y al femenino en plural), etcétera². Por este procedimiento pueden describirse todas las características morfológicas de una palabra a través de una secuencia formada por las cifras 1 y 0 ó a través de un cierto número de válvulas abiertas y cerradas. Quedan, claro está, numerosos problemas gramaticales que resolver, generalmente con ayuda del contexto (en rumano, p. ej., la distinción entre genitivo y dativo (cf. 4.5)), pero por el momento nos basta con haber presentado el funcionamiento básico de estos recursos de automatización (cf. 7.2.).

2.9. Otras operaciones al alcance de la máquina

Después de haber analizado gramaticalmente el texto introducido en la máquina, se inscriben todas las informaciones obtenidas en la "memoria" de la misma. La memoria constituye una de las grandes innovaciones que han allegado las modernas máquinas de cálculo. Se caracteriza por la capacidad de conservar los resultados de ciertas operaciones intermedias para ser utilizados ulteriormente. La memoria de la máquina se construye, por lo general, magnéticamente, de modo que la cifra 0 corresponda a la magnetización negativa y 1 a la positiva. El empleo de células fotoeléctricas parece gozar de un amplio futuro en la construcción de memorias automáticas.

2. [Estas precisiones se refieren a la estructura nominal del rumano. En español, claro está, habría que reajustar estas variables. Cf. para el género en rumano, 3.29, n. 74].

Continúa después la operación de síntesis, es decir la composición del texto correspondiente en la lengua destinataria de la traducción. Así, pues, en la máquina se establece la correlación entre las palabras de una lengua con las palabras de la otra, fase en la que entra en funcionamiento la segunda parte del diccionario correspondiente a la lengua a la que se traduce. Se elabora asimismo una serie de reglas con base binaria para la composición de las formas flexivas equivalentes a las formas de la lengua inicial. Se establece, además, una correspondencia en la configuración sintáctica de las dos lenguas (cf. 3.13). Si una palabra de la lengua inicial no tiene equivalente en la otra lengua, se reproduce sin traducción. Y después de todas estas operaciones, la máquina imprime el resultado sobre una banda³.

2.10. La importancia del conocimiento de las leyes estadísticas del vocabulario

Para un feliz desarrollo de las traducciones automáticas resulta extremadamente útil conocer las leyes estadísticas que gobiernan los fenómenos de la lengua. Así, existen leyes estadísticas capaces de determinar la longitud de un texto, es decir el número de palabras del texto, una vez conocido el vocabulario del mismo o conjunto de palabras distintas que contiene. Y viceversa, conociendo la longitud de un texto, es posible determinar el vocabulario y, por lo tanto, la extensión del diccionario que debe introducirse en la máquina (cf. 2.20).

Una dificultad esencial para la traducción automática consiste en la existencia de palabras con más de un significado. Es uno de los motivos por los cuales, en la actualidad, sólo se experimenta con la traducción de textos científicos y técnicos, en los que suelen predominar palabras con significado único y preciso. Existen, sin embargo, leyes estadísticas capaces de determinar la proporción de palabras polisémicas de un texto cualquiera.

Para ilustrar la eficacia de las leyes estadísticas del vocabulario en la traducción automática, nos referiremos a un solo ejemplo: a partir de las investigaciones emprendidas por Zipf sobre la frecuencia de las palabras, se ha comprobado que la mayoría de las palabras polisémicas acusan, al mismo tiempo, una frecuencia elevada en el texto considerado⁴. Y como, por otra parte, las primeras 50 palabras, en orden decreciente de frecuencia, cubren aproximadamente la mitad del texto, de ello resulta que el esfuerzo para resolver el problema de la multiplicidad de significados debe concentrarse, en primer lugar, sobre aquellas 50 palabras.

3. Para una descripción más detallada del modo como se desarrolla una traducción mecánica, cf. PÁNOV, "Traducerea", y 7.2.

4. [Según ZIPF (*Human Behavior*), las diferentes acepciones de una palabra tienden a ser proporcionales a la raíz cuadrada de su frecuencia relativa (cf. 2.20)].

2.11. Dificultades relativas a la introducción de la gramática en la máquina

La traducción automática plantea una cantidad de problemas delicados, cuya resolución requiere el esfuerzo conjunto de lingüistas, ingenieros y matemáticos. Hemos visto que, para el desarrollo de la traducción en óptimas condiciones, es conveniente introducir en la máquina no sólo un vocabulario, sino también una gramática: una determinada morfología y una determinada sintaxis. Las palabras aparecen en el texto con ciertas formas flexivas, y las frases, a su vez, ofrecen una tónica particular, que la máquina debe captar a fin de encontrar las correspondencias adecuadas en la lengua en que traduce. Cuando se ha intentado introducir la gramática en la máquina, se ha comprobado que surgen dificultades considerables a causa del carácter impreciso y de las prolijas especulaciones descriptivas que abundan en la gramática tradicional. Para que la máquina “entienda” la declinación de los nombres, la conjugación de los verbos y —sobre todo— la sintaxis de la lengua en cuestión, es necesario reducir la totalidad del sistema morfológico y sintáctico a un número bien determinado de conceptos y reglas claros, precisos y, por tanto, faltos de cualquier equívoco o ambigüedad.

¿Es acaso posible esto? ¿Es compatible con el alto grado de complejidad que presentan los fenómenos de la lengua? Si la respuesta es afirmativa, entonces toda la gramática tradicional experimentará una renovación y sus bases científicas se consolidarán. Pero este problema es demasiado arduo para resolverlo de un golpe. El único método consiste en abordarlo por etapas. En una primera fase, pueden evitarse aquellos fenómenos que sólo aparecen de un modo esporádico. Esto presupone que se hayan estudiado de antemano las distintas alternativas morfológicas y sintácticas desde el punto de vista frecuencial. Sólo por este camino podemos averiguar cuáles son las situaciones excepcionales y, por tanto, prescindibles en una primera aproximación.

2.12. ¿Qué ayuda puede proporcionar la matemática?

¿En qué forma puede contribuir la matemática a resolver este problema? Por matemática solemos entender un cuerpo determinado de conocimientos especializados, un cierto número de definiciones y teoremas, una cierta técnica de demostración y una cierta técnica de cálculo. ¿Qué relación, entonces, podría tener con lo expuesto más arriba una fórmula cualquiera de trigonometría, un teorema relativo a los triángulos o a las derivadas, una ecuación o un cálculo algebraico? Evidentemente, una relación directa de este tipo no existe entre matemática y gramática. Pero si concebimos la matemática de un modo más amplio, como ciencia que no se circunscribe tan sólo a una porción de consideraciones for-

males de estricta especialidad, sino como ciencia que desempeña un papel bien determinado en el estudio de la realidad que nos rodea, entonces la cuestión cambia por completo. La matemática se ocupa de los aspectos numéricos de los fenómenos, pero también de la forma de las relaciones que entablan entre sí los elementos de un conjunto de objetos. Los aspectos cuantitativos y formales pueden estudiarse, hasta cierto punto, independientemente de la naturaleza específica de los fenómenos u objetos en cuestión. Consideremos, p. ej., la morfología. Desde el punto de vista morfológico, la lengua se presenta como un conjunto de palabras agrupadas en clases del siguiente modo: dada una palabra x , la clase que la contiene estará formada por todas las palabras que sean formas flexivas de x . Si, p. ej., x es un nombre, la clase de x se compone de todas las formas que se obtienen a partir de x al modificarle el número y el caso y al pasar de la forma que contiene el artículo indeterminado (o determinado) a la que contiene el artículo determinado (o indeterminado)⁵. En cuanto observamos de esta manera la flexión de las palabras accedemos ya al dominio de la matemática, no sólo en tanto que cuerpo de conocimientos especializados, sino más bien como un modo particular de abordar los problemas o una forma determinada de pensar.

La división en clases que acabamos de proponer es idéntica, desde el punto de vista de sus propiedades formales, a muchas otras que se efectúan habitualmente en álgebra y geometría y se denomina, en términos matemáticos, división en **clases de equivalencia**. En efecto, convengamos que dos palabras x e y se hallan en relación R si ambas pertenecen a la misma clase. Para mostrar que la relación R es una relación de equivalencia, es preciso verificar si es reflexiva, simétrica y transitiva⁶. Dejamos a cuenta del lector la verificación de estas propiedades en la hipótesis de la ausencia de homonimia.

De una manera análoga, si admitimos como equivalentes dos palabras x e y entonces y sólo entonces cuando la sustitución de x por y o la sustitución de y por x en cualquier frase correctamente construida produzca otra frase también correctamente construida, comprobamos que tiene lugar una división de las palabras en clases de equivalencia. Esta división tiene una importancia decisiva en el estudio de las estructuras lingüísticas. Pero reclama, por supuesto, definir con precisión el concepto de "frase correctamente construida".

5. [Se refiere a la estructura nominal del rumano, donde, como es sabido, el artículo determinado va puesto al nombre y forma con él un todo inseparable en la mayoría de modificaciones casuales oblicuas. El artículo indeterminado va antepuesto y acusa independientemente las desinencias declinacionales. El nombre rumano cambia, entonces, sus elementos terminales en virtud de esta característica].

6. [Estas características se expresan como sigue:

"para todo x , xRx " (reflexividad),

"si xRy , entonces yRx (simetría),

"si xRy e yRz , entonces xRz (transitividad)].

Ahora bien, esto no es más que un principio. Los intentos de formalización se han ido aplicando ya sobre aspectos gramaticales mucho más complejos que los morfológicos. El modo de pensar que depara la matemática nos brinda la posibilidad de separar más fácilmente los fenómenos lingüísticos puramente formales de aquellos que implican un significado. En este sentido llegamos a conceptos y definiciones del mismo tipo que los utilizados por la matemática y que “entiende”, en consecuencia, la máquina. Lo que necesita la traducción automática es precisamente una gramática construida sobre la base de estos razonamientos.

El programa para la traducción de una lengua a otra es comunicado a la máquina por medio de un algoritmo o sucesión bien determinada de operaciones que aquella debe ejecutar (cf. 4.1 y ss.). Para ello, hay que redactar este algoritmo en una lengua inteligible para la máquina y, por tanto, interpretable en los términos de su código electrónico. Es lo que se llama una *metalengua*. El conjunto de símbolos que constituyen la metalengua en que se redacta el algoritmo de traducción automática representa para el proceso lo mismo que las fórmulas H_2O y $NaCl$ para la química o para la matemática $\sin x \cos y + \sin y \cos x$ y $(-b \pm \sqrt{b^2 - 4ac}) / 2a$. La formación de una metalengua de esta suerte reclama un estudio de inspiración lógico-matemática para los hechos de lengua.

2.13. Antecedentes en el proceso de matematización de la lingüística

Por “matematización” no hemos de entender “panmatematismo”, una subordinación absoluta de las demás ciencias a la matemática por la que ésta se erigiría en objeto único de interés científico. Por matematización de una ciencia se comprende la separación y descripción adecuada de los aspectos cuantitativos, formales, relacionales o estructurales de la ciencia respectiva. La intervención de la matemática en tantos dominios de investigación no se explica por alguna superioridad frente a las demás ciencias, sino por el hecho de que estudia unos aspectos relativamente pobres y justamente por eso muy generales, casi comunes a todos los fenómenos.

La idea de aplicar el pensamiento matemático al estudio de los fenómenos lingüísticos es bastante vieja. La hallamos, ya en el siglo pasado, en la obra de V. I. Buniakovski.

He aquí lo que escribió este matemático el año 1847:

“...Permitaseme añadir algunas palabras sobre otra aplicación del análisis de probabilidades que al parecer nadie ha mostrado. Esta nueva aplicación se refiere a las investigaciones gramaticales y etimológicas de una lengua cualquiera y a la filología comparada... Cuando se trata sobre una lengua, ante todo suponemos tener su descripción aritmética detallada o, dicho de otro modo, su

estadística, es decir los índices numéricos relativos a la suma total de las palabras de aquella lengua, la división de estas palabras en partes del discurso, su subdivisión por el número de letras, por las letras iniciales, por la terminación, etcétera. Aquí también entrarán las informaciones sobre las reglas generales, las excepciones de distintos tipos, las palabras evidentemente prestadas de otras lenguas, etcétera. Son resultados numéricos, cuyo análisis severo exige, sin duda, consideraciones matemáticas. Disponiendo de tales datos estadísticos referentes a dos o más lenguas, podemos compararlas de distintas formas y los resultados obtenidos gozarán de una autoridad que no siempre pueden presentar los filólogos en su defensa, dada la situación actual de la ciencia.”⁷

En 1894, el gran lingüista Ferdinand de Saussure escribía: “Les quantités du langage et leurs rapports sont régulièrement exprimables, dans leur nature fondamentale, par des formules mathématiques”, y más tarde, refiriéndose a la expresión más simple de los conceptos lingüísticos, añadía: “L’expression simple sera algébrique ou elle ne sera pas”. También en relación con la lingüística, afirmaba Saussure en 1911: “On aboutit à des théorèmes qu’il faut démontrer”.

En cuanto al problema fundamental de las relaciones entre lo continuo y lo discreto en la lengua, otro gran lingüista, Baudoin de Courtenay, intentó servirse de algunas nociones básicas de la matemática. En 1909, también él expresó el convencimiento de que la lingüística tendría en el futuro vínculos estrechos con la matemática. Según la concepción de Baudoin de Courtenay, la lingüística precisa de métodos cuantitativos y de métodos deductivos lo más perfeccionados posible; la lingüística debe realizar, con vistas al perfeccionamiento de sus métodos, algo semejante al modo por el cual la matemática convierte el estudio de los conjuntos infinitos en el de los conjuntos numerables.

Con relación al nexo entre matemática y lingüística, cabe citar asimismo las afirmaciones de Emile Borel en el IV Congreso Internacional de Matemáticos de 1909 y las reflexiones de Jacques Hadamard, quien, en 1943, observando los progresos alcanzados por la lingüística con la promoción de los métodos estructurales, declaró, de un modo sin duda metafórico, que la lingüística es un puente entre la matemática y la humanidad.

El gran lingüista R. Jakobson observa que tanto el teórico de la lengua como el investigador de un determinado estadio o de la evolución entera de la misma, así como los que se interesan por las ramas modernas de la lingüística aplicada se sienten atraídos por disciplinas matemáticas como la lógica matemática —en especial por la teoría de las funciones recursivas y la teoría de la automoción—, la topología, el álgebra, la teoría de la comunicación y los métodos probabilísticos.

7. Cf. BUNIAKOVSKI. “O vozmozhnosti”.

2.14. El sentido del proceso de matematización

A lo largo de la historia de la ciencia se puede seguir un proceso constante de matematización que afecta a todos los dominios científicos. Al principio, esta influencia afectó a la astronomía, la mecánica y la física. Luego, a partir del siglo pasado, ha comenzado a manifestarse la matematización de las ciencias que se ocupan de la materia viva, en especial la biología. En la actualidad, se ha llevado tan lejos la matematización de la biología, que ya se plantea el problema de su axiomatización. Se ha llegado a describir matemáticamente la actividad nerviosa superior y se halla en el orden del día la cuestión de describir, también matemáticamente, todo el pensamiento humano. De este modo, la matemática ha penetrado en la psicología; no es raro, pues, que, con el estudio matemático del pensamiento y de la fisiología cerebral, resultara inevitable la posibilidad de ocuparse del lenguaje y de la lengua mediante procedimientos matemáticos.

La descripción matemática de determinados aspectos del pensamiento ha brindado al hombre la oportunidad de transferir a la máquina una inmensa cantidad de rutina acumulada en su actividad cotidiana. La rutina está ligada precisamente a las formas de trabajo intelectual que el hombre efectúa en virtud de reglas formales establecidas con precisión. Tales reglas admiten una algoritmización, es decir una reducción a procedimientos unitarios basados en una prescripción determinada que no comporte ambigüedad. Justamente esta algoritmización constituye el modo por el cual el hombre se dirige a la máquina. Dicho de otra manera: toda actividad humana es transferible a la máquina en la medida en que sea reductible a un algoritmo. Asignando a la máquina la mayor parte posible de su actividad, el hombre puede consagrarse cada vez más a la labor creadora, es decir a la actividad situada en el polo opuesto de la rutina.

El hecho de que ciertas ciencias se matematicen más tarde que otras se explica por razones de índole metodológica e histórica. Una ciencia como la biología estudia fenómenos más complejos que los fenómenos de que se ocupa la física. Resulta, así, que en un fenómeno físico la separación entre los aspectos cuantitativos y los aspectos espaciales es más factible que en un fenómeno biológico. Al mismo tiempo, esta separación requiere una cierta madurez en la ciencia respectiva, madurez que varía en función de la complejidad de los fenómenos estudiados. El estudio de los aspectos cuantitativos y formales es global, de conjunto, y exige un buen conocimiento previo acerca de las cosas particulares, del acervo de hechos que contiene la ciencia en cuestión.

La separación de los aspectos cuantitativos y formales no tiene en cuenta la naturaleza de las cosas; se hace únicamente en función de las necesidades indagatorias.

El proceso de matematización de las ciencias forma parte del proceso más general por el cual se intersecan progresivamente los distintos dominios de in-

vestigación. La lingüística, p. ej., confluye cada vez más no sólo con la matemática, sino también con la física, la psicología, la lógica, la técnica y otras.

La lingüística matemática acusa, a su vez, un proceso de síntesis, análogo al que se produjo en la aparición de la cibernética. Las siguientes observaciones del académico A. I. Berg, relativas a la cibernética, también son aplicables esencialmente a la lingüística matemática, con la diferencia, en todo caso, de que la síntesis desarrollada por esta última se halla inmediatamente ligada a la problemática de una ciencia social, la lingüística:

“Sin duda, la cibernética se ha constituido como ciencia que ha realizado la síntesis de una serie entera de disciplinas científicas. La aparición de nuevas ciencias en las zonas de contacto de distintas ramas del conocimiento es, tal vez, una de las particularidades más características del desarrollo científico de nuestros días. Así han aparecido la biofísica, la bioquímica, la química física, la radioquímica y muchas otras disciplinas que reúnen, en cierto sentido, los objetos de ciencias diferentes. Por lo que parece, la cibernética no es una excepción a este respecto. Lo significativo consiste en el hecho de que la cibernética haya aparecido en el punto de reunión de ciencias en apariencia bastante distintas: la matemática, la disciplinas de ingeniería y las biológicas. Evidentemente, el carácter de esta síntesis debe despertar un gran interés para la gnoseología del materialismo dialéctico.”⁸

2.15. Premisas lingüísticas de la lingüística matemática

La lingüística ha vivido mucho tiempo como ciencia descriptiva, de observación y comparación de hechos concretos y particulares. Pero los hechos se han acumulado más y más, ciertos tipos de relaciones y modos organizativos se han vuelto cada vez más patentes, hasta que se ha impuesto la necesidad de estudiarlos. Estas relaciones y modos de organización han aparecido, pues, sobre la base de un inmenso material de datos acumulado tanto en la lengua como en la lingüística. Así, también en lingüística han recibido un perfil cada vez más preciso y una importancia más y más grande los aspectos globales, de conjunto, los aspectos estructurales referentes a la forma de las relaciones que contraen mutuamente los elementos de la lengua. El *Curso de lingüística general* de Ferdinand de Saussure expone los fundamentos para una investigación de esta naturaleza. Las ideas de Saussure, la teoría de la formalización fonológica propuesta por Trubetskoi y algunos resultados de Jakobson —como, p. ej., sobre la posibilidad de expresar cualquier estructura fonológica a base de oposiciones

8. Cf. BERG, “Unele probleme”.

binarias— han preparado el terreno para la introducción de los métodos matemáticos en la lingüística.

Es verdad que afirmaciones como la que hizo F. de Saussure en el sentido de que “en la lengua no hay más que diferencias” no se conforman a la realidad. Los mismos que sostienen este tipo de afirmaciones no siempre son consecuentes con ellas, sobre todo si no se dan cuenta de ello. Al menos alguna vez, no estudian cualquier tipo de relaciones y diferencias, sino sólo las que aparecen con una base significativa, es decir a partir de los elementos mismos entre los que se ejercen las relaciones. Como resultado de esta situación, puede comprobarse hoy que las ideas y nociones introducidas por F. de Saussure se reencuentran en la teoría matemática de los códigos. Con el mismo valor que tiene en Saussure el concepto descriptivo de “oposición” se introduce aquí la noción cuantitativa, métrica, de “distancia” (cf. 10.5). Las analogías de este tipo llegan muy lejos. También se encuentran, con ayuda de esta distancia, las oposiciones asociativas y sintagmáticas. La regularidad de la lengua, su carácter sistemático, tal como lo concebía Saussure, corresponde perfectamente a la regularidad que hallamos en un código, la cual resume aquí el compromiso de coherencia las necesidades de codificación con las de descodificación.

La capacidad correctora de un código puede ponerse en relación con ciertas formas particulares de oposiciones lingüísticas. Los conceptos fundamentales de la teoría de Trubetskoï sobre las oposiciones, como, p. ej., el de correlación, se hallan aquí perfectamente ilustrados. Desde el punto de vista de la teoría de los códigos, surge igualmente una jerarquía bien determinada entre los diversos tipos de oposición. Las oposiciones proporcionales son preferibles a las aisladas, y las bilaterales a las multilaterales. En cuanto al problema de aumentar la eficacia de los códigos correctores reencontramos también muchas de las nociones de Trubetskoï.

Ni siquiera los conceptos básicos con que opera Hjelmslev son ajenos a los problemas de la teoría de los códigos. Por lo demás, las mismas clasificaciones que propone Hjelmslev se hallan muy cerca de las que propuso Trubetskoï. El concepto de “extensión”, establecido por Hjelmslev, está íntimamente relacionado con la llamada distancia distribucional y ésta, a su vez, tiene un parentesco lejano con la “distancia” de Hamming, bien conocida en la teoría de los códigos.

En fin, la posibilidad de expresar toda estructura fonológica en términos binarios tal como lo ha demostrado Jakobson y el hecho, indicado por él mismo, de que el número total de rasgos pertinentes fonemáticos es menor o igual a 12 en lenguas diversas encuentran un claro reflejo en el carácter binario de la mayoría de códigos utilizados y en la circunstancia de haber podido transportar al terreno de los fonemas algunas leyes estadísticas válidas para las palabras.

2.16. La importancia del análisis lingüístico estructural y sus relaciones con el análisis de la sustancia lingüística

Sin duda, el estudio evolutivo de la lengua y del modo como se verifican los cambios constituye uno de los cometidos básicos de la lingüística. Pero reducir la lingüística sólo a esto significaría angostar considerablemente el objeto que le compete de un modo natural. Una de las cuestiones fundamentales de la lingüística consiste en el estudio de las leyes que se hallan en la base de los distintos modos de uso lingüístico en la sociedad. Las necesidades sociales han llevado al primer plano el problema de automatizar los distintos procesos comunicativos que se sirven de la lengua: la información, la lectura y la traducción automáticas. Por consiguiente, la lingüística, si no quiere desprenderse de las necesidades sociales y convertirse en una disciplina anacrónica, debe incluir en su objeto de indagación los problemas lingüísticos que derivan de la automatización de los procesos lingüísticos. Y el principal de estos problemas, tal como han demostrado las investigaciones de los últimos años (N. Chomsky, T. Moloshnaia, I. A. Melchuk, T. Nikolaeva, etc.), se refiere a la descripción lingüística a partir de la posición recíproca de los distintos elementos, a partir de las relaciones, paradigmáticas o sintagmáticas, entabladas entre los elementos lingüísticos.

Se ha comprobado claramente que el problema de automatizar los procesos comunicativos establecidos por medio de una lengua no es, en primer lugar, de índole técnica, sino lingüística.

Revelar de un modo profundo la red de relaciones: he aquí una cuestión muy actual de la lingüística. En este sentido, reciben un interés decisivo los procesos de abstracción, gracias a los cuales, en determinadas etapas de la investigación, se omiten o se realzan algunas facetas del objeto experimentado no a causa de su naturaleza propia, sino de las exigencias de la investigación misma (del mismo modo, cuando un fisiólogo destaca en primer plano el estudio de las funciones del hígado, pongamos por caso, no significa que niegue la importancia de la estructura anatómica del mismo). La teoría leninista de la reflexión nos enseña que los procesos de abstracción por los cuales se extraen las propiedades generales no sólo no nos alejan de la realidad, sino que, por el contrario, nos acercan más a lo esencial y necesario.

La lingüística debe responder actualmente a la pregunta siguiente: ¿cuáles son los fenómenos lingüísticos susceptibles de ser descritos mediante recursos formales, posicionales y exactos, y en qué consiste esta descripción? En tal sentido, la lingüística elabora modelos lógicos, construcciones teóricas deductivas, que, justamente por su carácter unilateral, es decir por el hecho de que ignoran determinados aspectos de los fenómenos, revelan más profundamente la estructura de los demás aspectos. La función gnoseológica de la lingüística se manifiesta hoy, en buena parte, mediante modelos lógicos. Tal como han demostrado

las experiencias de los últimos años (Chomsky, Kulaguina, Revzin, etc.) el modelo lógico sirve no sólo para expresar de otra manera, accesible a la máquina, la estructura ya conocida de la lengua, sino para descubrir ciertos matices nuevos y más profundos de esta misma estructura. Ello no implica que las investigaciones sobre la sustancia sonora o semántica de la lengua sean inútiles. De hecho, este tipo de investigaciones se revalorizan hasta el punto de resultar esenciales, puesto que cuanto más se conoce la sustancia sonora y semántica de la lengua tanto mejor pueden aislarse los aspectos relacionales.

Por lo demás, no es posible indicar con precisión dónde termina la sustancia y dónde empiezan las relaciones. Un elemento sustancial, un sonido, p. ej., revela, en un análisis atento, un complejo de relaciones; y, por su parte, una relación tampoco existe en estado puro, sino más bien al contrario, se comprende mejor y adquiere nuevos atributos cuando se detalla la naturaleza de los miembros que la determinan. No tiene sentido, pues, separar del campo teórico de la lingüística la sustancia sonora o semántica, pero su estudio exclusivo es, a su vez, insuficiente, ya que para penetrar en la esencia de los fenómenos lingüísticos conviene estudiar también las relaciones. La descripción semántica de los casos gramaticales no ha sido infructuosa, sino que, por el contrario, ha constituido la base intuitiva a partir de la cual se ha podido elaborar otra descripción en el marco de un conjunto de rasgos diferenciales. En determinadas cuestiones y circunstancias, es precisamente esta última descripción la más necesaria gracias a que penetra más profundamente en la estructura del sistema casual⁹.

Una descripción sustancialista, como, p. ej., la de las partes del discurso, puede resultar no lingüística, en el sentido de que se vuelve inutilizable y contradictoria ante una serie de problemas fundamentales que plantea la lingüística misma.

Los procesos de abstracción pueden acercarnos, pues, al conocimiento de la realidad objetiva, pero pueden también apartarnos de ella si no se manipulan en una dirección eficaz y fecunda. ¿Cuál es, entonces, el criterio para verificar la autenticidad de los conocimientos? Como se sabe, el materialismo dialéctico proporciona una respuesta precisa a esta pregunta: la única verificación posible sobre la verdad contenida en los conocimientos logrados depende de su confrontación con la práctica. Ahora bien, debemos entender la práctica en sentido amplio si no queremos caer en perversiones neopositivistas. La práctica lingüística se registra en el conjunto de usos sociales de la lengua. El análisis estructural de la lengua se ha mostrado esencial en la elaboración de lenguas de información, en la composición de algoritmos de traducción automática y en la codificación lingüística destinada a su transmisión por diversas líneas comunicativas. Y a su vez, estas actividades estimulan el análisis estructural de la lengua al

9. Cf. PADUCHEVA, "Descrierea".

orientarlo por vías fecundas. El análisis estructural ha obtenido así grandes éxitos. Ahora ya no se trata de “escuelas estructuralistas”, sino de una tendencia general de la lingüística. Las relaciones que el análisis estructural pone de manifiesto constituyen el reflejo, en el pensamiento humano, de algunas relaciones que existen en la realidad objetiva de la lengua. En esto reside la clave de los éxitos y de las perspectivas del análisis lingüístico estructural.

2.17. La lingüística matemática como etapa superior en el desarrollo de la lingüística estructural

La lingüística estructural ha conseguido poco, a pesar de todo, en lo que se refiere a la explicación de las nociones fundamentales. Una situación característica a este respecto la constituye el problema del fonema. Conceptos como los de “neutralización” o “archifonema” no han recibido aún una claridad y una precisión suficientes para manejarlos de un modo sólido. En este punto, como en el del sintagma, del morfema, o de las estructuras lingüísticas fundamentales, la lingüística estructural ha sufrido ciertas deficiencias, falta de perspectiva y ausencia de un método evolucionado. En principio, la lingüística estructural y la lingüística matemática se ocupan, en buena parte, del mismo objeto: la descripción de los aspectos formales, relacionales, estructurales, de los fenómenos de la lengua. (La lingüística matemática estudia, además, los aspectos estadísticos.) Pero mientras la lingüística estructural no dispone todavía de un método específico y adecuado a este propósito, la lingüística matemática tiene a su disposición el conjunto de métodos elaborados por una de las ciencias más antiguas, los métodos excepcionalmente perfeccionados de la lógica matemática, de la teoría de los conjuntos, del cálculo de probabilidades, del análisis matemático, del álgebra, de la topología. En segundo lugar, aunque como derivación del primer factor, la lingüística estructural se halla encorsetada en los esquemas de la lógica aristotélica y no se plantea el cometido de resolver ningún problema concreto, práctico, en tanto que se halla, en todo caso, desposeída de contactos directos con la realidad; por su parte, la lingüística matemática tiene ante sí la perspectiva vasta y precisa de resolver algunos de los problemas más agudos que actualmente plantean la práctica y la técnica. el problema de aumentar la capacidad de los canales de transmisión, de leer automáticamente, de construir aparatos estenográficos, de recopilar y conservar mecánicamente la información, de traducir electrónicamente, de reseñar automáticamente los trabajos científicos y técnicos, de transformar lenguas orales en escritas, etc.

Así, pues, la lingüística matemática, adoptando y valorificando algunas nociones, esquemas y sugerencias de la lingüística estructural, enriquece el objeto estudiando, además, los aspectos estadísticos, establece un vínculo estrecho con

las exigencias prácticas y técnicas y utiliza el conjunto de métodos elaborados por la matemática clásica y moderna.

2.18. La lingüística estructural no matemática como premisa necesaria para el desarrollo de la lingüística matemática

Ante esta situación, vale la pena hacer un balance sobre las realizaciones de la lingüística estructural y precisar sus perspectivas de progreso. ¿Ha allegado resultados efectivamente nuevos, desconocidos para la lingüística tradicional, o acaso se ha limitado a proveer de nueva formulación datos ya conocidos? Desde luego que en una ciencia lingüística como ésta es difícil delimitar lo que constituye un dato o un hecho nuevo de una organización, una estructuración o una formulación nueva sobre unos hechos conocidos. Pero aun cuando sea cierta la segunda de estas posibilidades, habremos de admitir que la lingüística estructural no carece de transcendencia. Incluso aunque la lingüística estructural no haya hecho más que preparar los primeros elementos y esquemas de la lengua por los que el hombre accede a comunicarse con la máquina, su importancia resulta fuera de toda duda. Al margen de ello, conviene observar que el valor de algunos resultados referentes a la organización y a la sistematización de un inmenso material de datos termina por comprenderse, por lo general, más difícil y tardíamente, se abre camino con mayor esfuerzo y el aprovechamiento de algunos de estos resultados se demora más todavía. La comprensión de estos mismos resultados exige la comprensión previa de la nueva problemática que plantean los aspectos globales de la lengua. Muchos malentendidos provienen del hecho de que, a menudo, se aprecien los resultados de la lingüística estructural con relación a la problemática tradicional, de la lingüística como ciencia descriptiva y de los hechos particulares.

¿Qué dirección tomará el futuro desarrollo de la lingüística estructural no matemática? ¿Acaso lo único que podría mantenerse es el hecho, incontestable por el momento, de que estas investigaciones están en manos de hombres dotados de una formación lingüística superior a la formación de los que se ocupan de la lingüística matemática? Si así fuese, como esta situación no durará mucho, sucedería que la lingüística estructural y la lingüística matemática llegarían a confluir en una sola línea de investigación.

Un fenómeno característico de la actual situación de la lingüística estructural viene representado por las investigaciones en torno al fonema. Las teorías estructuralistas sobre el fonema se hallan actualmente en plena renovación. Shau-mian ha demostrado la falsedad de la afirmación según la cual la relación entre los conceptos de “sonido” y de “fonema” sería de particular a general. Kuznetsov y Revzin, aplicando la simbolización binaria de los matemáticos, han aduci-

do una definición formalizada sobre el fonema partiendo de la concepción en virtud de la cual el fonema constituye un conjunto finito de rasgos fonológicos pertinentes. También en la descripción del fonema, V. Belevitch ha introducido el instrumento más especializado de las funciones booleanas. En fin, B. Mandelbrot ha desarrollado una amplia analogía entre algunos fonemas de la lengua y algunos fenómenos físicos, concibiendo el análisis subfonemático como un análisis del átomo en núcleo y electrones. Naturalmente, una analogía de este tipo puede parecer una simple alegoría. Sin embargo, en principio, la idea de que los hechos lingüísticos adoptan una existencia *sui generis* e incomparable a ninguna otra cosa, debido a su complejidad, está equivocada. La analogía representa uno de los más poderosos métodos de investigación y una expresión luminosa de la audacia creativa. La analogía no sustrae nada de lo que es específico en los hechos; sólo destaca lo que es común a unos hechos que, a primera vista, parecen mutuamente adventicios. Evidentemente, las analogías deben efectuarse con prudencia. Pero, en suma, la analogía manifiesta en muy alto grado la unidad material del mundo y de los distintos tipos de movimiento.

Un ejemplo sencillo y significativo para la valoración de los métodos matemáticos en lingüística aparece en la clasificación de las oposiciones en privativas, graduales y equipolentes, debida a Trubetskoï. J. Cantineau ha criticado con fundamento esta clasificación por su carácter incompleto, pero ni él mismo observó una deficiencia esencial de la clasificación, deficiencia que surge tan pronto como intentamos expresar estos tipos opositivos con ayuda de las operaciones y relaciones de la teoría de los conjuntos. Procediendo de esta manera, nos damos cuenta inmediatamente de que, mientras las oposiciones privativas se expresan a través de relaciones de inclusión, las graduales lo hacen a través de relaciones de orden. Esta clasificación de Trubetskoï no se apoya, por tanto, en un punto de vista bien determinado ya que mezcla el criterio del orden con el de la inclusión¹⁰.

2.19. La estructura probabilística de la lengua

Otro aspecto que conforma el objeto de investigación de la lingüística matemática está vinculado con la estructura probabilística de la lengua. Hace ya mucho tiempo que se ha observado que la lengua se presenta como un conjunto estadístico. En la lengua actúan numerosos y muy variados factores, que no siempre permiten una evaluación por la que se pueda emprender un análisis causal de los fenómenos. Pero más allá de esta apariencia desordenada, el hombre de ciencia se siente atraído por la estabilidad frecuencial de determinados hechos lingüísticos, la cual tiene origen en ciertas propiedades objetivas, de estructura.

10. Cf. APOSTEL-MANDELBROT-MORF, *Logique*.

Aquí se abre, sobre todo para el lingüista, un amplio campo de experimentación. Pero ningún aspecto matemático ofrece menos interés. El matemático soviético A. I. Jinchin elaboró, hace unos 25 años, el llamado “método de las funciones arbitrarias”, por el cual se prevén tanto la estabilidad frecuencial de ciertos datos como el valor de esta frecuencia.

Jinchin expone su metodología del siguiente modo: “Partiendo de una distribución arbitraria de los datos iniciales, se demuestra que la estabilidad y el valor numérico de la frecuencia de un determinado evento en una serie grande de experimentos pueden ser establecidos a partir de las particularidades objetivas del fenómeno mismo y que el valor de la frecuencia no depende de la distribución arbitraria inicial”. A causa de las dificultades matemáticas que contiene, este método sólo se ha aplicado hasta ahora a ciertos fenómenos de naturaleza mecánica. En principio, sin embargo, tal como indica el propio Jinchin, el método es capaz de tratar cualquier fenómeno que se desarrolle según ciertas leyes.

El análisis macroscópico de los textos lingüísticos ha puesto en evidencia determinados esquemas probabilísticos, que se han convertido en objeto de experimentación para ciertas teorías matemáticas especiales. Un esquema de esta índole fue descubierto, hace unos 60 años, por el matemático ruso A. A. Markov. La sugerencia provino, al parecer, de un texto lingüístico que casualmente correspondía al poema *Evgeni Onegin* de Pushkin. Cada letra tiene una cierta probabilidad de aparecer y una cierta probabilidad de ir seguida de otra letra dada. Así, pues, el fenómeno que denominamos “lengua” se presenta como una sucesión de estados $s_1, s_2, \dots, s_i, s_{i+1}, \dots, s_n$, en la cual existe una probabilidad determinada de que después del estado s_i siga el estado s_j . Esta clase de esquema se conoce, en la teoría de las probabilidades, por el nombre de “cadena de Markov”.

Fácilmente se comprende que el esquema de la cadena de Markov no registra el hecho de que, a veces, la aparición de una letra no depende sólo de la letra precedente, sino de las letras anteriores a la precedente (cf. 8.10). Los matemáticos rumanos O. Onicescu y G. Mihoc han presentado y estudiado cadenas que acusan esta incidencia, llamadas “cadenas con enlaces completos”.

Desde hace mucho tiempo se hallan ejemplos intuitivos sobre la estabilidad de la frecuencia lingüística. Edgar A. Poe dio pruebas de una intuición perfecta sobre la estabilidad frecuencial de las letras en textos de lengua inglesa cuando, en el cuento *El escarabajo de oro*, hace que Legrand descifre el mensaje del capitán Kidd.

Para la descripción de la estructura probabilística de la lengua ha sido necesario recurrir a algunas nociones de la física, en primer lugar a la entropía y a la ergodicidad (cf. 8.9). Ello demuestra una vez más la importancia de la analogía como método de investigación científica. El carácter ergódico de los procesos lingüísticos se refiere a que, en textos relativamente largos, las frecuen-

cias se estabilizan y se vuelven sensiblemente iguales. Así, p. ej., en dos textos de cinco líneas cada uno, la frecuencia relativa de una determinada letra puede resultar bastante diferente, pero en dos textos de 50 páginas que pertenezcan a la misma lengua, la frecuencia relativa de una letra determinada será aproximadamente igual en ambos casos.

La entropía mide la indeterminación que emana de un mensaje lingüístico. En una lengua cuyas letras tuvieran la misma probabilidad de aparición, es decir cuyos textos fuesen homogéneos desde el punto de vista de la probabilidad de sus elementos, la indeterminación que produciría la aparición de una nueva letra, o sea la entropía, sería mayor que en una lengua formada por letras con probabilidad diferente. Cuanto más semejante es la probabilidad de las letras, tanto mayor es el desorden, puesto que con mayor dificultad cabe prever la letra siguiente. Así, pues, la entropía consiste, en cierto modo, en una medida del desorden. A través del concepto de entropía, la estructura probabilística de la lengua se encuentra estrechamente vinculada a los axiomas y resultados de la teoría de la información.

C. E. Shannon, uno de los fundadores de la teoría de la información, se ocupó de la entropía lingüística. Indicó al propio tiempo algunos experimentos que permiten determinar la entropía de una lengua dada (cf. 8.11). También en Rumanía se han realizado experiencias de este tipo sobre la lengua rumana debidas a Edmond Nicolau (cf. cap. 8). Por otra parte, hemos de subrayar que la lingüística constituye un importante auxiliar para los investigadores que trabajan en el dominio de la teoría de la información, hecho especialmente natural si tenemos en cuenta que la lengua es el principal medio de comunicación entre los hombres.

2.20. La lingüística estadística

Uno de los campos de la lingüística matemática, el más cultivado por los lingüistas, es el de la lingüística estadística. Existe ya una abundante bibliografía sobre esta actividad, que tanta trascendencia tiene para la confección de diccionarios de frecuencia y para la determinación del fondo principal de la lengua. Recientemente se habla con una insistencia creciente sobre las particularidades estadísticas de los estilos, de cuyo acopio se esperan nuevos procedimientos para la crítica literaria. La resolución de este tipo de problemas presupone en apariencia muy pocos conocimientos y hábitos matemáticos y ofrece un campo de acción y de interpretación muy amplio. Esta engañosa comodidad explica —en buena parte— la concentración, en la especialidad, de un número inmenso de cultivadores. Sin embargo, el riesgo del empirismo estéril, no iluminado por la teoría, no es en ninguna parte mayor que aquí. Y puede que en ninguna parte

más que en ésta se manifieste la ignorancia del significado real de las cosas. Las múltiples oscilaciones que van desde la generalización absoluta de los métodos estadísticos hasta su repulsa total marcan las dos actitudes características de aquellos que no comprenden el significado de los métodos matemáticos en el estudio de la lengua. Algunos rechazan incluso la idea misma de su utilización, mientras otros tienden a ver en una fórmula matemática un horóscopo, una solución milagrosa que invalidará o subordinará cualquier otra investigación lingüística hecha con recursos no matemáticos. Estos últimos ven las Facultades de Filología como futuras “sucursales” de las Facultades de Matemática y Física.

¿Qué significado real cabe atribuir a la estadística lingüística? Ésta no puede —en ningún caso— proporcionar una indagación exhaustiva sobre los fenómenos lingüísticos. Tampoco es, al menos de momento, tan significativa, como lo pretende P. Guiraud, ni tan trivial, como lo sostiene G. Matoré.

La estadística lingüística no es más que un auxiliar, y sus resultados deben sujetarse a un punto de vista omnicompreensivo. No debemos olvidar ni un momento que la separación de los aspectos cuantitativos se hacen sólo en función de las necesidades metodológicas. Es absolutamente necesario, pues, que, al cabo de esta fase analítica por la que se separan los distintos aspectos, se realice la operación de síntesis. Aquí se halla una de las deficiencias que acusan ciertos usos del método estadístico en lingüística. Muchas veces se extraen interpretaciones abusivas, sin una derivación rigurosa y necesaria de los datos aducidos. Otra insuficiencia, de naturaleza más profunda y probablemente ineludible por ahora, proviene del hecho de que las distintas fórmulas estadísticas —como, p. ej., las de Estoup, Zipf y Mandelbrot— tengan su origen en las propiedades objetivas de la estructura lingüística (cf. 9.3). A veces se pasa por alto que estas fórmulas no siempre ofrecen un carácter estrictamente empírico, sino que se infieren matemáticamente de algunos rasgos de los fenómenos lingüísticos. Ello no impide que estos rasgos tengan un especial interés teórico, ya que constituyen el punto de partida cuando se emplean estas fórmulas. Los experimentos, a su vez, han de actuar con arreglo a las condiciones impuestas por las fórmulas y, si no ocurre así, hay que fijar con exactitud las diferencias que concurren para comprobar eventualmente, por vía empírica, la validez de una fórmula en condiciones más generales o de otro tipo respecto de aquellas en que se estableció, homologando, si éste es el caso, las nuevas condiciones como una investigación teórica más.

El método estadístico resulta un auxiliar precioso para revelar el carácter sistemático de la lengua, puesto que permite determinar ciertas propiedades de regularidad con carácter global e inasequibles por otros métodos. La estadística reporta, además, en lo que concierne a algunos hechos conocidos, una suerte de precisión especialmente saludable.

No todos los aspectos cuantitativos admiten un estudio estadístico. Y, por otra parte, algunas leyes cuantitativas de carácter estadístico parecen permitir el establecimiento de ciertos nexos de relación con posibilidades muy fructíferas. En este sentido, si se comprobara que la extensión del artículo consagrado a una palabra dada del diccionario está relacionada de un modo determinado, matemáticamente expresables, con el número de acepciones que tiene aquella palabra de igual forma que en la hipótesis, empíricamente verificada, según la cual hay una relación matemática entre el número de acepciones de una palabra y la frecuencia de ésta, resultaría entonces que la extensión de un artículo del diccionario estaría también relacionada a través de una ley matemáticamente expresable con la frecuencia de la palabra respectiva en la lengua¹¹.

Conviene notar asimismo que algunas cuestiones aparentemente indescribibles con la ayuda de la frecuencia podrían resultar en definitiva problemas justamente de frecuencia. Así, las palabras disponibles que algunos oponen a las palabras frecuentes acusan, la mayoría de las veces, una frecuencia alta en cierto sector de la lengua¹². De ahí la necesidad de distinguir distintos sectores en el marco de la lengua, en relación con los cuales deben tratarse los problemas de frecuencia, homonimia y otros. Al menos para algunas necesidades, como en la traducción electrónica, este punto de vista es el más indicado.

También como las demás ramas de la lingüística matemática, las investigaciones de lingüística estadística pueden entrañar profundas repercusiones, capaces de afectar a aspectos no cuantitativos de la lengua. El simple hecho de numerar las palabras de un texto obliga a elegir una definición sobre la palabra. Esta definición debe evitar toda ambigüedad, aun con el riesgo de sacrificar ciertas verdades lingüísticas, siempre y cuando la parte sacrificada resulte suprimible a medida que los textos tomados en consideración se hagan suficientemente largos. Este tipo de situaciones sugiere, sin duda, que en una “macro-lingüística” —es decir, en un estudio lingüístico desde el punto de vista global— se recomiende utilizar definiciones que sacrifiquen aspectos particulares, aunque éstos ofrezcan el mayor interés, con la condición de que lo sacrificado sea tan sólo un fenómeno esporádico y que la definición resultante contenga una precisión matemática.

Pueden tratarse muchas otras cuestiones interesantes con relación a la estadística lingüística. Las más delicadas se refieren a los caracteres estadísticos de los estilos. Fenómenos tales como los descubiertos por Yule y Herdan, según los cuales ciertas expresiones matemáticas relativas a un texto lingüístico gravi-

11. [Cf. nota 4].

12. Las “palabras disponibles” aluden a las que, sin tener una frecuencia elevada en el habla común, son, no obstante, insustituibles en determinadas situaciones, como, p. ej., *termómetro, forro, alfil*, etc.

tan en torno a unos valores determinados, ocultan una realidad profunda, sobre la cual deberán reparar tanto lingüistas como matemáticos¹³.

Como hemos visto, la creación de algoritmos para la traducción automática, de diccionarios electrónicos, ha imprimido un nuevo impulso a las investigaciones de estadística lingüística. El diccionario electrónico sólo contiene algunos miles de palabras y la elección juiciosa de estas palabras implica la formación previa de un diccionario de frecuencias donde las palabras se incorporen en orden decreciente a partir de su frecuencia en la lengua. Conviene, al mismo tiempo, conocer la frecuencia de las distintas formas gramaticales y de los distintos tipos de configuraciones sintácticas (cf. 8.2). Sólo así se podrá discernir entre fenómenos sistemáticos y fenómenos esporádicos, omitibles, en la composición del algoritmo de traducción, es decir en la composición de las series operacionales que debe efectuar la máquina con vistas a la traducción.

La confección de diccionarios de frecuencia y el componente gramatical de los algoritmos de traducción automática se han mostrado especialmente eficaces para el perfeccionamiento pedagógico en la enseñanza de las lenguas. Para dar sólo una idea a este respecto, mencionaremos que las 100 primeras palabras del diccionario de frecuencias de una lengua cubren la mitad de cualquier texto escrito en la lengua respectiva.

La construcción de un diccionario de frecuencias de la lengua rumana es un problema sólo parcialmente resuelto. La obra de A. Juilland, P. M. H. Edwards e I. Juilland, muy estimable desde el punto de vista de la metodología estadística, resulta, sin embargo, insuficiente por lo que respecta a la selección de las muestras¹⁴. De todos modos, habrá que tenerlo en cuenta para toda investigación ulterior. Hay que mencionar, asimismo, la labor anterior de Al. Graur para el establecimiento del fondo léxico principal en rumano, basado en la importancia de las palabras en la lengua¹⁵. La importancia de una palabra puede medirse, con una aproximación relativamente buena, a partir de la extensión del artículo que le consagra el diccionario. Ahora bien, dicha extensión depende, en primer lugar, del número de acepciones de la palabra respectiva: cuanto más numerosas son las acepciones, tanto mayor es el artículo correspondiente en el diccionario. Y el número de acepciones depende, a su vez, por una ley estadística bien conocida, de la frecuencia de la palabra en la lengua: el número de acepciones de una palabra es directamente proporcional a la raíz cuadrada de su frecuencia. Entonces, si el fondo léxico principal deducido por la extensión del artículo del diccionario equivale aproximadamente al fondo establecido a partir de la frecuencia (según resulta, para el rumano, de un trabajo efectuado por V. Şuteu

13. [Cf. YULE, *Statistical Study*; HERDAN, *Language, Type-token Mathematics y Calculus*].

14. Cf. JUILLAND, EDWARDS y JUILLAND, *Frequency Dictionary*.

15. Cf. GRAUR, *Încercare*.

(cf. 8.4)), ello puede utilizarse como un indicio de que la distinción de acepciones en el diccionario y la extensión explicativa que se le atribuye obedecen a la aplicación de unos criterios homogéneos. Así, pues, la evaluación de la frecuencia de las palabras ayuda a establecer el valor objetivo de unos hechos, problemáticos de otro modo.

La estadística lingüística constituye una fase preliminar para las investigaciones de lingüística matemática. Sin ella apenas puede emprenderse nada en este dominio, pero, al propio tiempo, ella sola ofrece una imagen absolutamente incompleta sobre la lingüística matemática. Es necesario dirigir las indagaciones estadísticas sobre textos científicos en primer lugar, debido a que la lengua de estos textos está gobernada por leyes puramente lingüísticas. El tratamiento mecánico de la lengua, por lo demás, se limita, al menos por el momento, a este tipo de textos.

La aparición de la lingüística matemática requiere la colaboración de lingüistas, ingenieros y matemáticos, pero las dificultades para recabar una tal colaboración son evidentes. Así, la colaboración entre lingüistas y matemáticos exige que el matemático se familiarice con las estructuras lingüísticas fundamentales (morfológicas, sintácticas, fonológicas) y con determinados hechos de lengua; y el lingüista, con la perspectiva típica del matemático y con ciertas nociones matemáticas (probabilidad, entropía, principios matemáticos de los ordenadores electrónicos, conceptos y operaciones de la teoría de los conjuntos). Conviene subrayar que la lingüística y la matemática intervienen en la lingüística matemática no sólo por sus hechos elementales, de principio, sino también por cuestiones que incumben al desarrollo moderno de ambas ciencias. Posiblemente los problemas de modelación de la lengua sugieran al matemático el estudio de unas estructuras nuevas por las mismas razones por las que la lingüística se verá probablemente estimulada a introducir una óptica diferente en diversos problemas sobre los cuales el conservadurismo y la tradición ejercen una fuerza especialmente grande (p. ej., en la gramática).

Respecto a la colaboración entre lingüistas y matemáticos, el lógico Haskell B. Curry relata el siguiente símil con un elefante y unos ciegos. Cada uno de los invidentes se acercó a una parte distinta del cuerpo del elefante. El que se colocó junto a una pata tuvo la pretensión de que el elefante es una columna; el que tocó la cola aseguraba que se trata de una soga; el que lo palpó lateralmente lo imaginó como un muro; el que se acercó a la oreja supuso que se hallaba junto a un velero y el que anduvo por la trompa aseguraba estar al lado de una manguera. Según Curry, la gramática general semeja también un elefante junto al cual el lingüista, el psicólogo, el lógico, el matemático pasan por un lugar distinto y lo interpretan a su manera. Curry se propone divulgar las impresiones de un invidente situado, respecto del elefante gramatical, en un lugar poco acostumbrado. Opina que el progreso en el desarrollo de la gramática sólo es posible

mediante una cooperación audaz entre los mencionados ciegos, una colaboración en la que cada uno trate de decir —sin temor al yerro o al apuro y con el cuidado de hacerse comprender— todo lo que piensa acerca del elefante.

No puede negarse que, en lo referente al acercamiento de los lingüistas a la matemática, aún existen ciertos inconvenientes tanto objetivos como subjetivos y, a la postre, semejantes a los que aparecen, p. ej., en el acercamiento a la matemática por parte de biólogos y médicos. A este propósito son significativas para algunos lingüistas las siguientes palabras del matemático polaco H. Steinhaus escritas sobre las reticencias de los médicos ante los procedimientos matemáticos: “Muchos investigadores de la naturaleza, y médicos en mayor número todavía, no creen en la posibilidad de aplicar la matemática a la naturaleza viva. Ello se debe a que en las lecciones escolares de matemática se les ha inculcado una idea preconcebida sobre el carácter abstracto de los conceptos, sobre el formalismo de sus definiciones y sobre el carácter artificial de unos principios que recuerdan más bien charadas y jeroglíficos; la simpleza caricaturesca de las concepciones geométricas vistas por el ojo del matemático, por una parte, y la complejidad y la fluidez del mundo vivo, los árboles eternamente verdes, con frutos de oro, de la naturaleza, como decía Goethe, gran adversario de la matemática, por otra parte, determinan la antítesis que compromete a los ojos de los naturalistas la idea misma de un tratamiento matemático de la naturaleza viva como una idea ingenua y perniciosa por su simplista pretensión. Ellos se imaginan que los matemáticos persiguen una fórmula omnívota, que pretenden calcular horóscopos con anticipación igual que los clérigos astrólogos de la Edad Media y que luego querrán prever cada paso y cada oscilación del organismo vivo”¹⁶.

La lingüística matemática se halla todavía en los comienzos. De momento, debe apreciarse por sus perspectivas y sólo después por sus logros. Es difícil adivinar cuál su camino principal y en qué relaciones entrará, a lo largo de su evolución ulterior, con la lingüística, con la matemática, con la lógica, con la técnica y con los demás dominios en que radica. Algunos, como Shaumian, se hallan más bien dispuestos a creer que la lingüística matemática progresará, al principio, como lingüística abstracta, axiomática, puramente deductiva; otros, como Toporov, preludian una filiación a las concepciones probabilísticas y estadísticas. En todo caso, la lingüística matemática no podrá agotar el fenómeno lingüístico, pero será particularmente eficaz en el sector de los aspectos cuantitativos y formales. El criterio de apreciación de la lingüística matemática lo constituirán, en primer lugar, las necesidades prácticas a que atienda.

16. Cf. STEINHAUS, “The collaboration”.

2.21. Anexo bibliográfico

Durante los diez años que han transcurrido desde la edición rumana de este libro, la lingüística matemática se ha desarrollado vertiginosamente produciendo una bibliografía que incluye millares de títulos. Para una mejor orientación general del lector, indicaremos ahora algunas de las monografías y revistas más significativas con cuya ayuda puede identificarse la mayoría de los trabajos de especialidad en este dominio.

En el terreno de la fonología, hay que mencionar en primer lugar los libros de TADEUSZ BATOG, *The Axiomatic Method in Phonology*, Routledge and Kegan Paul, Londres, 1967, y de F. H. H. KORTLANDT, *Modelling the Phoneme*, Mouton, La Haya, 1972. El primero es una investigación, basada en la mereología de Leśniewski, sobre el modo de pasar de lo continuo a lo discreto en el material sonoro de la lengua; el segundo, a su vez, constituye una síntesis crítica, con muchos aspectos novedosos, de los distintos modelos matemáticos en fonología propuestos en la Europa del Este. A éstos cabe añadir, en una cualificación totalmente aparte, el libro de S. K. SHAUMIAN, *Problems of Theoretical Phonology*, Mouton, La Haya, 1968, donde la lógica formal y los elementos de la teoría de los conjuntos y de los grafos conforman el marco de una concepción fonológica especialmente original.

La teoría de las gramáticas formales viene sintetizada en monografías muy notables como: MAURICE GROSS y ANDRÉ LENTIN, *Notions sur les grammaires formelles*, Gauthier-Villars, París, 1968 (buena introducción para el lingüista que posee algunos rudimentos de lógica, teoría de los conjuntos y álgebra; en el libro se pasa revista a numerosos tipos de sistemas formales, gramáticas y autómatas sin entrar, no obstante, en detalles); JOHN E. HUPCROFT y JEFFREY D. ULLMAN, *Formal Languages and their Relation to Automata*, Addison-Wesley Publishing Company, Reading, Mass., 1969; ARTOO SALOMAA, *Formal Languages*, Academic Press, Nueva York-Londres, 1973. Los dos últimos libros requieren una preparación matemática superior y, desafortunadamente, apenas contienen referencias explícitas a las lenguas naturales. En cambio, desarrollan hechos matemáticos de base en los que se apoya la teoría de las gramáticas generativas. Más equilibrados resultan, a este propósito, los libros de JOHN P. KIMBALL, *The Formal Theory of Grammar (Foundations of Modern Linguistics)*, Englewood Cliffs, Prentice-Hall, 1973; MAURICE GROSS, *Mathematical Models in Linguistics*, Englewood Cliffs, Prentice-Hall, 1972 y ROBERT WALL, *Introduction to Mathematical Linguistics*, Englewood Cliffs, Prentice-Hall, 1972, donde las consideraciones matemáticas se combinan siempre con comentarios relativos a las lenguas naturales o artificiales. De la bibliografía soviética especializada citaremos el libro particularmente riguroso de A. V. GLADKI, *Formalnye grammatiki i iazyki*, Izd. Nauka, Moscú, 1973. Entre los libros tradicionales me-

rece recordarse el de SOLOMON MARCUS, *Gramatici si automate finite*, Editura Academiei, Bucarest, 1964, dedicado a las gramáticas de tipo 3 en la jerarquía de Chomsky (cf. 3.13) y el de SEYMOUR GINSBURG, *The Mathematical Theory of Context Free Languages*, McGraw-Hill Book Comp., Nueva York, 1966, dedicado a las gramáticas de tipo 2 en aquella jerarquía. Una teoría generativa original, que consiste en explotar y combinar distintos tipos de autómatas que operan a diferentes niveles lingüísticos, ha sido propuesta por PETR SGALL, LADISLAV NEBESKÝ, ALLA GORALČIKOVÁ y EVA HAJIČOVÁ en el libro *A Functional Approach to Syntax in Generative Description of Language*, Elsevier, Nueva York, 1969, donde el carácter mixto del equipo de autores (dos lingüistas y dos matemáticos) se refleja fielmente en la estructura del trabajo, en el que los modelos algebraicos discurren con naturalidad a partir de motivaciones de índole lingüística.

Otros libros se refieren a modelos matemáticos o lógicos en semántica: EMANUEL VASILIU, *Outline of a Semantic Theory of Kernel Sentences*, Janua Linguarum, Series Maior 71, Mouton, La Haya, 1972 y PETER ROLF LUTZEIER, *Der "Aspekt" Welt als Einstieg zu einem nützlichen Kontextbegriff für eine natürliche Sprache*, Stuttgart, 1974, en los que se hace característica la fuerte tendencia matematizante de la semántica. Con relación a estos dos libros, y a muchos otros de los que hemos hablado antes o de los que hablaremos en adelante, hay que observar que el empleo de los métodos matemáticos deja de indicarse explícitamente en el título. Este fenómeno, hoy frecuente, era una rareza diez o quince años atrás. Los métodos matemáticos se han extendido tanto en lingüística, que muchos autores no sienten la necesidad de advertir al lector sobre este hecho. De ahí que la expresión "lingüística matemática" tiende a convertirse en un pleonasma.

También las investigaciones estadísticas en lingüística han experimentado un gran progreso. Conviene tener en cuenta los libros de V. A. MOSCOVICH, *Statistika i semantika*, Izd. Nauka, Moscú, 1969; CHARLES MÜLLER, *Initiation à la statistique linguistique*, Librairie Larousse, París, 1968¹⁷ y BARRON BRAINERD, *Weighing Evidence in Language and Literature. A Statistical Approach*, Toronto Univ. Press, 1975, los cuales, respetando todo el rigor y las matizaciones de los métodos modernos de la estadística matemática, conceden una gran atención a la accesibilidad y a las motivaciones de orden lingüístico. Este hecho no podía atribuirse a los libros de estadística lingüística publicados en la década de 1960: los de PIERRE GUIRAUD, abundantes en sugerencias, eran insuficientes en cuanto a la fundamentación matemática de los datos, mientras que los libros de GUSTAV HERDAN resultaban inaccesibles al lingüista.

En torno a los métodos estadísticos tampoco pueden olvidarse los procedi-

17. [Hay traducción española; cf. Bibliografía].

mientos de la teoría de la información. Mencionaremos, a este respecto, los trabajos de R. G. PIOTROVSKI de Leningrado y, sobre todo, el libro de PETR BECKMANN, *The Structure of Language. A New Approach*, The Golem Press, Boulder-Colorado, 1972, donde por primera vez las lenguas naturales están ya enfocadas, de una manera sistemática, desde el punto de vista de la teoría de los códigos detectores o correctores de errores (cf. 10.6 y 7). Para los interesados en las cuestiones que se tratan en los capítulos 8 y 10 del presente libro, la labor de Beckmann supone una continuación natural.

Dentro de la perspectiva de los problemas tratados en los capítulos 4, 5 y 7, dedicados a la lingüística computacional, hoy extraordinariamente desarrollada, cabe destacar libros como los de DAVID HAYS, *Readings in Automatic Language Processing*, American Elsevier, Nueva York, 1966 y H. BRANDT CORSTIUS, *Exercises in Computational Linguistics*, Mathematisch Centrum, Amsterdam, 1970. Añadamos también a ARNOLD W. PRATT, A. HOOD ROBERTS y KATHLEEN LEWIS (eds.), *Seminar on Computational Linguistics*, U. S. Dept. of Health, Education and Welfare, octubre de 1966. El último decenio se caracteriza por la masiva penetración de los métodos matemáticos y de los ordenadores en filología (problemas de fechación, de localización, de paternidad de textos, de filiación de manuscritos, de parentesco idiomático, etc.). Hacia estos intereses se dirigen, significativamente, las conferencias internacionales que organiza periódicamente la Asociación Internacional de Lingüística Computacional (1967, Grenoble; 1969, Sanga-Saby, Suecia; 1971, Debrecen; 1973, Pisa), así como la Conferencia, con participación internacional, de Varna (Bulgaria), celebrada en mayo de 1975. La filología es un dominio apto por excelencia para el uso de los ordenadores electrónicos, por cuanto los problemas filológicos contienen un elevado número de datos cuyo tratamiento es manualmente irrealizable. Un libro típico en este sentido es el de Dom FROGER, *La critique des textes et son automatisations*, Dunod, París, 1968, en el cual ciertos métodos conocidos en principio quedan explicitados con ayuda de la teoría de los grafos y transformados virtualmente en algoritmos programables en el ordenador.

La aplicación de los ordenadores en lingüística marca el paso de un tipo de estudio basado en muestras relativamente reducidas a la confección de unos "corpus" cada vez más amplios, con tendencia a la exhaustividad.

Una gran cantidad de monografías se ha consagrado durante el último decenio bien a los problemas generales de la lingüística matemática (A. V. GLADKI y I. A. MELCHUK, *Elementy matematicheskoi lingvistiki*, Izd. Nauka, Moscú, 1969¹⁸; BARRON BRAINERD, *Introduction to the Mathematics of Language Study*, American Elsevier, Nueva York, 1972), bien a determinados problemas especiales (I. I. REVZIN, *Les modèles linguistiques*, Dunod, 1967 y *Metod mode-*

18. [Hay traducción española; cf. Bibliografía. Cf., asimismo, HOCKETT. *Language*].

lirovania i tipologia slavianskyj iazykov, Izd. Nauka, Moscú, 1967; SOLOMON MARCUS, *Algebraic Linguistics. Analytical Models*, Academic Press, Nueva York-Londres, 1967, *Introduction mathématique à la linguistique structurale*, Dunod, París, 1967, *Poetica matematica*, Editura Academiei, Bucarest, 1970); ERNESTO ZIERER, *La teoría de los gráficos en la lingüística*, Universidad Nacional de Trujillo, Perú, 1967; FERENC KIEFER, *Mathematical Linguistics in Eastern Europe*, American Elsevier, Nueva York, 1968; ALEJANDRO ORTIZ y ERNESTO ZIERER, *Set Theory and Linguistics*, Mouton, La Haya, 1968; LADISLAV NEBESKÝ, *Algebraic Properties of Trees*, Universita Karlova, Praga, 1969; HANS HEINRICH LIEB, *Sprachstadium und Sprachsystem*, Kohlhammer, Stuttgart, 1970; S. K. SHAUMIAN (ed.), *Matematicheskaia lingvistika*, Izd. Nauka, Moscú, 1973. Un volumen entero (J. HORECKÝ, L. KALMÁR y S. MARCUS, eds.) del *Recueil Linguistique de Bratislava* (vol. IV), Bratislava, 1973, se ha consagrado a la lingüística algebraica a partir de las comunicaciones del Simposium de Lingüística Algebraica de Smolenice, Checoslovaquia, celebrado del 10 al 12 de febrero de 1970, del mismo modo que el número 10 (S. MARCUS, ed.) de la Revista *Poetics* de Holanda se ha dedicado a los métodos matemáticos en la poesía. Mención aparte merece el libro de RENÉ THOM, *Stabilité structurelle et morphogénèse. Essai d'une théorie générale des modèles*, W. A. Benjamin, Inc., Reading, Mass., 1972, donde se ofrece una concepción nueva sobre el comportamiento biológico y lingüístico del hombre basada en la teoría matemática de las catástrofes que fundó precisamente el mismo René Thom.

Un gran incremento han experimentado asimismo las revistas consagradas en su totalidad o en gran parte a la utilización de los métodos lógico-matemáticos y de los ordenadores en lingüística y en poética. Consignemos de entrada que las revistas de recensiones *Mathematical Reviews* y *Zentralblatt für Mathematik und ihre Grenzgebiete* reseñan con regularidad, en la sección de *Computing Machines*, artículos de lingüística matemática, y que la revista soviética *Referativny Zhurnal-Matematika* hace lo propio dentro de la rúbrica *Matematicheskie Problemy Semiotiki*. A su vez, la revista *Referativny Zhurnal-Informatics, Abstracts* está dedicada por completo a la lingüística matemática y computacional y a la documentación automática. En 1974 aparecen la revista *Theoretical Linguistics* (HELMUT SCHNELLE, ed.), publicada por Walter de Gruyter, Alemania Federal, y el órgano de la Asociación Americana de Lingüística Computacional, *American Journal of Computational Linguistics* (DAVID G. HAYS, ed.). La revista *Computers and the Humanities* reserva regularmente algunas secciones al empleo de ordenadores en lingüística y en el estudio de la literatura. La revista peruana *Lenguaje y Ciencias* (ERNESTO ZIERER, ed.) está consagrada a los métodos lógico-matemáticos y a los ordenadores en lingüística. Otras revistas como *Foundations of Language, Linguistics, Voprosy Iazykoznanja, Linguistic Inquiry, Journal of Linguistics, Langages* o *Lingua e Stile* han publicado

numerosos artículos que utilizan métodos lógico-matemáticos. La publicación *Prague Studies of Mathematical Linguistics*, de la editorial "Academia" de Praga, que aparece cada dos años, incluye exclusivamente trabajos de lingüística algebraica, estadística y computacional. Otros artículos de lingüística con carácter decididamente matemático pueden hallarse también en *International Journal of Computer Mathematics* (Nueva York), *Discrete Mathematics* (Amsterdam), *Mathematical System Theory*, etc.

Últimamente surgen numerosas aplicaciones de la lingüística matemática a otras ciencias. Son típicos en este sentido los abundantes artículos de la revista *Semiotica* (TH. A. SEBEOK, ed., Mouton) y un libro como el de MARIE NOWAKOWSKÁ, *Language of Motivation and Language of Action*, Mouton, La Haya, 1974.

3. Modelos matemáticos en lingüística

3.1. Concepto de modelo y modelación

La unidad material del mundo o compenetración dialéctica de las distintas formas de acción material posibilita el empleo fructífero de la analogía en el proceso cognoscitivo y de investigación científica. Si podemos establecer una cierta analogía entre los objetos (o fenómenos) *A* y *B*, entonces el estudio de uno de ellos puede beneficiarse de los resultados obtenidos en el estudio del otro. Este método se hace interesante en tanto que los objetos *A* y *B* sean suficientemente distintos para que uno de ellos pueda ser investigado al menos por un método inaplicable al otro. Si, p. ej., *A* se presta a una prueba experimental inasequible a *B*, cabe, sin embargo, obtener algunas informaciones acerca de este objeto *B* a partir de las conclusiones extraídas del experimento sobre *A* y de los rasgos comunes que previamente se hubieran reconocido en ambos. He aquí cómo, por una vía indirecta, la investigación de un objeto o fenómeno, cuya naturaleza u otra condición no le hacen accesible a los experimentos, puede no obstante aprovechar las ventajas del método experimental. Así ocurre con los fenómenos astronómicos, cuyo conocimiento depende en cierta medida de los resultados obtenidos en experiencias analógicas. Tampoco las enfermedades del organismo humano permiten una experimentación directa, en la medida en que se respeten los imperativos humanitarios, pero gracias a los numerosos rasgos comunes que existen entre el organismo humano y el de otros seres vivos con los que cabe experimentar directamente, también la medicina se beneficia decisivamente del método experimental.

Sea un objeto o fenómeno *A*, sometido a una prueba como sustitutivo de otro objeto o fenómeno *B*, con el cual *A* mantiene una cierta correspondencia. El proceso por el cual, de las informaciones extraídas a partir de la investigación sobre *A* y de la correspondencia entre *A* y *B*, se obtienen ciertas informaciones sobre *B* constituye un proceso de modelación o, más abreviada-

mente, una *m o d e l a c i ó n*. Decimos, así, que *B* ha sido modelado a partir de *A*, que *B* es el original y *A*, un modelo de *B*¹.

La modelación será *m a t e r i a l* o *i d e a l*, según la naturaleza material o no de *A*: La modelación material en que la correspondencia entre *A* y *B* conste de naturaleza física e identidad de forma geométrica, pero no de dimensión, desarrollo, velocidad, etc. determina una *m o d e l a c i ó n d e s e m e j a n z a*. Y si, en una modelación material, la correspondencia entre *A* y *B* no deriva de la naturaleza física o de la forma geométrica común, sino del hecho por el que determinado aspecto funcional de *A* aparezca también en *B* (p. ej., que una misma ecuación describa a *A* y a *B*), entonces es que se trata de una *m o d e l a c i ó n a n a l ó g i c a*².

Entre los modelos ideales, los más interesantes se producen cuando *A* es un objeto o una construcción matemática. Se forma, así, una *m o d e l a c i ó n m a t e m á t i c a*, donde *A* resulta un modelo matemático de *B*.

Un modelo matemático se presenta, por lo común, del siguiente modo. Se consideran ciertos conjuntos de elementos y se definen, en estos conjuntos, determinadas relaciones y operaciones. Se postulan algunas propiedades de los conjuntos, relaciones y operaciones considerados. Luego, por vía deductiva, se obtienen nuevas propiedades que conducen a la definición de nuevos conjuntos, relaciones y operaciones, y se procede así sucesivamente. Como los conjuntos, relaciones y operaciones consideradas al principio se han elegido de tal modo que admitan una interpretación relativa al objeto modelado, los resultados obtenidos deductivamente admitirán también una interpretación del mismo tipo.

Una particularidad de la modelación matemática de un objeto o fenómeno proviene de su carácter indirecto. Un fenómeno natural no se presta por lo general a una modelación matemática directa. Al principio, se ensaya una descripción no modélica o una modelación no matemática del fenómeno como base de partida para intentar la modelación matemática. Dicho de otro modo, en una modelación matemática, *B* no es normalmente el propio fenómeno natural estudiado, sino una descripción no matemática del mismo. A veces, el carácter indirecto de este tipo de modelaciones matemáticas resulta aún más pronunciado, de forma que entre el modelo y el fenómeno se intercalan más descripciones intermedias.

Existen ciertas diferencias entre el modo en que se realiza el conocimiento de los modelos materiales, por una parte, y el de los matemáticos, por otra. Pero como desde el punto de vista lingüístico la modelación material tiene una importancia menor (sólo interviene en fonética) y como esta modelación per-

1. Cf. FITIALOV, "Modelele", y ZINOVIEV-REVZIN, "Modelul".

2. Cf. GUTENMACHER, "Modelirovanie matematicheskoe"; KOLMOGOROV, "Modelirovanie"; SĂHLEANU, "Valoarea", y ZINOVIEV-REVZIN, "Modelul".

manece fuera del propósito de nuestra exposición, no trataremos aquí estas diferencias.

3.2. Necesidad de la modelación matemática de la lengua

Supongamos ahora que B es un fenómeno lingüístico. Como tal admite un estudio directo por medio de la observación inductiva. Sin embargo, el conjunto de los actos de habla es infinito. Las exigencias de la lingüística aplicada y del estudio de la estructura lingüística requieren la jerarquización y delimitación de un número mínimo de elementos y de unas propiedades atribuidas a ellos, a partir de las cuales las demás formaciones y propiedades se obtengan por deducción. Todas estas consideraciones reclaman la introducción de los modelos deductivos en la lingüística. Pero estos modelos deductivos no pueden aplicarse con éxito si no están integrados en los llamados sistemas formales. Los fenómenos lingüísticos se hallan muy lejos de la estructura típica de esta clase de sistemas. ¿Cómo resolver, entonces, este problema? En este punto, la teoría materialista dialéctica sobre el conocimiento brinda una respuesta precisa: a base de la modelación matemática de los fenómenos lingüísticos. En efecto, si el objeto matemático A es un modelo del fenómeno lingüístico B , entonces A puede someterse a la investigación deductiva y, como A es un sustitutivo o imitador parcial de B , los resultados extraídos tendrán una cierta significación con respecto a B . De esta manera, la modelación matemática de los fenómenos lingüísticos permite estudiar a éstos de un modo mediano y por métodos incompatibles a su estructura propia. Y al mismo tiempo, el estudio de los fenómenos lingüísticos se beneficia de los métodos y resultados de la matemática. Ahí radica, en esencia, la función gnoseológica de los modelos matemáticos aplicados a la lengua. Desde luego, para un fenómeno lingüístico B cabe encontrar una infinidad de modelos; tales, que cada uno de ellos proporciona un conocimiento parcial sobre el fenómeno en cuestión y contribuye a la elaboración de la teoría del fenómeno lingüístico considerado.

3.3. Modelaciones concomitantes

La necesidad de modelaciones concomitantes sobre el mismo objeto, fenómeno o proceso surge no sólo porque cada uno de los aspectos posibles reclama una modelación especial que no puede atender, por lo común, a otros aspectos; incluso para un mismo aspecto es recomendable y necesaria la participación de más de una modelación efectuada desde los dominios más variados. En este sentido cabe establecer, para los distintos elementos de un modelo, la capacidad

respectiva en reflejar algunas propiedades de la realidad modelada, puesto que la confrontación de distintos modelos sobre el mismo aspecto de un objeto —una vez establecida la correspondencia entre los modelos empleados— separa lo que no pertenece al original, sino al dominio de valores en cuyos términos se expresa el modelo. Así, para los llamados lenguajes con un número finito de estados se han construido modelos a partir de la lógica matemática y de la teoría de conjuntos, a partir de la teoría de los grafos y de la teoría de los autómatas finitos³. La comparación de estos modelos ha mostrado qué parte de ellos refleja aspectos de los lenguajes con un número finito de estados y qué pertenece al dominio matemático en cuyos términos se ha construido la modelación sin tener, por tanto, ningún valor definitorio acerca de aquellos lenguajes.

También aquí se aplica la observación de antes sobre el carácter mediato de la modelación matemática de un fenómeno natural. En rigor, no se modela matemáticamente un fenómeno lingüístico como tal, sino una descripción de un fenómeno análogo o un modelo no matemático de una descripción de este tipo. Estas etapas intermediarias que se interponen entre el modelo y el original han sido facilitadas por el desarrollo entero de la lingüística y, sobre todo, por la lingüística estructural (cf. 2.18).

3.4. Analogía, modelo, teoría

La presente exposición indica claramente que el modelo es mucho más que una simple analogía, ya que cumple una función eurística; pero es mucho menos que la teoría misma del fenómeno en cuestión, por cuanto proporciona un conocimiento menos complejo que ella.

La analogía entre las palabras y el dinero no determina una modelación. Ni el cálculo de los tipos sintácticos de Bar-Hillel, Harwood o Lambek, ni el concepto de tipo introducido por Kulaguina, ni el modelo mixto paradigmático-generativo propuesto por Shaumian constituyen la teoría de las categorías gramaticales sobre las partes del discurso⁴. Pero cada uno de ellos es un modelo de las partes del discurso y todos contribuyen, junto con las descripciones no modélicas o semimodélicas de Hjelmslev, Kuryłowicz, Harris y otros, a elaborar la teoría general sobre las partes del discurso⁵.

3. Para la teoría de conjuntos, cf. BAR-HILLEL y SHAMIR, "Finite-state languages"; CHOMSKY-MILLER, *Finite State Languages*, y MARCUS, *Grammatici*; para la teoría de los grafos, cf. ČULÍK, "Some notes", y ZIERER, *Teoría*, y para la teoría de los autómatas finitos, cf. BAR-HILLEL y SHAMIR, "Finite-state languages", y RABIN-SCOTT, "Finite automata".

4. Cf. BAR-HILLEL, "Quasi-arithmetical notation"; HARWOOD, "Axiomatic syntax"; LAMBEK, "Mathematics", "Contributions" y "Calculus"; KULAGUINA, "Ob odnom sposobe" y SHAUMIAN, "Lingvisticheskie vo-prosy" [y *Strukturnaia Lingvistika*].

5. Cf. HJELMSLEV, *Prolegomena*; KURYŁOWICZ, "Isomorphisme"; HARRIS, *Methods* [y *Structural Linguistics*]; FITIALOV, "Formalno-matematicheskie modeli", y ZINOVIEV-REVZIN, "Modelul".

3.5. El modelo omite aspectos esenciales

Con respecto a la modelación matemática de objetos y procesos más complejos, como los biológicos o de disciplinas humanísticas, se han hecho cada vez más evidentes la posibilidad y la necesidad no sólo de que el modelo refleje algunos aspectos esenciales del objeto, sino de que omita, al mismo tiempo, otros aspectos esenciales. Así, los modelos construidos a través de la teoría de conjuntos y de la lógica matemática por parte de Kulaguina, Bratchikov, Fitalov, et-cétera, revelan aspectos esenciales de la lengua, concretamente relacionados con los sistemas lingüísticos potenciales; en cambio, los mismos modelos soslayan un aspecto tan esencial a la lengua como el probabilístico⁶. Por su parte, los aspectos probabilísticos de la lengua están incluidos en modelos como los de B. Mandelbrot, Herdan o Guiraud, pero estos mismos modelos ignoran la estructura lógica de la lengua⁷. Los modelos de Kolmogorov para la poesía de Pushkin y Maiakovski presentan una situación similar, ya que atienden de un modo profundo a aspectos relativos al ritmo, entonación y acento de los poemas considerados, pero también silencian, incluso al tratar de los aspectos probabilísticos, aspectos tales como los que se refieren a la estructura en "cadena de Markov" en la cual los eventos son las apariciones de palabras.

3.6. La importancia de la modelación matemática para la lingüística aplicada

La autenticidad de los conocimientos proporcionados por la modelación matemática ha pasado con éxito la prueba de la confrontación con la práctica. En una modelación de este tipo se ha basado la elaboración de lenguas lógicas de información para distintos dominios científicos, así como también la composición de los algoritmos para la traducción automática. A su vez, la lingüística aplicada, en la acepción cibernética del término, ha estimulado profundamente la construcción de modelos matemáticos de la lengua, en especial de modelos sintagmáticos. De hecho, cualquier intento de promover una de estas esferas sin tener en cuenta a la otra es perjudicial. La modelación matemática, en ausencia de los sugerimientos y los problemas de la lingüística aplicada, termina por perder la vitalidad. Y a su vez, la lingüística aplicada tampoco puede prescindir de la modelación y de la teoría de la lengua⁸.

6. Cf. KULAGUINA, "Ob odnom sposobe"; BRATCHIKOV, "Nekotorye teoremy", y FITALOV, "O modelirovaniya sintaksisa" y "O postroeni morfologii".

7. Cf. MANDELBROT, "Informational theory" y "Structure"; HERDAN, *Type-token Mathematics*, y GUIRAUD, *Caractères statistiques*.

8. Cf. ANDREEV-ZINDER, "Osnovnye problemy"; BAR-HILLEL, "Present status" y "The future"; BRATCHIKOV-FITALOV-CHEITIN, "O strukture informatsii"; FITALOV, "Formalno-matematicheskie modeli"; IVANOV, "Lingvisticheskie voprosy"; MOLOSHNAIA, "O poniati grammaticheskoi konfiguratsi", y SHAUMIAN, "Lingvisticheskie voprosy".

No es posible comprender con exactitud la utilidad y las perspectivas de la modelación matemática de la lengua sin analizar atentamente el desarrollo que ha experimentado, en época reciente, la lingüística aplicada. Antes de aparecer la cibernética, las aplicaciones lingüísticas se limitaban a actividades tales como la composición de diccionarios, la enseñanza de lenguas, la creación y perfeccionamiento de la escritura, el establecimiento de normas para la transcripción y la transliteración, la búsqueda de propuestas convencionales para reglamentar la terminología científica, etcétera. Junto a estas actividades, que conservan, por lo demás, su importancia, han aparecido nuevas esferas de aplicación para la lingüística a causa del progreso científico y técnico y también de las nuevas condiciones sociales. Nos referimos, en especial, a la necesidad y posibilidad de mejorar los recursos de comunicación (telefonía, telegrafía, radio, estenografía, etc.) y de crear nuevos procedimientos de este tipo. Tenemos, al mismo tiempo, en cuenta la construcción de dispositivos para introducir información lingüística en las máquinas (entre las que hay que contar los aparatos eléctricos de escucha, de lectura, etc.), para extraer de las máquinas la información en forma lingüística y para elaborar la información verbal en el interior de las máquinas electrónicas de cálculo. Se han hecho patentes también la necesidad y posibilidad de construir lenguas lógicas de información para distintos dominios científicos, y algoritmos de traducción automática de una lengua a otra. Las lenguas lógicas de información son lenguas formadas con la ayuda de la lógica matemática y cuyo propósito consiste en permitir la introducción y elaboración, en máquinas electrónicas de cálculo, de la información contenida en un determinado dominio científico. Una lengua lógica de este tipo registra, con sus procedimientos específicos (sobre todo construyendo un cálculo dado de predicados), las construcciones típicas de los textos respectivos junto con sus caracteres léxicos y gramaticales. El establecimiento de las lenguas de información es una tarea urgente, en especial para dominios como la biología y la química, donde la bibliografía especializada se ha hecho ya tan vasta que a sus respectivos estudiosos pronto les faltará tiempo hasta para hojear todo lo que se publica. La automatización del proceso de información científica es, pues, un importante problema de nuestra sociedad.

Acaso esta manera de presentar los hechos dé la impresión de que los problemas fundamentales que aparecen en los nuevos dominios de la lingüística aplicada afectan, en primer lugar, a la ingeniería y a la técnica y que, en el mejor de los casos, apenas ofrecen relaciones con la lingüística teórica. En realidad, tal como se ha demostrado con creces en los últimos años, las principales dificultades planteadas en el perfeccionamiento de los algoritmos de traducción automática y en la composición de las lenguas lógicas de información provienen del insuficiente desarrollo de la lingüística, en su nivel estructural y de modelación formal y matemática sobre los fenómenos de la lengua. Esto es explicable:

mientras se emplea la lengua como medio de comunicación directa entre los hombres, la comprensión se realiza por el hecho de que cada uno de los interlocutores apela a las relaciones significativas que hay entre las palabras de las frases. Pero en cuanto debe introducirse la información lingüística en la memoria de una máquina electrónica de cálculo o transmitirse por una determinada línea de comunicación, el nuevo interlocutor —la máquina o la línea— ya no es capaz de utilizar las relaciones de significado, y debe remitirse exclusivamente a las relaciones formales, codificables y traducibles al lenguaje de las válvulas electrónicas o a los impulsos eléctricos. Es, por tanto, necesaria una descripción previa de los distintos tipos de relaciones gramaticales sobre la base única de la posición mutua de los elementos en la frase, y un estudio sistemático de la lengua desde el punto de vista de sus aspectos de código, de la teoría de la información. Cuando, hace unos veinte años, se efectuaron los primeros experimentos de traducción automática, se intentó confeccionar *ad hoc* unas descripciones de este tipo con las que revelar las necesidades inmediatas de la experiencia en que se trabajaba. Sin embargo, las posibilidades de este tipo de expedientes se agotaron rápidamente. Más tarde se ha comprobado que sin el correspondiente desarrollo de la lingüística estructural y matemática y, por consiguiente, de la teoría de la lengua al nivel más abstracto, no pueden obtenerse progresos sustanciales ni se puede avanzar con cierta perspectiva en el terreno de la lingüística aplicada.

Esta situación ha provocado un gran ímpetu en las investigaciones de lingüística estructural y, sobre todo, en lo que se refiere a la modelación matemática de la lengua. La colaboración de estas dos ramas, la lingüística estructural y matemática por una parte y la lingüística aplicada por la otra, y el modo en que se han estimulado recíprocamente ilustra de una manera lúcida la teoría materialista dialéctica sobre la interdependencia y la unidad de la teoría y la práctica. Muchos de los modelos matemáticos de la lengua elaborados en los últimos años son producto de sugerimientos inspirados en distintas actividades de la lingüística aplicada: el modelo basado en la teoría de conjuntos por O. S. Kulaguina, el modelo probabilista de N. D. Andreev, los modelos de Fitalov y Bratchikov sobre morfología, los modelos proyectivos (sobre los que hablaremos luego) de Lecerf, Ihm, Camion y otros, el modelo de Yngve sobre las estructuras regresivas, así como muchos otros entran en esta categoría (cf. 3.12 y ss.).

Por otra parte, la teoría y la práctica de algunas disciplinas matemáticas y cibernéticas orientan, a su vez, la modelación matemática de la lengua hacia vínculos cada vez más sólidos entre los compartimentos más diferentes de la ciencia. Así, el desarrollo de la teoría de los autómatas finitos ha provocado la elaboración de las llamadas gramáticas con un número finito de estados, mientras que en el ámbito de la teoría de los algoritmos, de las funciones recursivas y de la máquina de Turing han llegado a aparecer modelos tales como las gra-

máticas de constituyentes inmediatos y las gramáticas transformativas. Al propio tiempo, estos mismos modelos contribuyen al progreso de muchas ramas de la lógica matemática y la cibernética: no es una casualidad que trabajos como los de N. Chomsky, el principal autor de las gramáticas generativas, vengán citados cada vez más a menudo en artículos de lógica matemática o de teoría de los autómatas. En la compleja dialéctica de estas relaciones planteadas entre dominios aparentemente tan heterogéneos, están involucradas incluso otras disciplinas, como las teorías de los semigrupos y de los números.

La modelación matemática de la lengua ha empezado a ejercer una influencia decisiva sobre la lingüística aplicada. Durante los últimos años ha comenzado la elaboración de una teoría científica de la traducción (cf. 7.1). La operación de traducir de una lengua a otra, considerada por mucho tiempo como una cuestión de simple destreza y habilidad por parte del traductor, se ha convertido en objeto de una investigación científica donde nociones tales como “traducción palabra por palabra”, “traducción libre”, “traducción artística”, etc. están modeladas matemáticamente. Merecen citarse, en esta dirección, experiencias tan interesantes como las realizadas por I. I. Revzin, V. I. Rozentveig y otros, en la Unión Soviética. No cabe duda que a través de la clasificación de las bases teóricas o lingüísticas de la traducción de una lengua a otra y a través de una comprensión más lúcida de lo que ocurre en el proceso de la traducción, incluso el tratamiento de las lenguas de información va a experimentar un gran progreso, ya que también aquí se trata de una traducción de una lengua natural a una lengua lógica.

3.7. Clasificación de los modelos matemáticos de la lengua

Los modelos matemáticos de la lengua admiten una clasificación con arreglo a distintos puntos de vista. Si se considera dado un cierto conjunto de secuencias (interpretado como una parte del conjunto de las secuencias aceptadas de la lengua) y se establecen, por vía axiomático-deductiva, las reglas que conforman aquellas secuencias a partir de un análisis de los elementos componentes que, ulteriormente, son clasificados y ordenados en virtud de una determinada jerarquía, entonces nos hallamos en presencia de un modelo analítico. A este tipo de modelos pertenecen: el modelo de O. S. Kulaguina para la gramática, el modelo compuesto por S. I. Fitialov y I. L. Bratchikov para algunos aspectos del análisis morfológico, el modelo de A. F. Parker-Rhodes para la sintaxis, etcétera⁹. Si, por el contrario, se parte de un conjunto de reglas distribuidas en una construcción formalizada, generalmente algorítmica, cuyo re-

9. Cf. KULAGUINA, “Ob odnom sposobe”; FITALOV, “O postroeni morfologuii”; BRATCHIKOV, “Nekotorye teoremy”; PARKER-RHODES, *Lattice Properties*, y PARKER-RHODES *et al.*, *Lattice Model*, “A new model” y “Derivation”.

sultado final determine la obtención de un cierto conjunto de secuencias, entonces se produce un modelo sintético o generativo. En él se incluyen las gramáticas con un número finito de estados, las gramáticas de constituyentes inmediatos y las gramáticas transformacionales elaboradas por N. Chomsky, G. A. Miller, M. P. Schützenberger, las gramáticas categoriales (bidireccionales o unidireccionales) y secuenciales estudiadas por Y. Bar-Hillel, M. Perles, C. Gaifman y E. Shamir, R. P. Mitchell, etc.¹⁰

Tal como hemos observado antes, la clasificación de los modelos en analíticos y sintéticos ha tomado por criterio la posición, en el modelo, del conjunto de secuencias aducidas. Otra clasificación se basa en el punto de vista de la teoría matemática empleada en el proceso de la modelación. Existen, a este propósito, modelos fundados en el álgebra y en la teoría de conjuntos, en el cálculo de probabilidades, en la lógica matemática, en las funciones booleanas, en la teoría de los retículos, en la de los grafos, etc.¹¹

Los modelos de la lengua se clasifican todavía en sintagmáticos y paradigmáticos. Los primeros reflejan las relaciones contraídas en la cadena o flujo del habla y, por tanto, las relaciones de carácter contextual. Las otras, en cambio, revelan relaciones entabladas por los elementos lingüísticos en el marco del sistema. El modelo de V. Yngve sobre las estructuras regresivas (cf. 3.17), las gramáticas proyectivas referentes a las relaciones entre el orden lineal y el orden estructural, los modelos contenidos en los trabajos de K. I. Babitski y S. I. Fitialov, el modelo configuracional de O. S. Kulaguina son otros tantos ejemplos de modelos sintagmáticos¹². Los modelos relativos al fonema, contruidos por J. H. Greenberg, T. Batog, I. I. Revzin, el referente a las partes

10. Cf. CHOMSKY, "Three models", *Syntactic Structures*, "Formal properties" y *Aspects*; CHOMSKY-MILLER, "Finite state languages"; SCHÜTZENBERGER, "A problem" y "Context-free languages"; BAR-HILLEL, "Recent results"; BAR-HILLEL, GAIFMAN y SHAMIR, "Categorial grammars"; BAR-HILLEL, PERLES y SHAMIR, "Formal properties"; GAIFMAN, *Dependency Systems*; MITCHELL, "Properties", y SHAMIR, *Sequencial Languages* y "A remark".

11. Para los modelos basados en el álgebra y la teoría de conjuntos, cf. KULAGUINA, "Ob odnom sposobe"; USPENSKI, "K opredelenin chastí rechi" y "K opredelenin padezha"; REVZIN, *Modelli*, "O nekotoryj voprosaj" y "Ob odnom podjode"; DOBRUSHIN, "Grammaticheskaia kategorija" y "Matematicheskie metody", y MARCUS, "O analiză", *Lingvistică matematică*, "Modèles mathématiques" e *Introduction*. Para el cálculo de probabilidades, cf. ANDREEV, "Modelirovanie iazyka"; MANDELROT, *Informational Theory* y "Structure formelle"; GUIRAUD, *Caractères statistiques*; HERDAN, *Type-token Mathematics*, y AJMANOVA-MELCHUK-PADUCHEVA-FRUMKINA, "O tochinij metodaj". Para los modelos basados en la lógica matemática, cf. KUZNETSOV-PADUCHEVA-ERMOLEEVA, "Ob informatsionnom iazyke"; QUINE, "Logic"; LAMBEK, "Mathematics" y "Calculus"; BAR-HILLEL, "Quasi-arithmetical notation", "Three remarks", "Recursive definitions", "Recent results", "Present status" y "The future"; BATOG, "Logiczna rekonstrukcja" y KALUZHNNIN, "Ob informatsionnom iazyke". Sobre los modelos basados en las funciones booleanas, cf. el modelo del fonema, elaborado por BELEVITCH, *Langage*. Para los modelos basados en la teoría de los retículos, cf. PARKER-RHODES, *Lattice Properties*. Y para la aplicación de la teoría de los grafos, cf. ČULIK, "Some notes"; GOODMAN, "Graphs"; MARCUS-VASILU, "Matematică și fonologie"; ZIERER, *Teoria*, etc.

12. Cf. YNGVE, "A model"; LECERF-IHM, "Éléments"; HIRSCHBERG-LYNCH, *Discussions*; CAMION, "Analyse"; BABITSKI, "K voprosu"; FTITALOV, "O modelirovania" y KULAGUINA, "Ob odnom sposobe".

del discurso elaborado por O. S. Kulaguina y V. A. Uspenski pertenecen, a su vez, al tipo de modelos paradigmáticos¹³.

Estas clasificaciones, a excepción de la primera, contemplan más bien tipos límite, ideales, de modelos. Por lo común, en un mismo modelo se hallan implicados distintos aparatos matemáticos (p. ej., el modelo de N. D. Andreev (cf. nota 11) se basa no sólo en el cálculo de probabilidades, sino también en la teoría de conjuntos), del mismo modo que en un mismo modelo se entrelazan el análisis sintagmático y el paradigmático. Así ocurre, p. ej., en los modelos del caso compuestos por A. N. Kolmogorov y V. A. Uspenski y el modelo de S. Marcus para el género¹⁴. Al decir que un modelo es paradigmático se entiende que en él predomina el análisis paradigmático, pero de ninguna manera que está desprovisto por completo de un análisis sintagmático. Una observación similar sirve asimismo para las demás clasificaciones de más arriba.

Mencionemos, por fin, la clasificación de los modelos en *algorítmicos* y *no-algorítmicos*, en tanto que se presenten como un proceso algorítmico o no (cf. cap. 4)¹⁵. Los modelos pueden ser también *constructivos* o *descriptivos*, según que pretendan construir o describir un objeto.

3.8. Ejemplos de modelos analíticos en gramática

El estudio de los modelos analíticos de la lengua se ha desarrollado especialmente en la Unión Soviética gracias a matemáticos como A. N. Kolmogorov, A. A. Liapunov, O. S. Kulaguina, V. A. Uspenski, G. S. Cheitin, R. L. Dobrushin, S. I. Fitialov, I. L. Bratchikov, A. V. Gladki y a lingüistas como N. D. Andreev, V. V. Ivanov, I. I. Revzin, entre otros. Para conocer las concepciones de estos estudiosos conviene presentar algunos conceptos preliminares.

Consideremos un conjunto E de elementos llamados palabras. A cada palabra $x \in E$ se le asocia una parte $\Gamma(x) \subset E$. Los elementos de $\Gamma(x)$ son *formas* de x ¹⁶. Todo conjunto finito y ordenado de elementos $a_1, a_2, \dots, a_i, \dots, a_n$, que pertenezca a E , se denomina *frase*, y se indica a base de $a_1 a_2 \dots a_i \dots a_n$. En el conjunto F de frases consideramos un subconjunto F_m . Todo elemento de F_m es una *frase marcada*. Dicho de un modo más simple, en función de frases marcadas se consideran, por lo común, las frases correctamente cons-

13. Cf. GREENBERG, "An axiomatization"; BATOG, "Logiczna rekonstrukcja" y "Critical remarks"; REVZIN, *Modeli*; KULAGUINA, "Ob odnom sposobe", y USPENSKI, "K opredelenin chasti rechi".

14. Cf. USPENSKI, "K opredelenin padezha", y MARCUS, "O analizä".

15. Cf. FITIALOV, "Formalno-matematicheskie modeli".

16. [Siendo x una palabra base, en tanto que elemento estable e identificable, las formas de x equivalen a las modificaciones (morfemas de género, número, tiempo, etc.) introducidas por las distintas alternativas sintácticas en el ámbito de la frase].

truidas a partir de una lengua natural dada, siendo E el vocabulario de la lengua en cuestión. Los ejemplos que ofreceremos de frases marcadas se referirán a esta interpretación. Una frase de la que se ha extraído una palabra, y sólo una, se denomina *frase unilacunar*, que indicamos así: $a_1 \dots a_{i-1} 0 a_{i+1} \dots a_n$, donde el signo "0" indica el lugar en que se ha eliminado una palabra. Una palabra b es *admisibile* para la frase unilacunar considerada si la sustitución de 0 por una forma determinada de b conduce a una frase marcada o, en otros términos, si existe $b' \in \Gamma(b)$ tal que la frase $a_1 \dots a_{i-1} b' a_{i+1} \dots a_n$ sea marcada. Así, p. ej., para la frase *Los hombres construyen — grandes*, la palabra *casa* es admisible en su forma *casas*. Una *frase unilacunar marcada* es, por definición, una frase unilacunar para la cual existe por lo menos una palabra admisible. Dos frases unilacunares f_1 y f_2 son *equivalentes con relación a una palabra b* si b es admisible, bajo la misma forma, tanto para f_1 como para f_2 . Por ejemplo, para las frases *Él — en la sala* y *Yo — en la sala*, el verbo *entrar* es admisible en la forma *entraba*. Así, pues, estas dos frases unilacunares son equivalentes en relación al verbo *entrar*¹⁷. Dos frases unilacunares f_1 y f_2 son *inmediatamente equivalentes* si existe al menos una palabra admisible para ambas y si f_1 y f_2 son equivalentes en relación con toda palabra admisible para ambas. Por ejemplo, las frases unilacunares *Estoy en una ciudad — y hermosa* y *Vivo en una aldea — y lejana* son inmediatamente equivalentes. Dos frases unilacunares f y φ son *absolutamente equivalentes* si existe un sistema finito de frases unilacunares

$$\psi_1, \psi_2, \dots, \psi_i, \dots, \psi_n$$

tal que:

- 1) para todo i ($1 \leq i < n$), ψ_i y ψ_{i+1} sean inmediatamente equivalentes;
- 2) $\psi_1 = f$;
- 3) $\psi_n = \varphi$.

La totalidad de las frases unilacunares se divide en clases disjuntas, de modo que dos frases de la misma clase sean absolutamente equivalentes y, en cambio, dos frases de clase distinta no sean absolutamente equivalentes. Estas clases se denominan *casos* (A. N. Kolmogorov).

Diremos que una palabra b es *posterior* a una palabra a ($a < b$) si en toda frase marcada que contenga a la sustitución de a por b produce una frase también marcada. Por ejemplo, la palabra *grande* es posterior a la palabra *pequeño*, ya que en toda frase correctamente construida que contenga *pequeño* la permutación de ésta por *grande* conduce a una frase también correctamente construida. Pero no así al contrario: en la frase *Tengo una casa grande*, la sus-

17. [Nótese que, en este ejemplo, *entraba* es la misma forma de una palabra base única].

titución de la última palabra por *pequeño* produce una frase ya no correctamente construida. *Pequeño*, pues, no es posterior a *grande*. Este ejemplo es típico para la relación de orden tal como la hemos definido más arriba. El hecho de que una palabra *b* sea posterior a una palabra *a* indica que en el paradigma de *b* la homonimia morfológica resulta más rica que en el paradigma de *a*. En el caso del adjetivo *grande* hay homonimia entre las formas de masculino y de femenino singular, mientras que esta homonimia falta en el paradigma de *pequeño*. Dados dos conjuntos de palabras *A* y *B*, diremos que el conjunto *B* es posterior al conjunto *A* y escribiremos $A < B$ si toda palabra de *B* es posterior a toda palabra de *A*. Diremos entonces de un conjunto *A* de palabras que es un conjunto inicial si no existe ningún conjunto no vacío *X* tal que $X < A$. La reunión de un conjunto inicial con todos sus conjuntos posteriores se denomina categoría gramatical elemental (R. L. Dobrushin).

I. I. Revzin ha comparado la definición de Dobrushin con la situación existente en las lenguas concretas y ha llegado a la conclusión de que debe modificarse la definición de “conjunto inicial” en el sentido siguiente: “un conjunto inicial es un conjunto que no es posterior a ninguna palabra que no le pertenezca”. Al mismo tiempo, Revzin ha mostrado que en las lenguas concretas existen algunas categorías gramaticales que no pueden obtenerse por el procedimiento de Dobrushin. Por ejemplo, el nominativo plural en alemán no está generado por ningún conjunto inicial. Ello plantea la necesidad de considerar categorías gramaticales generadas por conjuntos no iniciales¹⁸.

Consideremos una partición α de un conjunto de palabras *E* tal que $E = \bigcup_{i=1}^n A_i$ ¹⁹. Toda sucesión finita de conjuntos $A_{i_1}, A_{i_2}, \dots, A_{i_j}$ de la partición α se llama α -estructura. Denominaremos α -estructura de una frase $f = x_1 \dots x_i \dots x_n$ a la siguiente sucesión: $A(x_1) A(x_2) \dots A(x_i) \dots A(x_n)$, donde $A(x_i)$ es aquel conjunto a cuya partición pertenece la palabra x_i . La α -estructura de *f* la indicamos mediante $\alpha(f)$. Una α -estructura es marcada si existe al menos una frase marcada que la admita como α -estructura. Dos conjuntos A_i y A_j de la partición α son α -equivalentes si:

- 1) para toda α -estructura marcada de forma $\alpha(f_1) A_i \alpha(f_2)$, la α -estructura $\alpha(f_1) A_j \alpha(f_2)$ es también marcada;
- 2) para toda α -estructura marcada de forma $\alpha(\varphi_1) A_j \alpha(\varphi_2)$, la α -estructura $\alpha(\varphi_1) A_i \alpha(\varphi_2)$ es también marcada.

La relación de α -equivalencia divide los conjuntos de la partición en clases disjuntas. Consideremos, para cada clase de este tipo, la reunión de todos los

18. Cf. DOBRUSHIN, “Elementarnaia grammaticheskaia kategorija” y “Matematicheskie metody”, y REVZIN, “O nekotoryj voprosaj” y *Modeli*.

19. [Esta expresión se lee: “*E* es igual a la unión de todos los conjuntos A_i desde $i = 1$ hasta $i = n$ ”].

conjuntos de que aquélla está formada. Estas reuniones constituyen los miembros de una nueva partición del conjunto E : la partición derivada de la partición α , que se indica mediante α' , la cual admite, a su vez, una nueva partición derivada α'' , y así sucesivamente. Un teorema de O. S. Kulaguina postula que $\alpha' = \alpha''$, cualquiera que sea la partición α del conjunto E .

Consideremos ahora dos casos particulares de la partición α . Sea al principio la llamada partición unidad del conjunto E , es decir una partición tal de E en la cual cada término sea un conjunto formado por un elemento único. La partición derivada de la partición unidad se denominará S -partición del conjunto E . Y los conjuntos que constituyen los términos de la S -partición se llamarán familias de palabras. La noción de familia tiene un significado lingüístico especial. Si un conjunto A de palabras forman una familia, entonces $A < A$, donde $<$ es la relación de orden introducida antes. Dicho de otro modo, si dos palabras a y b pertenecen a la misma familia, entonces en toda frase correctamente construida que contenga a , la sustitución de a por b produce otra frase correctamente construida y viceversa, en toda frase correctamente construida que contenga b , la sustitución de b por a produce asimismo una frase correctamente construida. Así, p. ej., las palabras *casa* y *mesa* pertenecen a la misma familia, mientras que *casa* y *hombre* no pertenecen a la misma familia. El concepto de familia comprende, de un modo simplificado pero preciso, lo esencial en el llamado método de análisis distribucional o análisis de los ámbitos o contextos en que aparece un elemento dado. En él se inspira la lingüística descriptiva moderna desarrollada en los Estados Unidos por Z. S. Harris y H. A. Gleason, entre otros (cf. 1.12)²⁰.

Otra partición importante del conjunto E es la partición en vecindades, que indicaremos por Γ -partición del conjunto E . En términos más simples, diremos que una vecindad es la totalidad de palabras que tienen un mismo significado léxico y distintos significados gramaticales. Por ejemplo, las palabras rumanas *om* ('hombre'), *omului* ('del/al hombre'), *oamenii* ('hombres'), *oamenilor* ('de/a los hombres'), *omul* ('el hombre'), etc. pertenecen a la misma vecindad. A causa de la homonimia léxica parcial, la pertenencia a la misma vecindad no constituye una relación de equivalencia, de forma que la vecindad compuesta por las palabras *bandă* ('banda'), *bande* ('bandas'), *bandei* ('de/a la banda'), *bandelor* ('de/a las bandas'), etc. y la vecindad compuesta por las palabras *bandă* ('cinta'), *benzi* ('cintas'), *benzii* ('de/a la cinta'), etc. contienen palabras comunes, pero no coinciden. En lo que sigue haremos abstracción de este fenómeno y admitiremos que las distintas vecindades son siempre disjuntas, pues sólo en este sentido podremos operar con la partición en vecindades del conjunto E . La partición derivada de la partición en vecindades se denomina T -partición del conjunto E , y los conjuntos que determinan la

20. [Cf. HARRIS, *Structural Linguistics*, y GLEASON, *Introduction*].

T-partición, tipos. El concepto de tipo constituye una aproximación a lo que en lingüística se llama “parte del discurso”. Sin embargo, aquí se manifiesta una falta de concordancia entre la realidad lingüística y su modelo lógico-matemático. En tanto que en la lengua la pertenencia de dos palabras a la misma familia implica la pertenencia de estas palabras a la misma parte del discurso, en el modelo lógico, tal como ha demostrado V. A. Uspenski, no ocurre así²¹. Uspenski llama *lengua adecuada* (*pravilnyi iazyk*) a toda lengua (desde el punto de vista lógico-matemático, naturalmente) en que la pertenencia de dos palabras a la misma familia implica su pertenencia al mismo tipo. Por otra parte, I. I. Revzin ha introducido las siguientes nociones: dos palabras x e y son *R-equivalentes* si existe un sistema finito y ordenado de palabras $x_1, x_2, \dots, x_i, \dots, x_n$, tal que:

1) para todo $i (1 \leq i < n)$, las palabras x_i y x_{i+1} pertenezcan bien a la misma vecindad, bien a la misma familia;

2) $x_1 = x, x_n = y$.

El conjunto E de palabras, factorizado por la relación de *R*-equivalencia indicada antes, da lugar a un conjunto resultante, cuyos elementos se denominan *retículos*. Así, pues, dos palabras pertenecen a los mismos retículos si, y sólo si, son *R*-equivalentes. Un ejemplo de palabras pertenecientes al mismo retículo lo forman, en rumano, las palabras $x = scaun$ (“silla”) e $y = caietului$ (“del/al cuaderno”). En efecto, consideremos las palabras x_1, x_2 y x_3 , donde $x_1 = x, x_2 = caiet$ (“cuaderno”) y $x_3 = y$; *scaun* y *caiet* pertenecen a la misma familia, mientras que *caiet* y *caietului* pertenecen a la misma vecindad.

V. A. Uspenski ha demostrado que, para las lenguas adecuadas, de la pertenencia de dos palabras al mismo retículo resulta su pertenencia al mismo tipo. El propio Uspenski ha demostrado también que, para las lenguas adecuadas, la partición derivada de la partición en retículos es justamente la partición en tipos.

Las relaciones entre retículos y tipos varía de una lengua a otra. Así, como observa I. I. Revzin, en ruso un retículo formado por sustantivos es, al mismo tiempo, un tipo, ya que todos los sustantivos del mismo caso y que se hallan en plural forman una familia. Ello hace posible la unión, por medio de una cadena de la misma naturaleza que la utilizada en la definición del retículo, de dos sustantivos de género distinto. En cambio, en francés, dos sustantivos de género distinto pertenecen al mismo tipo, pero a retículos diferentes (I. I. Revzin).

En el mismo orden de ideas expuesto antes, V. A. Uspenski ha introducido el concepto de *lengua conforme* (*soglasovannym iazyk*). Una lengua se llama conforme si, indicando por medio de $\Gamma(u)$ la vecindad de la palabra u y

21. La noción de “familia” que aquí consideramos no guarda ninguna relación con las “familias de palabras” de la lingüística tradicional.

por medio de $S(u)$ la familia de la palabra u , a partir de $\Gamma(x) \cap S(y) \neq 0$ resulta $\Gamma(y) \cap S(x) \neq 0$. Tal como hace notar I. I. Revzin, la lengua rusa no es conforme, puesto que $\Gamma(stul) \cap S(lampy) = stulia$, mientras que $S(stul) \cap \Gamma(lampy) = 0$ ²². El mismo Revzin ha observado que la lengua rumana tampoco es conforme, debido, entre otras razones, a la existencia del género neutro en ella.

3.9. Modelación y lingüística estructural

La modelación matemática de la lengua se halla estrechamente ligada a la lingüística estructural y especialmente a la descriptiva. La lingüística estructural se encuentra en el origen de muchos modelos elaborados durante los últimos años: el modelo de los constituyentes inmediados de Chomsky fue precedido por la labor de Pike y el de Rulon Wells se compuso también a partir de un tema similar; el modelo transformacional de Chomsky tuvo otro antecedente en el análisis transformacional de Harris; los modelos proyectivos confeccionados por Lecerf, Ihm, Camion, Lynch y Hirschberg se hallan en germen ya en L. Tesnière, y así sucesivamente²³. El contacto entre la modelación matemática de la lengua y la lingüística estructural se manifiesta incluso en la interpretación de los modelos. Las llamadas vecindades características sobre las que habla Harris sugieren cómo enriquecer paulatinamente el conjunto de las frases marcadas y cómo seleccionar estas frases para que la interpretación del modelo sea lo más inspirativo y fecundo posible.

La importancia creciente que adquiere la modelación matemática sobre el plano del contenido también capacita la reactualización de algunas de las técnicas utilizadas en la glosemática.

No sólo la lingüística estructural, sino también la no estructural, ejerce una destacada influencia sobre la modelación matemática de la lengua. Muchos modelos se han elaborado con éxito a partir de sugerencias indirectas de la gramática intuitiva.

Parece incomprensible la actitud de algunos (p. ej., la de V. A. Zvegintsev) que, aun reconociendo la importancia práctica de la modelación matemática de

22. [Para la comprensión de este ejemplo, hay que tener en cuenta los siguientes datos: *stul*, 'silla', es nombre masculino singular y su plural es *stulia*; *lampy*, 'lámparas', es nombre femenino plural. Si se asigna, entonces, a los significados *st* = silla(s) y *la* = lámpara(s) unos valores convencionales tales como 0 para el masculino y 1 para el femenino, en la primera posición, junto con un valor 0 para el singular y 1 para el plural, en la segunda posición, puede escribirse: $stul =_{st} 00$; $stulia = st 01$, y $lampy = la 11$. Así, se ve que, efectivamente, $\Gamma(st 00) \cap S(la 11) = st 01$, mientras $S(st 00) \cap \Gamma(la 11) = \emptyset$, ya que tanto la vecindad de *stul* como la familia de *lampy* tienen en común a *stulia*, en tanto que la familia de *stul* y la vecindad de *lampy* no tienen ningún elemento común].

23. Cf. PIKE, "Taxemes"; WELLS, "Immediate constituents"; HARRIS, "Transformational model"; CAMION, "Analyse"; HIRSCHBERG-LYNCH, *Discussions*; LECERF-IHM, *Éléments*, y TESNIÈRE, *Éléments*.

la lengua, niegan su trascendencia para la lingüística²⁴. ¿Acaso esto no constituye una violación de la unidad entre la teoría y la práctica? ¿Cómo es posible que lo que es importante en la lingüística aplicada no deje ninguna huella en la teoría de la lengua?

Tal como se ha comprobado en los últimos años, la construcción de las lenguas lógicas de información y de las lenguas intermedias, realizada al principio con propósitos estrictamente aplicativos, también ha empezado a imponer una especial influencia en la investigación, sobre bases objetivas, de los aspectos semánticos de las lenguas naturales así como en la modelación matemática de estos mismos aspectos (cf. 7.2). Sobre esto hizo asimismo hincapié el lingüista V. V. Ivanov en su amplio comunicado sobre la lingüística matemática presentado en el IV Congreso de los Matemáticos Soviéticos, de 1961.

3.10. La dialéctica de la aproximación a través del modelo

La dialéctica de la aproximación, a través del modelo, sobre los fenómenos lingüísticos ofrece un interés particular. La contradicción dialéctica de esta aproximación consiste en el hecho de que, por una parte, un modelo debe ser simple a fin de prestarse lo mejor posible a una investigación deductiva, pero por otra parte un modelo debe ser complicado a fin de que registre la mayor parte de lo que es esencial en el fenómeno respectivo. G. S. Cheitín ha introducido el concepto de *c o n v e r g e n c i a* de una sucesión de modelos hacia su original²⁵. En una sucesión de este tipo, cada modelo está contenido en el siguiente, en el sentido de que toda propiedad obtenida por medio de un modelo de la sucesión puede ser obtenida también por medio del modelo siguiente y toda propiedad del original se obtiene por medio de un modelo de la sucesión. Cheitín aplicó este concepto sobre todo a los modelos generativos o algorítmicos. En tales condiciones, cabe decir que cada modelo de la sucesión es más adecuado al original que el modelo precedente. Y como el modelo es siempre susceptible de una investigación inaplicable al original, resulta que no puede existir identidad entre el modelo y el original, de modo que el modelo dejaría de ser ya modelo. Así, pues, dado un modelo A' de B , existe siempre un modelo A más adecuado a B que A . De ahí resulta que el modo con que Chomsky presenta sus tres famosos modelos no se halla en conformidad con la dialéctica de las relaciones entre un modelo y su original. Chomsky deja entrever que el modelo transformacional es adecuado en su totalidad. En rigor, lo que él denomina "modelo transformacional" representa una colección infinita de modelos correspondientes a una familia infinita de clases de transformaciones. El grado reduci-

24. Cf. ZVEGUINSEV, "Ce se înțelege".

25. Cf. CHEITIN, "K voprosu".

do de formalización que ofrecen las gramáticas transformacionales impide determinar su posición²⁶.

3.11. Los modelos generativos y su función explicativa

Es importante notar que la modelación matemática no sólo sirve para describir la lengua, sino también para explicar determinadas leyes internas de la lengua. En los modelos cualificados como “gramáticas generativas”, aparecidos en los últimos años, se explican los procedimientos internos por los que, partiendo de un inventario reducido de frases fundamentales y de una colección de reglas, se obtienen proposiciones y frases más complejas. Estos modelos permiten una comprensión más profunda del modo como el cerebro humano elabora frases a partir de las palabras conservadas en la memoria. La posibilidad de modelar matemáticamente este tipo de procesos implica la posibilidad de transmitir a las máquinas algunas de las actividades fundamentales relativas al lenguaje.

Haremos una presentación sucinta de los principales tipos de modelos generativos poniendo de relieve algunas de sus cualidades significativas. En el marco de estos modelos, se considera la lengua como un dispositivo cibernético que genera frases de la clase que normalmente denominamos correctas. La concepción más rudimentaria consiste en asimilar la gramática a un autómata finito. Se toma un vocabulario básico formado por un número finito de elementos (llamados símbolos o palabras). El autómata posee un número finito de estados, de modo que el paso de un estado a otro vaya acompañado por la producción de una palabra. El estado denominado inicial desempeña en el proceso un papel particular. Toda actividad del autómata empieza y termina en el estado inicial; por otro lado, el estado inicial no puede ser estado de transición. La totalidad de las sucesiones finitas de palabras (es decir de las frases) obtenidas de este modo constituye, por definición, la lengua generada por el autómata finito considerado. A su vez, las lenguas elaboradas así se llaman lenguas con un número finito de estados.

Sin duda, la sencillez y la comodidad que presenta este tipo de modelo son evidentes; se sabe sólo que la teoría de los autómatas abstractos es una teoría algebraica en pleno desarrollo, no excesivamente compleja y, por el momento, abundante en resultados. Para adquirir una idea sobre el grado de adecuación de las gramáticas con un número finito de estados en relación a las lenguas naturales, introduciremos de momento el concepto de *clase distribucional*. Llamaremos lengua sobre el vocabulario V a un determinado conjunto L de sucesiones finitas de elementos tomados de V . Las sucesiones de L se llamarán frases correctas, frente a las sucesiones finitas que no pertenezcan a la

26. Cf. también VASILIU, “Niveluri lingvistice”.

lengua L y que serán consideradas frases incorrectas. Así, p. ej., si V es el conjunto de las palabras españolas, entonces la lengua española es el conjunto L de todas las sucesiones finitas correctamente construidas y formadas a partir de elementos de V . Dadas dos construcciones (o sucesiones finitas de palabras), s_1 y s_2 , diremos que s_1 y s_2 se hallan en la misma clase distribucional si para toda pareja de frases, f_1 y f_2 , las frases $f_1 s_1 f_2$ y $f_1 s_2 f_2$ son o bien ambas correctas o bien ambas incorrectas. Tomando, p. ej., como L la lengua española, las sucesiones $s_1 = \text{muy guapo}$ y $s_2 = \text{guapo}$ se hallan en la misma clase distribucional, ya que en toda frase correcta que contenga s_1 la sustitución de s_1 por s_2 determina otra frase correcta, y en toda frase correcta que contenga s_2 la sustitución de s_2 por s_1 determina asimismo otra frase correcta²⁷. En cambio, las sucesiones s_1 y $s_3 = \text{hombre muy guapo}$ no se encuentran en la misma clase distribucional por cuanto en la frase correcta *He encontrado un hombre muy guapo*, la sustitución de s_3 por s_1 produce la expresión incorrecta **He encontrado un muy guapo*. La relación definida es, evidentemente, una relación de equivalencia en el conjunto de las frases (correctas o no). Las clases de equivalencia correspondientes son, por definición, clases distribucionales. Se puede demostrar que una lengua contiene un número finito de estados si, y sólo si, el número de las clases distribucionales es finito. Por otro lado, en las lenguas naturales existen ciertos procedimientos recursivos que determinan la existencia de una infinidad de clases distribucionales. Pues bien, justamente estos procedimientos marcan la incongruencia entre las lenguas con un número finito de estados y las lenguas naturales. He aquí un ejemplo de procedimiento recursivo: sea $f_1 = \text{Si vienes, entonces acepto}$; $f_2 = \text{Si } f_1, \text{ entonces acepto}$; ... $f_n = \text{Si } f_{n-1}, \text{ entonces acepto}$, y así sucesivamente. Un procedimiento de este tipo puede aparecer artificial, pero en matemática aparecen a menudo situaciones de este tipo cuando, p. ej., se afirma que si un cierto teorema es verdadero tiene lugar una cierta propiedad.

Un modelo más adecuado a las lenguas naturales que las gramáticas con un número finito de estados es el llamado modelo de los constituyentes inmediatos. En un modelo de este tipo se considera un vocabulario finito V , un conjunto finito Σ de sucesiones finitas de elementos de V y una colección finita F de reglas (órdenes) con la forma $X \rightarrow Y$ ("sustituir X por Y ", siendo X y Y sucesiones finitas de símbolos de V), de modo que un solo símbolo de X pueda ser sustituido en el paso $X \rightarrow Y$. Partiendo de un elemento de Σ , se aplica sucesivamente una serie de órdenes procedentes de F hasta que se llega a una sucesión que no admite ninguna otra orden de F . A esta sucesión se la denomina *t e r m i n a l*, y el conjunto de las sucesiones terminales constituyen una *l e n g u a t e r m i n a l*. Se puede demostrar que toda lengua con un número finito de estados es una lengua terminal, pero no viceversa. Las lenguas terminales

27. Hacemos abstracción de las formas de comparativo y superlativo del adjetivo *guapo*.

modelan mejor las lenguas naturales, ya que, p. ej., los procedimientos recursivos del tipo que hemos aducido antes se mantienen en el marco de una lengua terminal.

El modelo de los constituyentes inmediatos cubre sensiblemente el análisis dicotómico de una oración, sobre la base de las relaciones subordinativas, pero también deja al margen algunos aspectos esenciales. La posibilidad de distinguir entre una oración afirmativa y otra interrogativa, así como la posibilidad de distinguir entre el genitivo del sujeto (como en *la salida del sol*) y el genitivo del objeto (como en *la recolección del maíz*) sobrepasan las posibilidades del modelo de los constituyentes inmediatos. Por este y otros motivos, se ha adoptado un modelo con mayor cobertura, el llamado modelo transformacional, que corresponde, en cierto modo, a la máquina de Turing de la lógica matemática.

Tal como hemos visto en la exposición anterior, las gramáticas generativas explican la composición de las frases de las lenguas naturales estableciendo, para distintos tipos de frase, los tipos correspondientes de dispositivos cibernéticos que modelan el proceso respectivo de generación. Así se obtienen informaciones profundas sobre la naturaleza de las reglas que se hallan en la base de la composición de las frases más complejas a partir de las más simples. El concepto de *regla gramatical*, que durante mucho tiempo ha sido algo intuitivo y vago, se ha convertido en un concepto rigurosamente definido y estudiado por vía deductiva.

La gran importancia de los modelos generativos también reside en su carácter algoritmizable. Como se sabe, la modelación algorítmica es una condición preliminar esencial para que un proceso pueda ser programado en la máquina electrónica de cálculo. Se manifiesta aquí una estrecha interacción entre la función teórica, o explicación de los resortes profundos del proceso generativo de las frases, y la función aplicativa de los modelos generativos, que facilita directamente la composición de los algoritmos de traducción automática y elabora las lenguas de información. Evidentemente, la teoría lingüística no puede ignorar estas implicaciones, por las que la lingüística general se enfrenta a una problemática nueva que aún espera un tratamiento apto.

En contraste con las gramáticas transformacionales, que Chomsky considera adecuadas a las lenguas naturales y dotadas de poder explicativo con respecto a ellas, las gramáticas con un número finito de estados y las gramáticas de constituyentes inmediatos han sido calificadas de inadecuadas a las lenguas naturales. Esta situación ha alejado injustamente la atención de los lingüistas hacia los dos últimos tipos gramaticales. Aquí se ha dogmatizado el hecho de que los modelos transformacionales son más adecuados a las lenguas naturales. Adoptando una actitud practicista, las gramáticas con un número finito de estados son suficientes, por la sencilla razón de que toda colección finita de secuencias es una lengua con un número finito de estados²⁸. Sin embargo, la modelación mate-

mática de la lengua se propone como objetivo principal no el análisis directo de un fragmento concreto de una lengua natural, sino la explicación de los conceptos básicos de la lingüística y la resolución de algunos problemas relativos a la estructura de la lengua. Sólo indirectamente, por medio de estas ventajas teóricas, se obtienen también facilidades en el dominio de la lingüística aplicada donde, desde luego, puede a veces aceptarse el supuesto de que la lengua es finita.

Una vez establecido en qué consiste la incongruencia entre las lenguas con un número finito de estados y las lenguas naturales, podemos encontrar porciones infinitas de las lenguas naturales que son *sublenguas* con un número finito de estados; podemos, al mismo tiempo, plantear el problema de obtener una gramática natural como límite de una sucesión convergente de gramáticas con un número finito de estados. El hecho de que una gramática con un número finito de estados opere como un autómata finito aumenta todavía más la importancia lingüística de estas gramáticas, tan poco estudiadas hasta ahora desde el punto de vista lingüístico²⁹. El autómata finito está reconocido como un modelo incomparablemente más cómodo y más simple que las máquinas de Turing asociadas a las gramáticas de constituyentes inmediatos³⁰. Esta situación también confirma el interés lingüístico de las gramáticas con un número finito de estados.

La analogía entre las gramáticas de constituyentes inmediatos y los algoritmos normales, en el sentido de Markov (cf. cap. 4, para la definición de algoritmo normal), es evidente. Sin embargo, la relación entre ambas nociones no está aún bien determinada.

La incorporación de los modelos de Chomsky al proceso dinámico de aproximación a las lenguas naturales por medio de una sucesión infinita de modelos también conducirá, sin duda, a una reevaluación de su poder explicativo³¹.

3.12. La jerarquía de Chomsky

Una gramática generativa tal como la entiende Chomsky es un sistema G formado por cuatro objetos $\langle V_T, V_N, A, \mathcal{R} \rangle$, donde V_T es un conjunto finito llamado *vocabulario terminal*, V es un conjunto finito llamado *vocabulario no-terminal*, V_T y V_N contienen por lo menos un elemento cada uno y no tienen elementos comunes, A es un elemento determinado de V_N llamado *axioma de la gramática*, y \mathcal{R} es una colección fini-

28. Cf. CHOMSKY-MILLER, "Finite state languages", y MARCUS, "Gramatici și automate".

29. Cf. BAR-HILLEL y SHAMIR, "Finite-state languages".

30. Cf. RABIN-SCOTT, "Finite automata".

31. [Cf. BAEZ, *Introducción*; NIVETTE, *Principios*, y QUESADA, *La lingüística*].

ta de reglas de forma $x \rightarrow y$, donde x y y son sucesiones finitas de elementos del vocabulario total (definido como la reunión del vocabulario terminal y del vocabulario no-terminal). Los elementos de V_T se denominan símbolos terminales, y los de V_N , símbolos no-terminales o auxiliares. Toda sucesión finita de elementos tomados de V_T se llama frase terminal. Una sucesión finita de elementos sobre un determinado vocabulario V se denomina frase sobre el vocabulario V . Una frase sobre V tendrá, entonces, la forma $x = a_1 a_2 \dots a_i a_{i+1} \dots a_n$, donde a pertenece a V para todo valor de i comprendido entre 1 y n . Al número n se le llama aquí extensión de la frase x . También se conviene en llamar frase nula, y se nota por medio de ω , a la frase igual a cero. Toda colección (finita o infinita) de frases sobre el vocabulario V constituye un lenguaje sobre el vocabulario V . Un par ordenado $\langle u, v \rangle$ de frases sobre V constituye un contexto sobre el vocabulario V . Una regla $x \rightarrow y$ en que y sea una frase sobre V_T se llama regla terminal. Una regla que no sea terminal se denomina no-terminal.

Diremos que la frase z deriva directamente, por medio de G , de la frase w si existe una regla $x \rightarrow y$ en \mathcal{R} y un contexto $\langle u, v \rangle$ tales que $w = u x v$ y $z = u y v$. Diremos asimismo que la frase z deriva, por medio de G , de la frase w si existe una sucesión finita de frases $t_1, t_2, \dots, t_i, t_{i+1}, \dots, t_n$ tal que $t_1 = w$, $t_n = z$ y que t_{i+1} derive directamente de t_i para cualquier valor de i desde 1 hasta $n - 1$. Definiremos el lenguaje $L(G)$ generado por la gramática G como la totalidad de las frases terminales que deriven del axioma A de la gramática G .

En lo que sigue, notaremos por medio de $w \xrightarrow{A} z$ el hecho de que la frase z derive directamente de la frase w por aplicación de la regla que tenga el número de orden n en la gramática G (en la hipótesis de que las reglas de G vayan enumeradas).

Una regla $x \rightarrow y$ de \mathcal{R} es independiente de contexto si la frase x coincide con un símbolo no-terminal y la frase y no tiene una extensión nula. Una regla $x \rightarrow y$ es dependiente de contexto si la extensión de la frase x es positiva, pero no sobrepasa la extensión de la frase y . Está claro, pues, que toda regla independiente de contexto es dependiente de contexto (aunque la terminología es aquí paradójica, se ha consagrado y hay que respetarla). Una regla independiente de contexto $B \rightarrow y$ es una regla con autoinclusión si el símbolo no-terminal B de la izquierda reaparece a la derecha, en la expresión de y . La presencia, en una gramática independiente de contexto, al menos de una regla con autoinclusión asegura la infinitud del lenguaje generado por la gramática en cuestión. Una regla $x \rightarrow y$ independiente de contexto es simple a la derecha si $y = Ba$, donde $B \in V_N$ y $a \in V_T$. La regla $x \rightarrow y$ independiente de contexto es simple a la izquierda si $y = aB$, donde $B \in V_N$, $a \in V_T$. Se dice de una gramática G que es regu-la-

da o con un número finito de estados si tiene lugar la siguiente alternativa: 1) todas las reglas no-terminales son simples a la derecha; o bien 2) todas las reglas no-terminales son simples a la izquierda. Además las reglas terminales presentan la forma $C \rightarrow c$ (donde $C \in V_N$, $c \in V_T$). Se dice de una gramática G que es dependiente (y, respectivamente, independiente) de contexto si todas sus reglas son dependientes (y, respectivamente, independientes) de contexto. Resulta, así, que toda gramática con un número finito de estados es independiente de contexto y que toda gramática independiente de contexto es dependiente de contexto.

Un lenguaje se denomina regulado o con un número finito de estados si puede ser generado por una gramática con un número finito de estados. Un lenguaje se denomina independiente de contexto si puede ser generado por una gramática independiente de contexto. Está, entonces, claro que todo lenguaje con un número finito de estados es independiente de contexto y que todo lenguaje independiente de contexto es dependiente de contexto. Cabe demostrar que existen lenguajes independientes de contexto pero no con un número finito de estados y lenguajes dependientes de contexto pero no independientes de contexto.

Las gramáticas generativas consideradas hasta ahora no imponen ninguna restricción en cuanto al orden en que deben ser aplicadas las reglas. Una posibilidad de introducir un orden de este tipo la ofrecen las gramáticas matriciales. A cada tipo de gramática chomskiana se le puede asociar un tipo correspondiente de gramática matricial. En lo que sigue, sólo tomaremos en cuenta las gramáticas matriciales asociadas a las gramáticas independientes de contexto.

Sea $G = \langle V_T, V_N, A, \mathcal{R} \rangle$ una gramática independiente de contexto. Llamaremos matriz (de orden m) sobre G ($m \geq 1$) a todo conjunto ordenado $\{R_1, R_2, \dots, R_m\}$ de reglas R_i de \mathcal{R} ($1 \leq i \leq m$). Una gramática matricial GM asociada a G es, por definición, un conjunto finito de matrices sobre G . Diremos que y deriva directamente de x a través de GM si existe una matriz $M = \{R_1, R_2, \dots, R_m\}$ de GM y una sucesión de frases z_1, \dots, z_m sobre el vocabulario total tales que

$$x \xrightarrow{R_1} z_1 \xrightarrow{R_2} z_2 \longrightarrow \dots \xrightarrow{R_m} z_m = y.$$

En este caso escribiremos

$$x \xrightarrow{M} y.$$

Diremos que una frase y deriva de la frase x a través de GM si existe una sucesión de matrices M_1, M_2, \dots, M_k de GM y una sucesión z_1, \dots, z_k de frases sobre el vocabulario total tales que

$$x \xrightarrow{M_1} z_1 \xrightarrow{M_2} z_2 \implies \dots \xrightarrow{M_k} z_k = y.$$

En este caso escribiremos $x \xrightarrow{GM} y$.

El lenguaje $L(GM)$ generado de GM será, por definición, el conjunto de las frases terminales (es decir, sobre el vocabulario V_T) que deriva del axioma A por medio de la gramática GM .

Se dice de un lenguaje que es **m a t r i c i a l** si existe una gramática matricial que lo genere.

No es difícil ver que toda gramática independiente de contexto es una gramática matricial de un determinado tipo, en el cual cada matriz está formada por una sola regla y a cada regla le corresponde una matriz. En este caso particular, el lenguaje generado por la gramática matricial coincide con el lenguaje generado por la gramática independiente de contexto a la cual ésta se halla asociada. Resulta que todo lenguaje independiente de contexto es un lenguaje matricial. Existen, sin embargo, lenguajes matriciales que no son independientes de contexto (p. ej., $\{a^n b^n c^n\} \ 1 \leq n < \infty$).

Todo lenguaje matricial es un lenguaje dependiente de contexto, pero existen lenguajes dependientes de contexto que no son lenguajes matriciales. Resulta entonces que la clase de los lenguajes matriciales es estrictamente intermedia entre la clase de los lenguajes independientes de contexto y la de los lenguajes dependientes de contexto. En este punto reside el especial interés que ofrece esta clase de lenguajes, comprendida entre una clase demasiado restrictiva y una clase demasiado amplia.

Como trataremos con lenguajes no independientes de contexto, hemos de hallar un criterio por el que podamos elucidar con relativa facilidad la pertenencia de un lenguaje a la clase de lenguajes que no son independientes de contexto. El siguiente teorema nos provee de un criterio de este tipo:

“Si L es un lenguaje infinito e independiente de contexto sobre el vocabulario V , entonces existen dos números naturales p y q tales que, para toda frase x de L de extensión más grande que p , tengamos una representación de forma

$$x = y_1 z_1 w z_2 y_2$$

donde y_1, z_1, w, z_2 e y_2 son frases sobre el vocabulario V , la frase w es no-nula, al menos una de las frases z_1, z_2 es no-nula, la extensión de la frase $z_1 w z_2$ es menor o igual que q , y la frase de forma

$$y_1 z_1^n w z_2^n y_2$$

pertenece al lenguaje L cualquiera que sea el número natural n .”

Tal como se observa, este criterio da una condición necesaria para que un lenguaje sea independiente de contexto. Tal criterio se aplica en su forma negativa, es decir para determinar los lenguajes que no lo satisfacen y, por tanto, no son independientes de contexto.

3.13. La modelación de las relaciones subordinativas

El orden en que se encuentran las palabras en el discurso o en un texto escrito es lineal; las palabras se sitúan formando una sucesión en la cual cada una de ellas tiene un predecesor inmediato y un sucesor inmediato, a excepción de la primera y la última que carecen, respectivamente, de predecesor y sucesor. Así, en la secuencia *Veo muy pocas cosas pequeñas*, las palabras están colocadas en orden lineal, de izquierda a derecha. Si investigamos las relaciones existentes entre los términos de esta proposición, comprobamos que al margen de las relaciones de contigüidad existen otras que conciernen más de cerca a la estructura de la proposición considerada. De este modo, *pequeñas* se halla en estrecha relación con *cosas* no sólo porque es contigua a ella, sino porque la explica y completa su sentido; dicho de otra manera, *pequeñas* está subordinada a *cosas*. Por otra parte, *cosas* está, en el mismo sentido, subordinada a *veo*, como objeto directo, a pesar de que *cosas* y *veo* no son contiguas (desde el punto de vista del orden lineal). A base de consideraciones similares, llegamos a la conclusión de que *pocas* está subordinada a *cosas*, y *muy* a *pocas*.

Las relaciones que acabamos de establecer se fundan en ciertas consideraciones relativas a los significados de los distintos términos de la proposición y a las relaciones entre estos significados. Podemos acceder a un grado superior de formalización y de objetivación describiendo la misma estructura proposicional exclusivamente a base de las relaciones mutuas que entablan sus términos, en la hipótesis de que podamos distinguir una secuencia correcta de otra incorrecta. Diremos que en una proposición correcta un término *A* está subordinado a un término *B* si la supresión de *A* con la conservación de *B* produce otra proposición correcta. (Esta definición corresponde, la mayoría de las veces, al concepto intuitivo de subordinación.) De este modo, en la proposición anterior, *pequeñas* está subordinada a *cosas*, por cuanto la proposición *Veo muy pocas cosas* es correcta; *muy pocas* está subordinada a *cosas pequeñas*, porque la proposición *Veo cosas pequeñas* es correcta; *muy* está subordinada a *pocas*, porque la proposición *Veo pocas cosas pequeñas* es correcta. En esta última proposición, las palabras *pocas* y *pequeñas* están subordinadas a *cosas*, mientras que, en *Veo cosas*, *cosas* está subordinada a *veo*. Todas estas relaciones subordinativas pueden presentarse de un modo más sugestivo por medio del siguiente esquema:

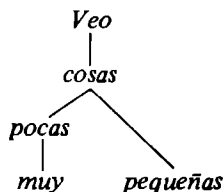


Figura 6

que define el orden estructural, “bidimensional”, de los términos de la proposición considerada. Cuando leemos u oímos una proposición, nuestro pensamiento la elabora convirtiendo su orden lineal en su orden estructural, el único que nos permite determinar el sentido de la proposición; y cuando [escribimos o] hablamos, el pensamiento dispone el orden estructural en un orden lineal. El proceso por el que se pasa de una lengua *A* a una lengua *B* puede realizarse bien a base de convertir directamente el orden lineal de la lengua *A* al orden lineal de la lengua *B* (en este caso estamos ante una traducción “palabra por palabra”), o bien a base de pasar del orden lineal de *A* al orden estructural de la misma *A* para luego pasar del orden estructural de *A* al orden estructural de *B* y, por fin, del orden estructural de *B* al orden lineal de la misma *B*. Esta última es la traducción propiamente dicha, basada en el significado. Una de las diferencias fundamentales entre el orden lineal y la estructura relacional consiste en que, mientras en el orden lineal cada término está vinculado como máximo a otros dos términos (el anterior y el posterior), en el aspecto relacional existe la posibilidad de que un término esté vinculado con tres o incluso más términos. Este tipo de fenómenos, muy frecuentes en cualquier lengua, hacen que el orden lineal sacrifique una parte del orden estructural. Es deseable que dos términos contiguos en el orden lineal sean también “contiguos” en el orden estructural; pero existirán términos estructuralmente contiguos que no aparecerán juntos en la proposición. Este fenómeno se halla estrechamente ligado a otro: el paso del orden estructural al lineal o del lineal al estructural no siempre permite mantener inalterable el orden. Así, en el análisis proposicional de más arriba, *cosas* está comprendida entre *pocas* y *veo* desde el punto de vista del orden estructural, pero no en cambio desde el punto de vista del orden lineal.

Consideremos, por otra parte, la frase *El hombre consigue progresos continuos*. Aquí, *progresos* está subordinada a *consigue*, mientras que *continuos* lo está a *progresos*, por lo que *progresos* se halla comprendida entre *consigue* y *continuos* tanto desde el punto de vista del orden estructural como desde el punto de vista lineal.

Consideremos ahora la proposición *El alumno muy aplicado aprende bien las lecciones*. Entre sus términos cabe observar las siguientes relaciones subordinativas:

El alumno muy aplicado aprende bien las lecciones (0)

Adoptando los modos de expresión de la gramática elemental, podemos decir que *muy* explica, completa el significado de *aplicado*, que *aplicado* completa el significado de *alumno* y que *lecciones* completa, a su vez, el significado de *aprende*. Sin embargo, entre estas relaciones subordinativas existen diferencias importantes. En efecto, por lo que se refiere a la primera relación de subordina-

ción, la de *muy* respecto de *aplicado*, observamos que admite una presentación como la de más arriba en tanto que se remite al hecho de que la supresión de *muy* no altera la corrección gramatical de la proposición, mientras que la supresión de *aplicado* sí la altera. Una situación semejante acontece en la subordinación de *bien* con respecto de *aprende*.

No cabe decir, en cambio, lo mismo acerca de la subordinación de *aplicado* con relación a *alumno*, ya que la supresión de *aplicado* altera la corrección gramatical del resultado.

Consideremos ahora la proposición *El alumno lee un libro muy interesante*. Aquí se producen las siguientes relaciones subordinativas:

El alumno lee un libro muy interesante (1)

En toda relación subordinativa distinguimos un término *regente* y un término *dependiente*. Ambos forman un grupo que llamaremos *sintagma*. Si la supresión del término dependiente no altera la corrección gramatical, diremos entonces que se trata de una *subordinación de primer rango*. En este caso, el término regente constituye la *resultante* o el *núcleo* del sintagma y el término dependiente, el *satélite*. En la proposición aducida, sólo una de las tres relaciones subordinativas es de primer rango: la subordinación de *muy* respecto de *interesante*, ya que si sustituimos el sintagma *muy interesante* por su resultante, obtenemos la proposición

El alumno lee un libro interesante (2)

La subordinación de *interesante* respecto de *un libro*, que evidentemente es de primer rango con relación a la proposición (2), se llamará *subordinación de segundo rango* con respecto a la proposición inicial (1). Como la supresión, en (2), de *un libro* altera la corrección gramatical (se obtiene **El alumno lee interesante*), resulta que la subordinación de *un libro* con respecto a *lee* no es tampoco de segundo rango. Si sustituimos, en (2), el sintagma *un libro interesante* por su resultante, obtenemos la proposición

El alumno lee un libro (3)

En relación con (3), la subordinación de *libro* respecto de *lee* es de primer rango; convenimos en considerar que este tipo de subordinación es de tercer rango respecto de (1).

3.14. Configuraciones

Estos análisis revelan una suerte de jerarquía en las relaciones subordinativas. Dicha jerarquía es susceptible de una modelación matemática gracias al concepto de configuración propuesto por O. S. Kulagina y estudiado por I. I. Revzin y A. V. Gladki. He aquí una presentación sumaria y simplificada de este modelo.

Sea V un conjunto finito denominado vocabulario. Toda sucesión finita de elementos de V se llama *frase*. El número de términos de una frase constituye la *extensión de la frase*. Distinguiremos una determinada colección de frases, a las que llamaremos *marcadas*. Toda frase se denominará *frase de rango 0*.

Sea F una colección de frases y sean f y g dos frases arbitrariamente elegidas. Diremos que f y g tienen la *misma distribución con respecto a F* si para todo par de frases φ y ψ tales que las frases $u = \varphi f \psi$ y $v = \varphi g \psi$ pertenezcan a F , tiene lugar la siguiente situación: o bien u y v son ambas marcadas, o bien u y v son ambas no-marcadas.

Sea F_0 el conjunto de las frases de rango 0. Diremos que una frase f es una *configuración de primer rango* si su extensión es al menos igual a 2 y si existe una frase x de extensión igual a 1 tal que f y x tengan la misma distribución con respecto a F_0 . (Está claro que $x \in V$). El elemento x constituye una *resultante de la configuración f* .

Diremos que una frase es de *primer rango* si no contiene ninguna configuración de primer rango. Notemos por medio de F_1 el conjunto de las frases de primer rango. Diremos que una frase f es una *configuración de segundo rango* si su extensión es al menos igual a 2 y si existe una frase x de extensión igual a 1 tal que f y x tengan la misma distribución con respecto a F_1 . El elemento x constituye una *resultante de la configuración f* . Una frase que no contenga ninguna configuración de segundo rango será, por definición, una *frase de segundo rango*.

Sea n un número natural. Supongamos que hemos definido las frases y las configuraciones de cualquier rango $\leq n$. Diremos de una frase f que es una *configuración de rango $n + 1$* si su extensión es al menos igual a 2 y si existe una frase x de extensión igual a 1 tal que f y x tengan la misma distribución con respecto a F_n , donde por F_n notamos el conjunto de las frases de rango n . Una frase que no contenga ninguna configuración de rango $\leq n + 1$ será, por definición, una *frase de rango $n + 1$* .

Toda configuración de rango n es también de cualquier rango mayor que n . El menor rango de una configuración constituye su *orden*. Toda frase de rango n es también de cualquier rango menor que n . El mayor rango de una frase (si es que existe este tipo de rango) constituye su *orden*.

Si tomamos el vocabulario de la lengua española en función de V y las frases correctas del español en función de las frases marcadas, entonces el sintagma $f = \text{muy interesante}$ es una configuración de primer rango. En efecto, presenta una extensión igual a 2 y existe una frase x de extensión igual a 1 (p. ej., $x = \text{interesante}$) tal que f y x tengan la misma distribución con respecto a F_0 . La frase (1) de más arriba es —como toda frase— de rango 0, pero no es de primer rango porque contiene la configuración de primer rango *muy interesante*. Sustituyendo en (0) *muy aplicado* por *aplicado*, y *aprende bien* por *aprende*, y, en (1), *muy interesante* por *interesante*, las nuevas frases obtenidas serán de primer rango. Éste es el método general por el que se obtienen frases de primer rango. Así, pues, la frase (2) es de primer rango. El sintagma *un libro interesante* es una configuración de segundo rango porque *un libro interesante e interesante* tienen la misma distribución con respecto al conjunto de las frases que no contienen configuraciones de primer rango. Por otra parte, *un libro interesante* no es una configuración de primer rango, porque no existe ninguna frase x de extensión igual a 1 por la que pueda sustituirse *un libro interesante* en cualquier frase de rango 0 de tal modo que se conserve la corrección gramatical. En lo que concierne al sintagma *lee un libro*, se trata de una configuración de tercer rango, puesto que en la frase (2), que es de primer rango, este sintagma no puede ser sustituido por una palabra única de tal modo que se conserve la corrección gramatical.

3.15. Relaciones de dependencia

Definiremos ahora una relación de dependencia, que será más débil y más general que la de subordinación, en el sentido de que todo vínculo de subordinación será también un vínculo de dependencia, pero no viceversa.

Sea $f = x_1 x_2 \dots x_i \dots x_n$ una frase sobre el vocabulario V . Diremos que x_j es *d e p e n d i e n t e* de x_k ($1 \leq j \leq n$, $1 \leq k \leq n$, $j \neq k$) si existe una sucesión finita y_1, y_2, \dots, y_s de términos de la frase f con las siguientes propiedades: 1) $y_1 = x_j$; 2) $y_s = x_k$; 3) para todo número natural m tal que $1 \leq m \leq s$ existe un número natural p_m tal, que y_{m+1} sea resultante de una configuración de orden p_m que contenga a y_m . Si, además, ocurre que $p_1 \leq p_2 \leq \dots \leq p_{s-1}$, entonces decimos que la dependencia de x_j respecto de x_k es *m o n ó t o n a*. En caso contrario, la dependencia es *n o - m o n ó t o n a*.

El menor número $s-1$ para el cual exista una sucesión y_1, y_2, \dots, y_s con las propiedades 1), 2) y 3) se denomina *g r a d o d e d e p e n d e n c i a d e x_j c o n r e s p e c t o a x_k* . La dependencia de primer grado se denomina *s u b o r d i n a c i ó n*.

Para ilustrar estos conceptos nos referiremos a la frase (1) de más arriba, donde probaremos que *muy* es dependiente de *un libro*. En efecto, *interesante*

es la resultante en la configuración de primer rango *muy interesante*, que contiene a *muy*, mientras que *un libro* es la resultante en la configuración de segundo rango *un libro interesante*, que contiene a *interesante*. Hemos demostrado así que la dependencia de *muy* con respecto a *libro* es una dependencia monótona, de segundo grado. Teniendo en cuenta que, en la misma frase (1), *un libro* es la resultante en la configuración de tercer orden *lee un libro*, deducimos que, en la frase (1), *muy* se halla en dependencia monótona de tercer grado con respecto a *lee*.

3.16. La condición de proyectividad

Consideraremos ahora dos situaciones que presentan una importancia particular tanto para la teoría de la gramática como para la lingüística aplicada.

α) Sea una frase $f = x_1 x_2 \dots x_i \dots x_n$. Si para todo par i, j ($1 \leq i \leq n$, $1 \leq j \leq n$, $i \neq j$) para el cual x_i está subordinado a x_j tiene lugar también la subordinación de x_k con respecto a x_j , cualquiera que sea k comprendido entre i y j , entonces decimos que la frase f es una frase proyectiva. Con una variante distinta y semiformalizada, esta noción fue introducida por Y. Lecerf y P. Ihm³². Se ve fácilmente que las frases (0) y (1) no son proyectivas. En efecto, la condición de proyectividad se reduce al hecho de que si un término x de la frase está subordinado a un término y , entonces todo término comprendido entre x e y está igualmente subordinado a y . Pero en la frase (0), pese a que *aplicado* está subordinado a *alumno*, la palabra *muy*, comprendida entre *alumno* y *aplicado*, no está subordinada a *alumno*. Y en la frase (1), aunque *interesante* está subordinado a *libro*, la palabra *muy*, comprendida entre *libro* e *interesante*, no está subordinada a *libro*. He aquí un ejemplo de frase proyectiva:

*El alumno aplicado prepara bien las lecciones difíciles*³³.

Una lengua en que todas las frases sean proyectivas se denomina lengua proyectiva. Y de la gramática de una lengua proyectiva decimos que se trata de una gramática proyectiva.

β) Sea una frase $f = x_1 x_2 \dots x_i \dots x_n$. Si para todo par i, j ($1 \leq i \leq n$, $1 \leq j \leq n$, $i \neq j$) para el cual x_i sea dependiente de x_j tiene lugar también la dependencia de x_k con respecto a x_j , cualquiera que sea k comprendido entre i y j , entonces decimos que la frase f es una frase cuasiproyectiva. Se

32. Cf. LECERF-IHM, *Éléments*.

33. [En todos estos ejemplos, consideramos que el artículo forma un bloque común con su determinante, tal como lo sugiere el original rumano].

puede demostrar que toda frase proyectiva es, al mismo tiempo, una frase cuasi-proyectiva. Lo recíproco, en cambio, no es verdadero; efectivamente, la frase (0), a pesar de no ser proyectiva, es cuasiproyectiva: *muy* se halla en dependencia de segundo grado con respecto a *alumno*.

Una lengua en que toda frase sea cuasiproyectiva constituye, por definición, una lengua cuasiproyectiva, y su gramática es una gramática cuasiproyectiva.

Se han postulado diversos teoremas para caracterizar las lenguas proyectivas. Más aún, se han construido incluso modelos matemáticos más adecuados a las lenguas naturales que los proyectivos y que, por tanto, registran ciertas situaciones que se apartan de la condición de proyectividad. A este orden de ideas se inscribe también la condición de cuasiproyectividad. Dichos modelos ofrecen una importancia singular para la lingüística aplicada, por cuanto la condición de proyectividad simplifica esencialmente la composición de los algoritmos de traducción en la máquina electrónica de cálculo.

Las lenguas proyectivas han podido estudiarse por vía deductiva y con ello se ha allegado una inestimable información sobre las diferencias estructurales que existen entre una lengua natural y una lengua proyectiva (cf. 3.28).

Los sistemas descriptivos que tienen en cuenta las relaciones, tal como han sido presentados por la lingüística tradicional y sobre todo por la lingüística estructural, han desempeñado un papel intermedio, pero esencial en la construcción del modelo matemático de las gramáticas proyectivas.

3.17. La hipótesis de Yngve

Convengamos en representar en forma arbórea el análisis configuracional de una frase. Así, p. ej., si representamos por medio de *f* la proposición (1), obtenemos entonces:

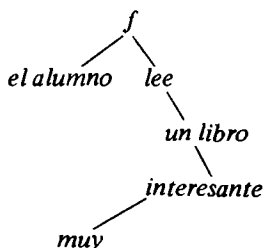


Figura 7

Los principios de representación son los siguientes: cuanto más pequeño es el orden de una configuración, tanto más baja se halla ésta en el diagrama ar-

bóreo. El término resultante de una configuración está situado encima y el término subordinado, debajo, a la derecha o a la izquierda según que la posición del término resultante esté a la izquierda o a la derecha, en la configuración. Las configuraciones con el término resultante a la derecha se denominan *configuraciones regresivas* y, a su vez, las configuraciones con el término resultante a la izquierda se denominan *configuraciones progresivas*.

En un artículo especialmente interesante, Víctor H. Yngve pone de manifiesto el carácter asimétrico de los diagramas arbóreos que describen el análisis en constituyentes inmediatos de las proposiciones en lengua inglesa³⁴. En tanto que las ramificaciones regresivas (orientadas hacia la izquierda) comportan, en cada etapa, una carga temporal de la memoria (del hombre o de la máquina) con un símbolo suplementario, las ramificaciones progresivas (orientadas hacia la izquierda) se hallan libres de toda restricción. Yngve emite la hipótesis según la cual, en inglés, una ramificación regresiva no puede comportar más de siete etapas, y considera que la validez de esta hipótesis es igualmente plausible para las demás lenguas naturales. Con respecto a estas consideraciones, Yngve introduce el concepto de *profundidad* de una estructura regresiva como equivalente al número de las ramificaciones de la estructura respectiva que implican una carga memorística. Así, la estructura regresiva *Very clearly projected pictures appeared* ('Aparecieron imágenes proyectadas muy claramente') tiene una profundidad igual a cuatro y corresponde al siguiente esquema:

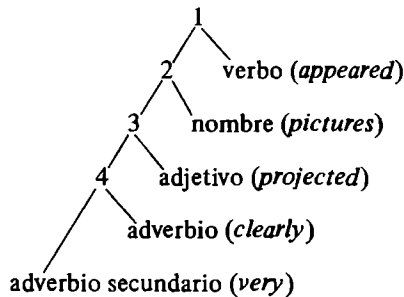


Figura 8

La pérdida en la memoria del verbo *appeared* ('aparecieron') determina que la expresión *Very clearly projected pictures* ('imágenes proyectadas muy claramente') no sea ya una proposición; la pérdida en la memoria de *pictures* ('imágenes') determina una expresión **Very clearly projected* ('proyectado muy claramente') ya no correcta, y así sucesivamente. En cambio, las estructuras

34. Cf. YNGVE, "A model".

progresivas pueden ser ilimitadamente largas. De este modo, *El alumno escribe*, *El alumno diligente escribe*, *El alumno diligente escribe bien los temas*, *El alumno diligente de nuestra clase escribe bien los temas pedidos en el examen*, etc. son estructuras progresivas cada vez más largas.

Yngve llama la atención sobre el hecho de que las notaciones matemáticas y de la lógica simbólica, en tanto que son lenguas escritas, no comportan la asimetría estructural indicada antes, la cual constituye una propiedad de las lenguas habladas.

3.18. El cálculo de los tipos sintácticos

En un ciclo de artículos, J. Lambek plantea el siguiente problema, singularmente interesante: así como en la física clásica la exactitud de una igualdad se verificaba mediante la comparación de los dos miembros implicados desde el punto de vista de sus dimensiones, ¿no es posible asociar a las palabras de una lengua natural ciertos tipos, de modo que la corrección de una frase pueda determinarse mediante un cálculo efectuado sobre los tipos?³⁵ Estos tipos primarios, asociados a unas palabras, deberían desempeñar aquí el papel que L (la longitud), T (el tiempo) y M (la masa) desempeñan en la física. Mediante un cálculo aplicado a los tipos primarios, deberían obtenerse determinados tipos derivados, del mismo modo que a partir de las magnitudes primitivas L , M y T se obtienen determinadas magnitudes derivadas: L^2MT , LM^2T^{-1} , etcétera. La analogía se detiene aquí, ya que, como veremos, frente a las magnitudes obtenidas a base de L , M y T , que forman un grupo abeliano (con tres generadores: L , M y T), los tipos obtenidos en el cálculo sintáctico asociativo no forman un grupo abeliano³⁶.

J. Lambek construye en principio un llamado cálculo sintáctico asociativo tal como sigue: se parte atribuyendo a las palabras y a las cadenas de palabras ciertos tipos primitivos. De esta manera, si nos referimos al inglés, todas las proposiciones correctamente formadas tienen el tipo s ; todos los nombres y grupos nominales, el tipo n ; todos los infinitivos de los verbos

35. Cf. LAMBEK, "Mathematics", "Contributions" y "Calculus".

36. [Por "grupo" se entiende todo conjunto de elementos que verifican las siguientes leyes:

- 1) $(x \star y) \star z = x \star (y \star z)$;
- 2) Existe un 0 tal que: $x \star 0 \star x = x$;
- 3) Existe un x' tal que: $x \star x' = x' \star x = 0$,

donde $\star = \cap, \cup, \neg$ (en la teoría de conjuntos, el álgebra de sentencias y el álgebra de circuitos, respectivamente), y para cualesquiera x, y, z . Si, además, se cumple la ley

- 4) $x \star y = y \star x$,

se trata, entonces, de un "grupo abeliano". Esta última ley (conmutativa) no se cumple en el modelo sintáctico de Lambek, donde, como se verá, la ordenación de los tipos es distintiva.]

intransitivos, el tipo *i*, y así sucesivamente. Con ayuda de estos tipos primarios se forman tipos compuestos a base de tres operaciones formales: la multiplicación, la división a la izquierda y la división a la derecha. Indicando por medio de $X \rightarrow x$ que la cadena *X* tiene el tipo *x*, los siguientes ejemplos ilustran el significado de las operaciones propuestas:

Juan $\rightarrow n$

Juan trabaja $\rightarrow s$

trabaja $\rightarrow n \setminus s$

Juan trabaja $\rightarrow n (n \setminus s)$

el hombre laborioso trabaja $\rightarrow s$

el hombre laborioso $\rightarrow s / (n \setminus s)$ ³⁷

Aun en su trabajo "Mathematics", mencionado antes, Lambek establece las siguientes propiedades (por $x \rightarrow y$ se entiende que toda cadena de tipo *x* es también una cadena de tipo *y*):

- | | | | | | |
|------|---|-----|---|------|---------------------------|
| (1) | $x \rightarrow x,$ | (2) | $(xy)z \rightarrow x(yz),$ | (2') | $x(yz) \rightarrow (xy)z$ |
| (3) | si $xy \rightarrow z,$ | | entonces $x \rightarrow z/y,$ | | |
| (3') | si $xy \rightarrow z,$ | | entonces $y \rightarrow x \setminus z,$ | | |
| (4) | si $x \rightarrow z/y,$ | | entonces $xy \rightarrow z,$ | | |
| (4') | si $y \rightarrow x \setminus z,$ | | entonces $xy \rightarrow z,$ | | |
| (5) | si $x \rightarrow y$ e $y \rightarrow z,$ | | entonces $x \rightarrow z.$ | | |

A la luz de este cálculo, el autor se ve obligado a considerar gramaticales cadenas como **John works today today* ("*Juan trabaja hoy hoy"), **John works yesterday today* ("*Juan trabaja ayer hoy"), debido a que todo procedimiento que conduce a estas cadenas conduce también a *John works today at lunch* y a *John works at lunch today*, que, para Lambek, son perfectamente aceptables³⁸.

En este cálculo, a cada una de las cadenas de palabras se les atribuye un tipo. Otro punto de vista viene expresado en el cálculo sintáctico no - a s o c i a t i v o, que presentamos a continuación.

37. [En estas operaciones, las barras inclinadas $\setminus, /$, indican, respectivamente, división a la izquierda y división a la derecha. El numerador de las fracciones se refiere al tipo (primario o no) con que se combina un elemento, y el denominador, a la combinación resultante. Así, p. ej., "*trabaja* $\rightarrow n \setminus s$ " quiere decir que el elemento *trabaja* se combina hacia la izquierda con un elemento del tipo *n* (*Juan, el hombre, etc.*) y forma un miembro de *s*, o proposición. Nótese que en inglés (y en todas las lenguas con adjetivo antepuesto al nombre) *hombre laborioso* sería algo así como *industrious man* y se notaría por medio de $s / (n \setminus s)$].

38. [Esta construcción resultaría tal vez viable en una situación imperativa, en la que alguien ordenase trabajar a John mientras toma el almuerzo: *John works today at lunch!* La principal dificultad no estriba, sin embargo, en la combinación de los adverbios temporales *today* y *at lunch*, como parece desprenderse del texto, sino en la contradicción que se plantea entre *today* y la forma simple del verbo *to work*; habría que decir *John is working today at lunch* con la forma imperfectiva *is working*].

Se consideran dos colecciones A y B , donde los elementos de A se denominan frases atómicas, y los elementos de B , tipos primarios. Se define, de modo recursivo, el conjunto F de las frases como sigue:

- 1.º toda frase atómica es una frase;
- 2.º si X e Y son frases, entonces XY es también una frase.

También por vía recursiva se define el conjunto T de los tipos, dotado de tres operaciones: la operación de composición ($\alpha\beta$), la división a la izquierda ($\beta\backslash\alpha$) y la división a la derecha de α por β (α/β). Se imponen, además, las condiciones:

- 1.º todo tipo primitivo es un tipo de T ;
- 2.º si $\alpha \in T$, $\beta \in T$, entonces $\alpha\beta \in T$, $\beta\backslash\alpha \in T$ y $\alpha/\beta \in T$.

Los tipos se asocian a las frases de tal modo que:

- 1.º si la frase X tiene el tipo α , y la frase Y el tipo β , entonces XY tiene el tipo $\alpha\beta$;
- 2.º si XY tiene el tipo γ para todo Y de tipo β , entonces X tiene el tipo γ/β ;
- 3.º si $X\gamma$ tiene el tipo γ para todo X del tipo α , entonces Y tiene el tipo $\alpha\backslash\gamma$.

Se demuestra así que las reglas (1), (3), (3'), (4), (4') y (5) también son aquí válidas, mientras que las propiedades de asociatividad (2) y (2') dejan de serlo. Por este motivo, el cálculo que acabamos de desarrollar se llama "cálculo sintáctico no-asociativo". Una aplicación interesante de este cálculo se refiere a los modelos contruidos por Chomsky. En efecto, una gramática de tipo 2 en la jerarquía que establece Chomsky en "Formal properties" puede describirse con ayuda de un diccionario que atribuye a cada frase atómica un número finito de tipos primarios, en tanto que la gramática consta de un número finito de reglas con la forma $P_i P_j \rightarrow P_k$, donde P_i son tipos primarios. Lambek desarrolla una discusión detallada sobre la lengua L_2 de Chomsky, mostrando de qué modo puede describirse con ayuda de los tipos.

3.19. Dificultades que aparecen en el proceso de modelación

Un problema particularmente delicado consiste en distinguir, en el marco de un modelo, los elementos faltos de capacidad representativa. Ello ocurre en especial cuando el proceso modelado pertenece a un dominio de la realidad donde la influencia de la intuición inmediata, del sentido común y de ciertos prejuicios puede crear determinadas imágenes deformadas que impidan una visión más clara sobre el proceso que se somete a modelación. Así, sólo cuando se ha asumido la modelación lógica de las estructuras lingüísticas, a menudo se descu-

bre que debería cambiarse el planteamiento mismo del problema o que debería presentarse de otro modo el fenómeno en sí mismo antes de ser modelado. Tal sucede, p. ej., en el modelo de las partes del discurso realizado por Kulaguina, Uspenski y Revzin, donde no es posible que una sola palabra pertenezca a más de una parte del discurso. Parece, entonces, que un modelo de este tipo requiere un perfeccionamiento a fin de registrar la pertenencia de una palabra a varias partes del discurso, fenómeno muy frecuente, p. ej., en inglés (cf. 3.29). Existe, no obstante, otro procedimiento que, por lo demás, utilizan algunos lingüistas y que consiste en rectificar, sobre esta base, el concepto mismo de parte del discurso.

Una situación análoga aparece también en la modelación de las partes de la proposición en el momento de dividirlas en partes principales y partes secundarias.

Al parecer, a veces hay que ajustar no el modelo al objeto modelado, sino la descripción no-modelada o modelada de una manera respecto de otra modelada de otra manera. Conviene no perder de vista que los objetos de la realidad no admiten una modelación matemática directa y sí, en cambio, una modelación basada en otra descripción modelada o en una descripción no-modelada.

3.20. Modelo y metáfora

En relación con todo lo dicho, el concepto de modelo ofrece una cierta analogía con el concepto de metáfora. El libro de Tudor Vianu relativo a la metáfora confirma esta analogía³⁹. La función del modelo en el proceso del conocimiento científico es de alguna manera semejante a la función de la metáfora en el proceso del conocimiento artístico. Así como no existe descripción matemática sin modelación, tampoco existe poesía fuera de la metáfora. Tanto en un caso como en otro, se trata de una transferencia de una noción a otra que permite describir una de las nociones a través de la otra. El término metafórico es en parte análogo al término que expresa, aun siendo parcialmente distinto, justamente igual que el modelo con respecto al objeto modelado.

Un objeto complejo de la realidad es susceptible de una pluralidad, de una infinitud de modelaciones, del mismo modo que un objeto puede admitir una infinitud de descripciones metafóricas. Se establecen, así, puntos de conexión entre el conocimiento científico y el conocimiento artístico.

Resulta, pues, que en el proceso de modelación de los fenómenos más complejos de la biología y de las disciplinas humanísticas surge la necesidad de estudiar las relaciones concomitantes y multilaterales que se entablan entre un gran número de formas de movimiento de la materia; precisamente aquí halla

39. Cf. VIANU, *Problemele*.

su origen la extraordinaria interferencia que actualmente presentan los diferentes dominios de investigación.

3.21. La función gnoseológica de la modelación matemática de la lengua

Por lo que acabamos de ver se desprende que los modelos matemáticos de la lengua cumplen una función gnoseológica importante. La modelación matemática de la lengua constituye, por lo demás, una nueva confirmación gnoseológica del materialismo dialéctico. Las abstracciones que intervienen en el proceso de abstracción no sólo no nos alejan de la realidad de la lengua ni restringen el poder explicativo del modelo, sino que nos proporcionan la posibilidad de profundizar más en la estructura de la lengua. La afirmación de estas verdades es singularmente actual, por cuanto la modelación matemática de la lengua representa una característica de la etapa contemporánea de la lingüística estructural. Algunos matemáticos y lingüistas, incluso entre aquellos que han allegado contribuciones destacadas a la modelación matemática de la lengua, desnaturalizan el significado de estos modelos. Así ocurrió, p. ej., en ciertas comunicaciones presentadas en el Congreso de Lógica, Metodología y Filosofía de la Ciencia, celebrado en Stanford, en 1960. En el ámbito de la Sección de "Modelos en Lingüística", Yuen Ren Chao presentó al Congreso una comunicación sobre "Modelos en lingüística y modelos en general"⁴⁰. Según Yuen Ren Chao, el hecho de que un objeto A sea adoptado como modelo de un objeto B y no B como modelo de A se debe exclusivamente a que A es más cómodamente manipulable que B ⁴¹.

En principio, según Yuen Ren Chao, es indiferente considerar que A es modelo de B o, al revés, que B es modelo de A . Esta concepción pragmática, de un carácter evidentemente neopositivista, resulta extraña a la teoría materialista dialéctica del conocimiento. El criterio de la comodidad puede actuar cuando debemos elegir entre dos modelos igualmente adecuados del mismo objeto. Pero en lo que concierne al hecho de que A sea considerado un modelo de B , no se trata de una cuestión pragmática, como cree Yuen Ren Chao, sino de una consecuencia de la capacidad de A para reflejar algunas propiedades de B . ¿Qué función gnoseológica cumple el objeto matemático A en relación con el objeto B ? He aquí la cuestión decisiva con respecto a este problema. Considerar que es indiferente determinar cuál es el modelo y cuál el original comporta admitir

40. Cf. CHAO, "Models".

41. "But how does one tell, given two entities, which is the thing and which is the model? The answer will have to be a pragmatic one. In general, that which can more conveniently be handled—that is, seen, heard, remembered, recorded, communicated, manipulated, experimented upon, inherited, etc. is the model and that about which corresponding information or results are hoped to be obtained through such handling (in the broad sense) is the thing". ("Models", p. 565).

que es indiferente averiguar si el objeto investigado en última instancia es *A* o *B*. No obstante, este punto no puede resultar indiferente, ya que entre *A* y *B* existen diferencias esenciales y cada uno de los fenómenos *A* y *B* tiene su propia teoría.

3.22. Función explicativa de los modelos probabilísticos

Consideremos ahora un significativo ejemplo para ilustrar la función explicativa de la modelación probabilística. Este ejemplo fue expuesto por N. F. Pelevina en la Conferencia consagrada a los problemas de lingüística aplicada celebrada, en 1960, en la ciudad de Chernóvtsy (URSS)⁴².

Dadas dos palabras con forma idéntica o semejante, se plantea el problema de averiguar si se trata de una etimología idéntica o sólo de una homonimia casual. Por regla general, este tipo de problemas se resuelven recurriendo a ciertas leyes semánticas. Pero a veces faltan las informaciones necesarias para un análisis semántico y hay que buscar otros métodos. En una situación de este tipo puede resultar útil la modelación basada en el concepto de probabilidad. Aunque en el ejemplo que aduciremos existen todas las condiciones para resolver el problema por vía semántica, nuestro propósito consistirá en reencontrar por vía matemática la identidad etimológica de las palabras en cuestión. Una vez comprendido el método, puede aplicarse también a los casos donde falte la base semántica.

Consideremos las palabras rusas *kosoi*, 'oblicuo', y *kosa*, 'guadaña'. Se plantea entonces el problema de si la raíz común de estas dos palabras se debe a la casualidad o proviene de una etimología común a ambas. Estimando en 50.000 el número de palabras de la lengua rusa y teniendo en cuenta que el número de palabras rusas que empiezan por *kos-* es, aproximadamente, de 100, resulta que la probabilidad de que una palabra rusa empiece por *kos-* es igual a

$$\frac{100}{50\ 000} = \frac{1}{500}$$

Ahora bien, *kosoi* es sinónimo con *krivoi*, mientras que *kosa* lo es con el dialectalismo *krivusha*. Las palabras *krivoi* y *krivusha* empiezan ambas por *kriv-*, y las palabras rusas que empiezan por *kriv-* son, aproximadamente, 50. Así, pues, la probabilidad de que una palabra de la lengua rusa empiece por *kriv-* es igual a

$$\frac{50}{50\ 000} = \frac{1}{1\ 000}$$

42. Cf. PELEVINA, "Ustanovlenie"; [C ERDÁ, "Probabilidad"].

La sinonimia entre *kosoi* y *krivoi* se establece entre una palabra que empieza por *kos-* y una palabra que empieza por *kriv-* y, por tanto, entre una palabra cuya probabilidad de aparición es de $1/500$ y una palabra cuya probabilidad de aparición es de $1/1.000$. Resulta entonces que la probabilidad de una sinonimia de este tipo se obtiene multiplicando las dos probabilidades respectivas, de modo que

$$\frac{1}{500} \cdot \frac{1}{1\,000} = \frac{1}{500\,000}$$

Pero junto a la sinonimia de *kosoi* y *krivoi* hay que considerar también la sinonimia de *kosa* y *krivusha*. Sucede, entonces, que las raíces *kos-* y *kriv-* son utilizadas dos veces para expresar el mismo significado, por lo cual la probabilidad de una doble sinonimia como ésta es igual a

$$\frac{1}{(500\,000)^2} = \frac{1}{250\,000\,000\,000}$$

lo que, en la práctica, puede asimilarse a 0. Así, pues, la raíz común de las palabras *kosoi* y *kosa* no pueden coincidir por casualidad, sino que provienen de una etimología única.

Tratemos ahora de generalizar el procedimiento. Sean dos pares de palabras (A, B) y (C, D). Supongamos que A y C tienen la misma raíz α , y que B y D tienen asimismo una raíz común β . Notaremos a base de p_1 la probabilidad de que una palabra empiece por α , y a base de p_2 la probabilidad de que otra palabra empiece por β . En estas condiciones, por medio de un razonamiento como el que hemos visto antes, se comprueba que la probabilidad de que A sea sinónimo de B o de que C sea sinónimo de D (la sinonimia puede modelarse matemáticamente) es igual al producto $p_1 p_2$, mientras que la probabilidad de que haya al mismo tiempo sinonimia entre A y B y entre C y D es igual a $(p_1 p_2)^2$. Precisamente esta última situación ocurre en el fenómeno aducido antes, donde $A = \textit{kosoi}$, $B = \textit{krivoi}$, $C = \textit{kosa}$, $D = \textit{krivusha}$. Cuanto menor es la probabilidad $p_1 p_2$, más probable es la etimología común de las palabras A y B (o de las palabras C y D); y cuanto más pequeña es la probabilidad $(p_1 p_2)^2$, más probable es la existencia concomitante de la doble etimología común: la de A y B , por una parte, y la de C y D , por otra.

En el procedimiento que acabamos de examinar, los hechos han sido de alguna manera simplificados y sólo parcialmente modelados, con el único objeto de poner en evidencia la idea principal. Observamos que, en contraste con la costumbre de dar reglas seguidas de excepciones o de comprobar vagamente que los hechos se producen de tal modo más a menudo que de tal otro, la probabilidad introduce, entre la imposibilidad y el determinismo puro, una escala de matices intermedios objetivamente establecidos que no podrían ser descubiertos por procedimientos intuitivos o por el sentimiento de la lengua. Innume-

rables situaciones reciben así una explicación que las proyecta sobre el sistema lingüístico del que forman parte. Hechos más significativos relacionados con la función del modelo probabilístico en la lingüística, pero que por su naturaleza más bien compleja se prestan difícilmente a una exposición simplificada, podrían citarse entre las aplicaciones, ya numerosas, al estudio genético de las lenguas y a la procedencia histórica de determinadas lenguas o dialectos. Particularmente interesante resulta, desde este punto de vista, el método de la glotocronología, relativo al problema de establecer la datación de ciertos procesos lingüísticos⁴³.

La modelación matemática de la evolución lingüística se verá también favorecida por el desarrollo de la dialectología estructural⁴⁴.

3.23. El carácter específico de las abstracciones matemáticas

La idea de que las abstracciones matemáticas se distinguen de las abstracciones no-matemáticas por alcanzar un nivel más cumplido es errónea. Bien es verdad que las abstracciones matemáticas acceden a un grado superior, en el sentido de que no se obtienen directamente a partir de los objetos reales, sino por medio de otras abstracciones. Sin embargo, las abstracciones con que opera la filosofía alcanzan, por lo general, un nivel más alto aún que el de las abstracciones matemáticas. Ni siquiera sobre las abstracciones habituales en ciertas teorías lingüísticas, como p. ej. en la glosemática, cabe decir que se hallan a un nivel inferior a las abstracciones matemáticas. Las características de las abstracciones matemáticas son de otra índole. Una de ellas consiste, p. ej., en el hecho de que se destacan de los objetos a partir de los cuales se han extraído, por lo que reciben una existencia autónoma que las convierte en susceptibles de un tratamiento matemático, axiomático-deductivo, mientras que los objetos que las han generado ya no intervienen más que en una de las muchas interpretaciones posibles que admiten los resultados matemáticos obtenidos. De este modo, y sólo a título de ejemplo, el modelo matemático de las gramáticas con un número finito de estados tiene una existencia enteramente independiente de la realidad lingüística sobre la que se inspira. Ello no obstante, esta realidad recibe una nueva luz en una de las interpretaciones posibles del modelo considerado. Este hecho no limita el poder explicativo de las abstracciones matemáticas, sino que, antes bien, lo

43. [Este método trata de determinar el grado de separación temporal entre lenguas, familias o grupos más amplios a partir de la fórmula:

$$t = \frac{\log C}{2 \log r}$$

donde t es el tiempo, contado en milenios; C el porcentaje de palabras comunes, y r , el índice constante de conservación. Cf. SWADESH. "Lexico-statistical dating"; LEES. "The basis", y COSERIU. "Critique"].

44. Para una discusión más amplia sobre la dialectología estructural, cf. A VRAM. "Despre dialectologie", [y ALVAR. *Estructuralismo*].

engrandece. Así, p. ej., el modelo matemático de las gramáticas con un número finito de estados, aun admitiendo una interpretación como autómatas finitos, establece un isomorfismo entre la teoría de los autómatas finitos y la teoría de las gramáticas⁴⁵. Por consiguiente, conviene subrayar que todo el problema del isomorfismo lingüístico queda planteado en una nueva perspectiva, como consecuencia del desarrollo de la cibernética y de la modelación matemática de la lengua.

3.24. Isomorfismo y modelo

El concepto de isomorfismo, utilizado por L. Hjelmlev para describir la conformidad entre los dos planos de la lengua, ha recibido una notable ilustración lingüística en la obra clásica de J. Kuryłowicz sobre la analogía entre la sílaba y la oración⁴⁶. Este mismo problema ha recibido un tratamiento ulterior en otros trabajos⁴⁷.

El isomorfismo, tal como lo expone Kuryłowicz, corresponde, por lo que veremos después, a la existencia de un modelo matemático común a la sílaba y a la oración, basado en la relación de subordinación y en el concepto de par predicativo, los cuales se definen en los modelos de la lengua que parten de la teoría de conjuntos. Otro ejemplo significativo de isomorfismo, revelado por medio de la modelación matemática de la lengua, se refiere al que existe entre la noción de clase distribucional y la noción de parte del discurso (cf. 7.2)⁴⁸.

Junto con el isomorfismo intralingüístico, importa considerar asimismo el isomorfismo entre las estructuras lingüísticas y las estructuras extralingüísticas. En esta dirección también se han obtenido resultados interesantes, relativos a los nexos entre la lingüística, por una parte, y la teoría de los autómatas, la termodinámica y la topología, por la otra. La significación filosófica de esta suerte de resultados no puede pasar inadvertida, ya que así se desvelan relaciones profundas entre formas de movimiento material aparentemente distintas.

La profundización de las relaciones entre isomorfismo y modelación en lingüística requiere una sólida investigación ulterior a la luz de los más recientes datos de la cibernética.

Como hemos dicho ya, en lingüística el término de isomorfismo fue empleado por primera vez por L. Hjelmlev para designar la correspondencia entre los dos planos fundamentales de la lengua, el del contenido y el de la expresión, o bien, lo que viene a ser lo mismo, la correspondencia entre el sistema gramatical y el sistema fonológico de la lengua. A causa de la imprecisión con que

45. Cf. BAR-HILLEL y SHAMIR, "Finite-state languages".

46. Cf. KURYŁOWICZ, "Notion de l'isomorphisme".

47. Cf. LEKOMTSEV, "K voprosu ob analoghii", y SHAUMIAN, "Lingvisticheskie voprosy".

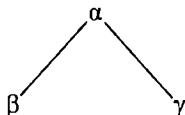
48. Cf. MARCUS, *Lingvistică matematică*.

Hjelmslev utilizó el término de isomorfismo, por cuanto permite diversas y muy variadas interpretaciones, la idea del isomorfismo lingüístico ha recibido tanto adhesiones entusiastas como protestas vehementes. Unos, considerando el isomorfismo entre *A* y *B* como una correspondencia por la cual se identifican ambos miembros, es decir por la cual *A* y *B* devienen indiscernibles, han puesto en tela de juicio la existencia del isomorfismo entre el sistema fonológico y el gramatical, argumentando que estos dos sistemas no admiten identificación alguna en tanto que presentan una diferencia fundamental: la ausencia de significado, en el primero, frente a la presencia del mismo, en el segundo. A esto se ha respondido que el isomorfismo no comporta la identificación entre *A* y *B*, sino tan sólo una conformidad o, en otras palabras, una constitución semejante. Las semejanzas existen —se ha replicado—, pero sólo afectan a aspectos secundarios de los dos sistemas, mientras que, en lo que tienen de esencial, el sistema gramatical se distingue del fonológico por la presencia del significado, ausente en este último. Las palabras contienen significado, pero los fonemas, no.

Una intervención interesante en las discusiones en torno al isomorfismo se debe al lingüista polaco J. Kurylowicz. Unos 25 años atrás analizó sistemáticamente lo que él mismo dio en llamar el “isomorfismo entre la sílaba y la proposición”. Para explicar mejor los hechos, consideremos un ejemplo en español: la sílaba *trans* y la proposición *El alumno aplicado aprende bien las lecciones*. Aparentemente, no puede establecerse ninguna analogía entre dos entidades de este tipo. Pero con un examen más atento, se observan ciertos puntos comunes en la estructura de ambas. En efecto, tanto en la sílaba como en la proposición existe un núcleo sin el cual ninguna de ellas sería posible. En la sílaba, este núcleo viene representado, en español, por un elemento vocálico cualquiera; no hay, desde luego, sílaba sin vocal en español. De la sílaba *trans* cabe separar el grupo consonántico *ns* sin que se pierda la condición de sílaba en el grupo resultante. Pero si se extrae la vocal *a*, la sílaba desaparece, ya que **trns* o **tr* es imposible en español. De un modo semejante, también la proposición contiene un núcleo sin el cual dejaría de ser proposición: este núcleo es el predicado y, concretamente, *aprende*, en el ejemplo de antes. De manera que, por tanto, no existe ni sílaba sin núcleo vocálico ni proposición sin núcleo predicativo. Pero vayamos más adelante. El grupo *aprende bien las lecciones* constituye lo que se denomina un grupo del predicado. Y en este grupo, el subgrupo *bien las lecciones* es un satélite del predicado *aprende*, ya que, al suprimirse, lo que queda sigue siendo una proposición: *el alumno aplicado aprende*. Esta misma situación tiene su contrapartida en la sílaba, donde el grupo de consonantes *ns* constituye un satélite de la vocal *a*, puesto que, al suprimirse, tampoco desaparece la naturaleza silábica del resultado. He aquí, pues, cómo en una proposición el grupo de los complementos guarda una relación con el predicado del mismo tipo que guarda, en la sílaba, el grupo final de consonantes respecto a la vocal nuclear.

Continuando esta analogía, cabe observar que el grupo del sujeto, *El alumno aplicado* en el ejemplo aducido, se corresponde, en la sílaba, al grupo inicial de consonantes, *tr* de nuestro caso. La analogía puede prolongarse y profundizarse más aún, pero los ejemplos ofrecidos bastan para lo que queríamos demostrar.

Así, pues, dos entidades lingüísticas pueden presentar, hasta un cierto punto, la misma estructura, a pesar de que, por su propia naturaleza, sean muy diferentes. Esquemáticamente, la estructura común de la sílaba y de la proposición es representable así:



α representa tanto un predicado como una vocal; β , un grupo de sujeto o un grupo inicial de consonantes y γ , un grupo de complementos o un grupo final de consonantes. Los elementos inferiores β y γ están subordinados al elemento superior α ; β y γ son susceptibles, a su vez, de un ulterior análisis dicotómico, que aquí omitimos.

Si tratamos de averiguar con qué concepto de isomorfismo ha operado Kuryłowicz en este trabajo, no es difícil demostrar que tiene que ver justamente con el utilizado en matemática. Así, una de las variantes más frecuentes del isomorfismo matemático se presenta del siguiente modo: se consideran dos conjuntos A y B y se define en A una determinada relación binaria R_1 y en B otra relación binaria R_2 . Si es posible establecer, entre los conjuntos A y B , una correspondencia biunívoca tal que, suponiendo que x_1 de A corresponda a x_2 de B y que y_1 de A corresponda a y_2 de B , y suponiendo que si x_1 e y_1 se hallan en relación R_1 , resulte que x_2 e y_2 se hallen en relación R_2 (del mismo modo que, suponiendo que x_2 e y_2 se hallen en relación R_2 , resulte que x_1 e y_1 se hallen en relación R_1), entonces decimos que A y B son isomorfos con respecto a la pareja de relaciones R_1, R_2 . Notando por medio de (A, R_1) el conjunto A estructurado con la relación R_1 y por medio de (B, R_2) el conjunto B estructurado con la relación R_2 , podemos decir que existe isomorfismo entre (A, R_1) y (B, R_2) . En el ejemplo de antes, A es el conjunto de los fonemas de la sílaba, B es el conjunto de las palabras de la proposición, R_1 es la relación de subordinación en el ámbito de la sílaba y R_2 es la relación de subordinación en el ámbito de la proposición. Justamente a causa de la importancia contenida en la relación de subordinación, el isomorfismo establecido por Kuryłowicz entre la sílaba y la proposición no carece de interés. Kuryłowicz emprendió en su análisis una verdadera modelación matemática, que aclara perfectamente la esencia conceptual del isomorfismo entre A y B : la existencia de un modelo matemático común para A y B . Con ello queda más preciso el carácter relativo del isomorfismo,

según el cual no tiene sentido hablar acerca de un isomorfismo absoluto entre sílaba y proposición si éstas no se conciben como estructuradas en base a la función de subordinación, función que la lingüística estructural de hoy puede formalizar totalmente.

Si volvemos ahora al isomorfismo entre el sistema fonológico y el sistema gramatical, al que se refería Hjelmslev, no será difícil precisar el sentido que conviene conceder a este isomorfismo. Sea un número de relaciones binarias $r_1, r_2, \dots, r_n, \dots$ definidas en el conjunto A de los fonemas de una lengua, y un número de relaciones binarias $R_1, R_2, \dots, R_n, \dots$ definidas en el conjunto B de los morfemas de la misma lengua tales que existe isomorfismo entre (A, r_1) y (B, R_1) , entre (A, r_2) y (B, R_2) , (A, r_n) y (B, R_n) y así sucesivamente. La precisión de estas relaciones binarias, es decir del carácter relativo del isomorfismo entre la fonología y la gramática, constituye una de las metas principales en la modelación matemática de la lengua. Tal como hemos visto, el isomorfismo y la modelación están estrechamente ligados entre sí. El valor gnoseológico del isomorfismo depende del grado de esencialidad que presenten, para las estructuras estudiadas, las relaciones binarias implicadas en el isomorfismo. El isomorfismo permite, hasta cierto punto, reducir el estudio de diversas estructuras al estudio de un solo modelo matemático. Por ello, no sólo se consigue economizar importantes recursos, sino que, al mismo tiempo, se penetra más profundamente en la esencia de los fenómenos estudiados, se consigue no sólo su descripción, sino también una explicación más cabal de su estructura interna.

Además del isomorfismo que cabe establecer en el marco de la lingüística, es también muy importante el que se establece entre las estructuras lingüísticas y las que estudian otras ciencias. Este problema, elevado a primer plano por el desarrollo de la cibernética, ha colocado también a un nivel superior las tareas de la modelación matemática en lingüística.

3.25. La modelación algorítmica

En cuanto a la modelación algorítmica, ofrece una especial significación el análisis algorítmico de los textos conservados del pueblo maya, que vivió hace varios siglos en las regiones de América Central. Este análisis, realizado en el Centro de Cálculo de Siberia, bajo la dirección del matemático S. L. Sobolev, ha desvelado aspectos nuevos sobre la oposición dialéctica entre cantidad y calidad. Se ha comprobado que una inmensa cantidad de operaciones elementales de carácter evidentemente cuantitativo, cantidad que hasta hace poco parecía inaccesible a una descripción matemática, se hace susceptible de modelación gracias a los ordenadores electrónicos de alta velocidad. Ello posibilita una mejor profundización sobre los aspectos cualitativos y favorece el paso de una esencia a otra superior, de acuerdo con la teoría leninista de la reflexión.

Lo que antes era cualitativo hoy admite una conversión a lo cuantitativo y se incorpora al dominio de la práctica cotidiana.

3.26. La modelación matemática y el desciframiento de la lengua maya

De la floreciente civilización que existía antiguamente en los territorios de América Central se han conservado tres manuscritos mayas (el de Madrid, el de Dresda y el de París) y un gran número de inscripciones lapidarias. Se conocen en la actualidad más de 1 000 inscripciones que aumentan en el curso de los años, aunque muchas de ellas se encuentran en mal estado a causa de las lluvias tropicales. Todos estos textos son muy cortos; el más amplio descubierto hasta ahora comprende 5 000 signos. Se han encontrado, además, más de cien ciudades mayas.

La principal dificultad en el desciframiento de estos textos radica, en primer lugar, en su carácter específico, concretamente sacerdotal y lacónico, al margen del elevado índice de deterioro. Una gran parte de la información —a veces, incluso, información básica— aparece en forma de dibujos alegóricos, cuyo significado es extremadamente difícil de comprender.

De este modo, para el análisis es imprescindible tomar en consideración todos los tipos de indicios contenidos en los manuscritos: textos jeroglíficos, dibujos y datos calendarísticos. Otras dificultades serias derivan de la ausencia de diccionarios y gramáticas bien confeccionadas sobre la lengua maya así como de la parvedad de noticias sobre los ritos y costumbres de aquellos pobladores.

Todas estas dificultades exigen la realización de un análisis profundo sobre los manuscritos y el allegamiento de una gran cantidad de materiales auxiliares (diccionarios, guías, etc.). Para alcanzar estos cometidos se invirtió muchísimo tiempo y, sin embargo, no se logró leer los manuscritos. Así, de una manera natural, se ha planteado entonces la oportunidad de utilizar las máquinas electrónicas de cálculo en el análisis de los manuscritos de los antiguos mayas.

En la etapa actual, no puede automatizarse completamente el proceso de investigar los antiguos sistemas de escritura; a lo sumo, se emplea la máquina para efectuar los procesos que comporten un trabajo de gran envergadura y que, al mismo tiempo, sean fácilmente formalizables.

El objeto de investigación del equipo de Novosibirsk se ha aplicado a los manuscritos con antiguos jeroglíficos mayas, conservados en Madrid y en Dresda, y que constituyen asimismo documentos históricos de la mayor importancia para los pueblos primitivos.

Durante el período colonial español, la población maya continuó escribiendo en su lengua, aunque utilizando los caracteres del alfabeto de los conquistadores. Del siglo *xvi* se ha conservado una cierta cantidad de materiales léxicos (crónicas, etc.) en los que se reflejan, hasta un cierto punto, las viejas tradicio-

nes mayas. Se han conservado, además, diccionarios mayas en transcripción española, algunos de ellos (p. ej., el diccionario Motul) compuestos justo al principio de la dominación hispánica. A partir de diversas fuentes etnográficas se han obtenido igualmente noticias sobre el sistema de escritura de la antigua lengua maya⁴⁹. Así, el análisis preliminar de aquellas fuentes y de los escritos investigados ha conducido a las siguientes hipótesis:

1) sobre el carácter de la escritura (que es jeroglífica, y los signos son fonéticos, ideográficos o determinativos);

2) Sobre la correspondencia que existe entre el significado de las palabras de los diccionarios y los materiales léxicos del período de colonización, por una parte, y el significado de las palabras de los antiguos manuscritos mayas, por otra.

La segunda hipótesis se halla en concordancia con el principio general en que se basa todo desciframiento, y que consiste en establecer una correspondencia entre los elementos del texto desconocido y los elementos de una lengua conocida al menos parcialmente. Champollion descifró textos egipcios estableciendo una correspondencia análoga entre éstos y la lengua de los coptos, un pueblo que en buena medida había mantenido, en el ámbito de su lengua, elementos de la antigua. Edgar A. Poe, en su cuento *El escarabajo de oro*, aprovechó, para descifrar el mensaje del capitán Kidd, la ventaja de conocer el orden en que se encuentran las letras, en inglés, desde el punto de vista de su frecuencia relativa contextual. Pero, mientras en el cuento de Poe se trataba de descifrar un texto escrito en una lengua conocida, aunque codificada, los textos mayas están escritos en una lengua en buena parte desconocida.

Partiendo, pues, de las dos hipótesis aducidas, el problema de analizar la antigua escritura maya consiste en establecer una correspondencia entre las palabras de los materiales léxicos y los signos de los viejos manuscritos y en determinar, sobre esta base, el tipo de uso y el significado de los signos jeroglíficos.

Para abordar con éxito esta tarea, es indispensable tomar en consideración un material lo más rico y diverso posible. El material léxico es suficientemente variado y accesible (diccionarios, crónicas, profecías, etc.). En cambio, los textos jeroglíficos son escasos, pese a que, utilizando de un modo juicioso los dibujos y los datos calendarísticos de los manuscritos, cabe compensar esta insuficiencia.

A fin de realizar la correspondencia entre las palabras de los textos jeroglíficos y las palabras escritas en transcripción española, surge el problema de elegir los criterios más seguros y adecuados. Para ello, éstos deben reunir las siguientes condiciones:

α) un signo que presenta en distintas palabras el mismo uso debe tener cada vez el mismo significado;

49. Cf. LANDA. *Relation*.

β) todas las palabras de los materiales léxicos deben encontrarse en el diccionario Motul;

γ) las frases de los textos léxicos deben corresponder a las fracciones, dibujos y datos calendarísticos de los textos jeroglíficos.

El proceso de investigación, a su vez, consta de dos fases:

a) la identificación de las palabras y determinación del significado y del tipo de uso de los signos (condiciones α y β);

b) establecimiento de la correspondencia entre las frases léxicas y las distintas fracciones, dibujos y datos calendarísticos de los textos jeroglíficos (condición γ).

Para la algoritmización y programación del problema de desciframiento se requiere: una guía para la búsqueda de las palabras en los textos de las crónicas mayas; un diccionario de las palabras que aparecen en estas crónicas; una guía para la búsqueda de las palabras en el diccionario Motul (en el cual no vienen dispuestas en orden rigurosamente alfabético; en él hay, además, muchas expresiones idiomáticas); diccionarios maya-ruso y ruso-maya compuestos a partir de los diccionarios maya-español y español-maya; un diccionario temático maya-ruso, etc.

En el curso de la investigación, deben identificarse ciertas palabras a base tan sólo de algunas de las letras que las componen. Evidentemente, la realización de este programa necesita una labor inmensa. Así, p. ej., para encontrar una palabra a partir de algunas letras componentes hay que consultar más de 900 páginas de diccionario.

Las máquinas electrónicas de cálculo ofrecen la posibilidad de resolver de una manera total y rápida los problemas de esta índole. Para ello, sin embargo, conviene ante todo redactar la información completa que contienen los textos mayas de un modo accesible a la máquina. En este tipo de redacción formalizada se emplea el siguiente simbolismo:

Sean $a_1, a_2, \dots, a_i, \dots, a_n$ elementos del material jeroglífico estudiado; n es el número de signos distintos y de datos calendarísticos que aparecen en los textos.

Todo elemento a_i se caracteriza mediante un esquema de forma:

$$a_i \rightarrow (A_j B_k)_i p_i q_i \alpha x_i,$$

donde A_j es el indicativo del complejo de elementos jeroglíficos en los manuscritos, de modo que j recorre tantos valores como palabras hay en los manuscritos de Madrid y Dresda; B_k es el indicativo del signo a_i en el complejo, en el que k toma valores de 1 a 5 (un complejo que compone un jeroglífico no puede tener más de 5 signos); p_i es la frecuencia del signo respectivo en el manuscrito de Madrid; q_i representa lo propio en el manuscrito de Dresda; α es un pará-

metro que caracteriza el tipo de uso del signo (fonético, ideográfico, cifra, determinativo) y x_i es el significado del signo, escrito en transcripción española.

Un complejo jeroglífico se escribe a base de $b_1, b_2 \dots b_5$, donde b_k ($k = 1, \dots, 5$) es un signo del complejo, es decir uno de los elementos a_1, \dots, a_n . Todo complejo se caracteriza por un esquema del tipo siguiente:

$$(b_1 b_2 b_3 b_4 b_5)_j \rightarrow (A_j)_i p_j q_j \beta (x_1 x_2 \dots x_q)_j,$$

donde $(A_j)_i$ es el indicativo, en el manuscrito, del complejo de rango j ; p_j es la frecuencia del complejo de rango j en el manuscrito de Madrid; q_j es, a su vez, la frecuencia del mismo complejo en el manuscrito de Dresda; β es un parámetro que caracteriza la estructura del complejo (el orden de lectura de los signos de que se compone), y $(x_1 x_2 \dots x_q)_j$ es el significado, en el diccionario, del complejo de rango j .

Para describir los dibujos se emplea también un determinado simbolismo. Los dibujos plantean el problema de traducir ciertas situaciones de la mitología maya. La composición de un dibujo maya refleja una estructura situacional de la mitología y una cierta sintaxis de situaciones. El conjunto de los distintos objetos del dibujo se pone en correspondencia con un conjunto de significados u_i . Todo elemento u_i se caracteriza por un esquema de forma:

$$u_i \rightarrow (C_j D_k)_i \gamma \delta d_s f_i (x_1 x_2 \dots x_l)_i,$$

donde C_j es el indicativo del dibujo en el manuscrito; D_k , el indicativo del objeto en el dibujo; γ , un parámetro que caracteriza las relaciones recíprocas entre los objetos del dibujo; δ , un parámetro que caracteriza el tipo de uso, en el dibujo, de los signos jeroglíficos o de sus elementos; d_s , un parámetro que caracteriza la clase de objetos a que pertenece el objeto con significado u_i ; f_i , un parámetro que caracteriza la complejidad del objeto, y $(x_1 x_2 \dots x_l)_i$, la denominación del objeto de rango i en el diccionario Motul.

Los objetos más complejos se expresan con ayuda de los símbolos que representan objetos más simples, del mismo modo que los complejos jeroglíficos se escriben mediante signos.

A toda sílaba del material léxico le corresponde un elemento d_i , que se caracteriza por los siguientes parámetros: B_k es el lugar de la sílaba en la palabra; p_{ik} , la frecuencia de la sílaba en un contexto dado, en el marco de las palabras; α , un parámetro que caracteriza la estructura de la sílaba; x_i , la sílaba que corresponde a d_i en la transcripción española (es decir, escrita en alfabeto español).

Una palabra del material léxico se representa en forma de $d_1 d_2 \dots d_l$. Toda palabra se caracteriza por un esquema del siguiente tipo:

$$(d_1 d_2 \dots d_l) \rightarrow (A_j)_i p_j (x_1 x_2 \dots x_l)_j,$$

donde $(A_j)_i$ es el indicativo de la palabra en el texto; p_j , la frecuencia de la palabra, y $(x_1 x_2 \dots x_l)_j$, el significado de la palabra de rango j .

Para analizar los sistemas antiguos de la escritura maya, donde se manipula un material heterogéneo, no es posible limitarse a un solo método de investigación. Es necesario recurrir a una gama entera de métodos para explotar en cada caso su particular eficacia en una dirección determinada. El método estadístico, inspirado en la regularidad estadística de la lengua, tal como ha demostrado A. A. Markov, conduce a un establecimiento suficientemente riguroso del significado de los distintos signos. El método basado en el aprovechamiento de las leyes existentes en la estructura de las diversas porciones, frases y complejos permite establecer el significado funcional de las frases, complejos y otros signos, e incluso, en algunos casos, el aspecto fonético exacto. El método para determinar el significado de un signo a partir de su contenido jeroglífico facilita la elección de la palabra equivalente. El método que utiliza la correspondencia entre el contenido de los dibujos y el sentido de las frases permite averiguar el significado de los distintos signos. Y el método basado en la composición de los complejos que constan de diferentes signos sin significado único determinado ofrece la posibilidad de poner en correspondencia los diversos grupos de complejos con grupos de palabras. En particular, cuando un grupo de complejos corresponde a una sola palabra del diccionario, cabe identificar con precisión la correspondencia fonética de los complejos.

El método más eficiente es el llamado método "rebusista"⁵⁰, inspirado en la búsqueda de la correspondencia entre complejos que contengan un número determinado de signos y palabras con un número igual de sílabas. La eficacia de este método reside sobre todo en el hecho de que al comparar los complejos y las palabras no se introduce ninguna limitación y, luego, en que se aplica a todos los tipos de materiales: léxicos, jeroglíficos, datos calendarísticos, dibujos.

El tratamiento del material estudiado mediante estos métodos se realiza del siguiente modo: en función del significado transmitido por los dibujos y del contexto, se establece el diccionario temático para una porción dada del manuscrito. Las palabras del diccionario temático se ponen, a su vez, en correspondencia con los complejos de la porción respectiva, a fin de que los significados que proporcionan los signos sean verificados en otros complejos, teniendo en cuenta las condiciones α y β que hemos expuesto antes.

Todos los métodos de cálculo se utilizan en la primera fase de la investigación. En esencia, esta fase consiste en elegir los textos en que se cumplen aquellas condiciones. Sigue luego la fase en que se confecciona, partiendo de los textos elegidos antes, un texto tal que corresponda al tema, a la porción, a los dibujos y a los datos calendarísticos o, dicho de otra manera, que satisfaga la condición γ .

La selección de los textos se efectúa a base de aprovechar los resultados ob-

⁵⁰ [De la expresión latina *de rebus quae geruntur* y con un sentido análogo al de los logógrafos y charadas].

tenidos en otras investigaciones de carácter histórico, etnográfico, arqueológico, etcétera. Se emplean algoritmos que describen el ritual, los sacrificios, el trabajo agrícola, la caza, las ceremonias y las distintas profesiones. La etapa final de esta segunda fase consiste en explicar el significado de las frases y de las palabras, lo que constituye una tarea posible a base de los métodos de la lingüística tradicional.

El problema de investigar la antigua escritura maya es, como se ve, una cuestión cibernética bastante compleja, cercana, por su naturaleza, a los problemas de la traducción automática y de la investigación de los sistemas de señales (cf. 7.2 y 10.2). En su primera fase, la indagación se parece a la traducción palabra por palabra (cf. 2.5), y en la segunda (establecimiento de correspondencias entre frases y porciones de los textos jeroglíficos) recuerda la operación misma de traducir.

3.27. El papel de la modelación lógica en la formalización del discurso indirecto

Hace unos quince años, el conocido matemático holandés Hans Freudenthal publicó un libro sobre la llamada “lengua cósmica” (cf. 6.1 y ss.)⁵¹.

En este tratado, el autor intenta aplicar el lenguaje de la lógica simbólica —utilizado, hasta hace poco, sólo en matemática— al habla cotidiana, a la conversación corriente. Surge, no obstante, una dificultad. En matemática, el criterio de la corrección consiste en formular definiciones, teoremas y demostraciones que la máquina pueda repetir. La máquina puede controlar la coherencia de una definición, la exactitud de unas fórmulas o de una demostración. Pero si aplicamos la lógica a objetos no-matemáticos, H. Freudenthal opina que ya no puede aplicarse como criterio la formalización propia de la máquina. Se impone, entonces, otro criterio, como el siguiente: hay que formularlo todo de tal modo que pueda interpretar cualquier texto incluso alguien que ignore todas las lenguas vivas del mundo —p. ej., un astronauta llegado a la tierra desde otro planeta—. Para ejecutar semejante programa, Freudenthal incluyó en su empresa numerosos compartimientos de la ciencia, como la lógica, el álgebra, la mecánica, la teoría de las probabilidades, etcétera. Con objeto de ilustrar las particulares dificultades que aparecen en relación con el discurso indirecto y con el modo de tratar las variables, espigaremos algunas de las consideraciones del mismo Freudenthal.

Sean p y q dos proposiciones. Notaremos a base de $p \vee q$ (p o q) su disyunción, de $p \wedge q$ (p y q) su conjunción, de $\neg p$ (no p) su negación, y de $p \rightarrow q$ (p implica q) la implicación de q por p .

51. Cf. FREUDENTHAL, *Cosmic Language* y también *Lincos*.

Por *variable* se entiende un nombre ambiguo. *César* es un nombre propio, pero *hombre* es un nombre ambiguo aplicable a todos los individuos humanos. *Roma* es el nombre propio de una ciudad, pero *ciudad* es un nombre ambiguo aplicable a todas las ciudades. *Yo, tú, él* son variables que sólo pueden indicar personas. *Hoy* es una variable que sólo se aplica a los días y *aquí* es otra variable sólo aplicable a lugares. Las variables de las lenguas vivas presentan dominios restringidos y diversos. En matemática, por el contrario, se emplean variables libres, expresadas mediante nombres muy sencillos, como x, y, z, \dots . Así, pues, no se limita a priori el dominio de una variable; tan sólo se explica, según el caso, el modo en que se limita. Si, p. ej., queremos indicar que la variable x debe referirse exclusivamente a la clase de los hombres, basta escribir $x \in \text{hombre}$.

En una proposición con sentido ninguna variable puede ser completamente libre. La libertad de las variables debe suprimirse, y cada variable ha de controlarse por un procedimiento cualquiera, p. ej., a base de cuantificación. De esta suerte, en la frase *todos los hombres son mortales*, la limitación se hace mediante generalización. En el lenguaje de la lógica, esta limitación se escribe así:

$$\Lambda x (x \in \text{hombre}) \rightarrow (x \in \text{mortal})$$

y se lee: “para todo x , si x es hombre, entonces x es mortal”. Como puede observarse, a través de Λ se ha notado el cuantificador universal. La negación de esta proposición se escribirá:

$$\forall x (x \in \text{hombre}) \wedge (x \notin \text{mortal})$$

y se leerá: “existe un x tal que x es hombre y x no es mortal”. Aquí, la variable x queda limitada por el cuantificador existencial (notado por \forall).

Hay además otros procedimientos para limitar una variable; p. ej., mediante el artículo definido (notado por \downarrow):

$$\downarrow x (x \in \text{hombre}) \wedge (x \notin \text{mortal})$$

(en caso de que no exista más que un solo hombre que no sea mortal).

Ahora bien, estos procedimientos de cuantificación resultan insuficientes. Por ello, H. Freudenthal introduce un nuevo recurso para suprimir la libertad de las variables, denominado de interrogación. Si formulamos la pregunta

$$(?x) (8x = 12),$$

leyendo: “¿qué x tiene la propiedad $8x = 12$?”, la variable queda igualmente limitada. Observemos otro ejemplo que aclara la importancia de este cuantificador. Sea la proposición

$$t_1 t_2 \text{ Ha Inq Hb } (?x) (8x = 12),$$

que se lee así: “del momento t_1 al momento t_2 , la persona Ha dice a la persona Hb: “¿Qué x tiene la propiedad $8x = 12$?””.

Entre los problemas relacionados con el uso de los pronombres relativos, H. Freudenthal destaca, en especial, el del discurso indirecto. Consideremos, a este propósito, una conversación de tipo:

Ha Inq Hb ($?x$) $8x = 12$,

Hb Inq Ha $x = 3/2$,

Ha Inq Hb Ben,

donde Ha (el profesor) formula un problema de cálculo aritmético, Hb (el alumno) contesta y Ha concede una nota, Ben (= bien) —como en el presente caso— o Mal, si el alumno da una respuesta errónea (aquí hay que prever la posibilidad de que éste conteste $12/8$). Un observador podría resumir estas conversaciones con la ayuda de unas reglas generales:

$$[t_1 t_2 \text{ Ha Inq } (?x) (8x = 12) \wedge (t_2 t_3 \text{ Hb Inq Ha } u/v)] \rightarrow V_{t_4, w} (t_3 t_4 \text{ Ha Inq Hb } w \wedge [w = \text{Ben} \leftrightarrow (u = 3 \wedge v = 2)] \wedge [(w = \text{Mal}) \leftrightarrow (u \neq 3 \wedge v \neq 2)]).$$

¿Qué relación guardan estos hechos con el discurso indirecto? Analicemos de momento qué cabe entender por discurso indirecto. Tomemos, p. ej., la frase *Ayer Juan me dijo: tú estás enfermo*. Mediante el discurso indirecto, esto se expresaría *Ayer Juan me dijo que yo estoy enfermo*. Según el parecer de Freudenthal, no existe una diferencia semántica entre el discurso directo y el indirecto, pese a que la diferencia en lo que se refiere a la estructura gramatical es evidente⁵². Pero lo esencial consiste en la sustitución de la variable *tú* por la variable *yo*, y de la variable *estás* por la variable *estoy*. En otras situaciones, es un nombre propio el que queda sustituido por una variable. Si le digo a Pedro: *Juan ha dicho: Pedro está enfermo*, la construcción indirecta sustituirá la variable *Pedro* por la variable *tú*. Así, la adopción del discurso indirecto se reduce a la sustitución de los nombres propios por variables y la de unas variables por otras o por nombres propios.

El discurso indirecto, al principio evitado, retorna de nuevo, a un nivel más alto. Comparando la conversación auténtica entre Ha y Hb con el relato hecho por el observador de antes, comprobamos que constantes como 8, 12, Ben, Mal han quedado sustituidas por las variables u , v , w , a , b . Entonces, al narrar una conversación hay que sustituir constantes auténticas por variables o variables por otras variables. Nos enfrentamos aquí con una analogía del discurso indirecto de las lenguas naturales en un lenguaje artificial. En cuanto queremos hacer aserciones sobre el comportamiento general de los hombres, ya no podemos

52. [Igual que en el original rumano, hacemos aquí abstracción del factor tiempo, ya que la versión indirecta de la primera frase también podría ser *Ayer Juan me dijo que yo estaba enfermo*].

respetar la autenticidad total de las cosas que relatamos, ya que una parte de nuestro relato proviene de haber sustituido ciertas constantes por variables y ciertas variables por otras variables. En un relato de este tipo, utilizamos un esquema de la siguiente forma:

$$t_1 t_2 \text{ Fit } p,$$

que se lee: “del momento t_1 hasta el momento t_2 ha ocurrido el evento p ”. Aquí, p puede ser una conversación, una melodía o cualquier otra cosa, p. ej., un ruido. El evento p , por lo demás, admite diversas formas de mención. Hay cosas que disponen de nombres convencionales; p. ej., Julio César o la Misa en Si Menor de Bach. A estas cosas basta asignarles su propio nombre para referirlas, pero existen también otros procedimientos: p. ej., podemos mencionar la Misa en Si Menor de Bach imitando la melodía (recurso difícil de aplicar en el caso de Julio César). Pero hay cosas que carecen de nombre convencional y que no pueden ser imitadas, o lo son sólo muy incómodamente. En estas situaciones se requiere inventar un nombre. Sea, p. ej., la frase *Roma es una ciudad*. No hay duda que nos hemos referido a la ciudad de Roma. Pero si decimos *Roma tiene cuatro letras*, dejamos de referirnos a esta ciudad. Para verificarlo, basta sustituir *Roma* por un equivalente semántico, como *la capital de Italia*. Después de efectuar esta sustitución en las dos frases de antes, comprobamos que la primera sigue siendo verdadera, pero no la segunda. En ésta, no se trata de la ciudad de Roma, sino de su nombre. Tales dificultades se superan habitualmente poniendo entre comillas lo que se refiere, no a un objeto, sino al nombre mismo del objeto. Así, pues, en lugar de escribir $t_1 t_2 \text{ Fit } p$, escribiremos $t_1 t_2 \text{ Fit } "p"$. Analicemos más de cerca este procedimiento. Para decir de una cosa que ha ocurrido del momento t_1 al momento t_2 , escribimos

$$\forall x \ t_1 t_2 \text{ Fit } x.$$

Supongamos, sin embargo, que el evento considerado consiste en que alguien ha pronunciado la palabra x . Para relatar este hecho, escribiremos

$$t_1 t_2 \text{ Fit } x,$$

donde x , contra toda apariencia, ya no es una variable. En todo caso, resulta incómoda la coexistencia de las dos expresiones aducidas hasta aquí. Puede evitarse, sustituyendo la última expresión por

$$t_1 t_2 \text{ Fit } "x".$$

Queda aclarado, de este modo, que esta x no requiere ninguna cuantificación puesto que ya no figura como variable en la fórmula. Como vemos, en este tipo de situaciones el artificio de las comillas resuelve el problema. Pero sólo un reducido número de problemas puede acogerse a una solución tan simple.

Consideremos, p. ej., eventos de tipo pxq , donde p y q son partes invariables, mientras x representa un número cualquiera. Como nombre ambiguo para eventos de este tipo ($p12q$, $p28q$, $p6q$) se podría elegir pxq , pero entre comillas. Pero ello no es suficiente. En efecto, " pxq " no es el nombre ambiguo de eventos del tipo $p12q$, sino el nombre del evento pxq . Para resolver este inconveniente, Quine inventó las "quasi-comillas" (notadas mediante $\ulcorner \urcorner$), con las cuales la relación de antes se escribe

$$\forall x \ t_1 t_2 \text{ Fit} \ulcorner pxq \urcorner.$$

Según Quine, $\ulcorner pxq \urcorner$ es un nombre ambiguo para eventos como $p12q$ ⁵³.

Pero ni siquiera las quasi-comillas aportan la solución definitiva al problema, ya que, p. ej., en un texto puede haber palabras con apariencia de variables, introducidas, en realidad, a causa del carácter mismo del relato. En este caso, las comillas atraen la atención sobre la existencia de unas variables no auténticas, producidas por la situación, pero no especifican cuál es la parte no auténtica.

Podría convenirse en que las variables no auténticas, debidas a la situación, se escribiesen en letras góticas. Pero entonces, si queremos referirnos a textos que contengan variables auténticas y no auténticas al mismo tiempo, surge la necesidad de recurrir a un nuevo tipo de caracteres, digámosle "hipergótico". De este modo, cada uno de los grados de complejidad en el habla reclamaría la introducción de un nuevo alfabeto, con lo que se sobrecargaría enormemente el lenguaje. Hay que rechazar, pues, esta suerte de solución debido a sus inmodisimas consecuencias.

Hay, sin embargo, otra salida, mucho más sencilla. De momento, renunciamos al uso de las comillas a fin de sustituirlas por la cita a base de "autonomía", es decir, por copia o reproducción. Si el evento citado por autonomía está dotado de significado y tiene una composición racional (p. ej., una exposición coherente), entonces no hay dificultad alguna. Pero si, por el contrario, el evento citado fuese algo incoherente, p. ej., un ruido, entonces lo único que interesa es saber si este evento falto de significado, incoherente, es o no auténtico. No hay dificultad en descubrir qué porciones de un relato son auténticas y cuáles obedecen a la introducción de unas variables no auténticas como consecuencia de las exigencias del relato. En cuanto, en un texto gobernado por "Fit", aparecen sustituciones, dicho texto adquiere coherencia desde el punto de vista contextual. Una parte no auténtica corresponde a una variable que no pertenece al dominio de "Fit", sino que viene impuesta desde fuera, y, por tanto, debe cuantificarse al margen del dominio de "Fit". Y a la inversa, si se cuantifica una variable fuera de este dominio, la variable es no auténtica, aparente. Al mismo tiempo, si al margen de "Fit" se impone una sustitución al texto gobernado por "Fit", en-

53. Cf. QUINE, "Logic".

tonces se recomienda no elegir una variable semejante a una parte auténtica del texto gobernado por este "Fit".

Volviendo a los ejemplos de antes, comprobamos que

$$t_1 t_2 \text{ Fit } pxq$$

es el relato de un evento pxq , por cuanto x está no cuantificado y es, pues, no auténtico. Del mismo modo,

$$t_1 t_2 \text{ Fit } \forall_x pxq$$

es el relato de un evento $\text{Fit} \wedge_x pxq$, ya que la cuantificación ha tenido lugar en el dominio de "Fit". (No interesa aquí si el evento que se relata es o no coherente o si tiene significado o no.) Por el contrario, en

$$\forall_x t_1 t_2 \text{ Fit } pxq,$$

x del dominio de "Fit" no puede ser auténtico, puesto que se halla cuantificado fuera de este dominio. En lo que atañe a pxq , aquí se comporta como un nombre ambiguo para eventos de tipo $p12q$, donde p y q son partes constantes; esta situación contrasta con la de más arriba, donde pxq era el nombre de un evento pxq .

En el lenguaje común se evitan las construcciones que presentan una complejidad lógica demasiado grande; se accede, a lo sumo, a tres o cuatro niveles. En cambio, el lenguaje artificial admite construcciones de una complejidad muy superior. En cada nivel pueden aparecer variables no auténticas impuestas por las necesidades del nuevo nivel. Para hallar la proveniencia de una variable no auténtica, basta controlar el lugar en que está cuantificada (por generalización, por existencia, por artículo, por signo de interrogación o por otro procedimiento).

El habla indirecta del lenguaje común se reduce al hecho de que las variables como *yo, tú, él, aquí, allí, ahora, ayer*, etc. pertenecen a un nivel superior, a veces difícil de identificar. No es necesario imitar esta singularidad idiomática cuando operamos con un lenguaje artificial. Al contrario, es natural postular que las variables de este tipo relativas a una persona, un lugar, un momento —por el simple hecho de que alguien las pronuncia y no por el uso de un signo lingüístico formal— no pueden pertenecer al discurso del que las pronuncia de un modo directo (cf. 6.2 y ss.).

La solución propuesta se encuentra cerca de los procedimientos empleados en las lenguas artificiales. A menudo, para resolver un problema relativo a las lenguas artificiales, se ensayan cantidades de métodos complicados. No obstante, en lo que sigue, comprobaremos que la solución viene indicada por las sencillas sugerencias que ofrecen las lenguas naturales, con la condición de aplicar una formalización determinada. El estudio de las lenguas naturales constituye, en la opinión de Freudenthal, una fuente particularmente rica en recursos y sugerencias aplicables a las lenguas artificiales.

3.28. Los modelos matemáticos en el estudio de la literatura

La idea de utilizar métodos matemáticos en el estudio de la literatura no es precisamente nueva. Ya Charles Baudelaire observaba que, para profundizar en el espíritu de un poeta, hay que buscar en su obra las palabras que aparezcan con mayor frecuencia. Este tipo de observaciones abrieron el camino a la aplicación de los métodos estadísticos en la investigación poética. El matemático A. Markov, partiendo del modo en que se alternan las vocales y las consonantes en *Evgueni Oneguín* de Pushkin, introdujo, a principios de nuestro siglo, el concepto, hoy célebre en el cálculo de probabilidades, denominado precisamente “cadena de Markov”. La importancia de los aspectos combinatorios en dramaturgia también ha sido apreciada desde hace mucho. Étienne Souriau hablaba, en 1950, sobre la necesidad de construir un “álgebra del teatro”⁵⁴. Diez años más tarde, Paul Ginestier elaboraba una “geometría de los personajes”⁵⁵.

Durante los últimos decenios, la matemática ha empezado a penetrar en el estudio de la poesía no sólo en sus aspectos cuantitativos, sino también en sus caracteres estructurales. Un precursor importante de la poética matemática estructural es el sabio francés de origen rumano Pius Servien (Pius Șerban Coculescu), el cual en la cuarta década del presente siglo confeccionó una representación estructural de la oposición entre el lenguaje lírico y el lenguaje científico⁵⁶. Mientras tanto, estas investigaciones progresaron y hoy cabe hablar ya de la Poética matemática como de un dominio interdisciplinario bastante bien perfilado⁵⁷.

La poética matemática es una provocación amistosa que los matemáticos dirigen a los poetas, críticos literarios, esteticistas y estilistas. Se propone, como toda obra poética, explicar la naturaleza del lenguaje poético y profundizar en su estructura. Pero su intervención es inédita. Naturalmente, toda investigación intenta alcanzar la mayor adherencia posible al objeto. Ahora bien, en el caso de la poesía, una tal adherencia puede realizarse de dos maneras; dicho en otros términos: hay dos formas de discutir sobre poesía. Una primera modalidad, la más habitual, consiste en realizar la adherencia al objeto a base de utilizar, como lenguaje de discusión —y, por tanto, como metalenguaje—, un lenguaje lo más cercano posible al lenguaje-objeto, en este caso, el poético. Este modo de abordar la poesía queda ilustrado por la crítica literaria en tanto que género propiamente literario. Pero existe además una alternativa a esta vía: en lugar de realizar la adherencia al objeto a base de su imitación, con el riesgo de confundir el metalenguaje con el lenguaje-objeto y perder de vista la diferencia entre el texto

54. Cf. SOURIAU, *Deux cent mille situations*.

55. Cf. GINESTIER, *Le théâtre*.

56. Cf. SERVIEN, *Le langage y Esthétique*; [MARCUS, *Gindirea*].

57. Cf. MARCUS, *Poetica matematică*.

en discusión y el texto con que se discute, cabe ensayar una delimitación, lo más rigurosa posible, entre los dos lenguajes. En el límite, obtenemos una situación por la que el metalenguaje empleado en la investigación de la poesía es el lenguaje de búsqueda, que se opone, con el mayor rigor, al lenguaje poético, y que no es otro que el lenguaje matemático. Precisamente esta situación límite es la que consigue la poética matemática.

Sin duda, la adopción de este punto de vista plantea de inmediato algunos problemas delicados. El lenguaje poético y el lenguaje matemático resultan, desde muy diversas perspectivas, incompatibles. Quisiera tan sólo mencionar aquí una de estas incompatibilidades que justamente la poética matemática pone de manifiesto: en el lenguaje matemático existe una concordancia entre la naturaleza de la expresión y la del significado, al presentar ambas un carácter discreto; mientras que en el lenguaje poético aquellas naturalezas se hallan en contradicción, puesto que la primera es discreta y, por tanto, numerable, y la segunda, continua o no numerable. La significación matemática es como un conjunto de puntos aislados, mientras que la significación poética sugiere una línea continua. Ahora bien, más allá de estas incompatibilidades, a un nivel superior de organización, el lenguaje matemático y el poético ofrecen una buena cantidad de rasgos comunes que les hacen muy provechosos recíprocamente. Ambos son lenguajes de búsqueda, de descubrimiento y, precisamente por ello, ambos promueven una máxima concentración de la expresión. Y ambos ponen en movimiento, en la consecución de sus respectivos fines, las figuras fundamentales del lenguaje, tanto las de tipo metafórico como las de tipo metonímico. Así, pues, en la poética matemática la adherencia al objeto se realiza por una vía indirecta, pero particularmente eficiente, porque apunta a los estratos profundos del lenguaje. Se revela con ello un hecho que viene a contradecir una concepción muy frecuente sobre la poesía, según la cual ésta se caracterizaría por la metáfora. La poética matemática desmiente esta tesis, puesto que la función metafórica es inherente a todo lenguaje humano, ya sea natural o artificial, con palabras, con gestos o con miradas, con propósitos científicos, poéticos o rutinarios. Lo que verdaderamente opone el lenguaje poético al lenguaje científico radica en que este último es puramente denotativo y, por consiguiente, rigurosamente regulado por las definiciones, mientras que el primero existe en la medida en que connota, es decir en la medida en que los significados de los términos no están reglamentados a priori, sino que se establecen ad hoc, en los contextos en que aparecen. Claro está que las funciones connotativa y metafórica se superponen a veces. Pero cada una de ellas puede existir con independencia de la otra: el lenguaje matemático es un ejemplo perfecto de metáfora sin connotación, al igual que muchos textos poéticos connotan aun en ausencia de lenguaje figurado.

Ahora bien, todo ello se refiere a la naturaleza del lenguaje poético, considerado desde un punto de vista más abstracto. ¿Qué podrían reportar, sin embargo,

los métodos matemáticos cuando se trata de un texto poético bien determinado? ¿Acaso nos ayudan a comprender algo de su “secreto de fabricación”? Si la poesía estuviera compuesta exclusivamente por elementos singulares, irrepetibles, si las piezas componentes de un poema se sustrajeran a toda organización, la respuesta sería evidentemente negativa. Pero la realidad es otra. Aduzcamos, p. ej., el poema *Ceasul de-apoi*, ‘*La hora postrera*’, de Tudor Arghezi. El poeta organiza la expresión a lo largo de una cadena de sintagmas nominales que mantienen invariablemente el nombre abstracto *ora*, ‘hora’, como elemento iterativo básico del poema, cuyo atributo viene siempre seleccionado desde otra esfera semántica, más o menos apartada de la esfera abstracta de la hora: de metal (*ora de bronz și de fier*, ‘hora de bronce y de hierro’), animal (*ora de catifea*, *ora de pîslă*, *ora de lînă*, ‘hora de terciopelo, hora de fieltro, hora de lana’), vegetal (*ora de hîrtie*, ‘hora de papel’) o pluralidad de reinos (*ora de praf*, ‘hora de polvo’). El transcurso del tiempo queda sugerido por una sucesión de elementos diferenciales injertados al elemento invariante e iterativo de la hora. Es necesario, pues, estudiar esta organización dotada de una complejidad racional que hace imprescindible el empleo de las técnicas modernas de investigación de estructuras; técnicas que proporciona precisamente la matemática.

En lo que concierne al presente ejemplo de Arghezi, estas técnicas están constituidas por el conjunto de conceptos y resultados de la teoría de los árboles, importante capítulo de la teoría matemática de los grafos. Cabe obtener con ello posibilidades muy matizadas para rastrear el modo en que se modifica la separación semántica entre los elementos iterativos y los elementos singulares del poema. Así, p. ej., puede comprobarse que, a lo largo del poema, la distancia paradigmática entre los términos de los sintagmas metafóricos crece, ya que equivale sucesivamente a 4 (abstracto-metal), a 5 (abstracto-animal) y a 6 (abstracto-vegetal). Esta evolución se logra a costa de aumentar la diferencia de generalidad entre los dos términos, que pasa sucesivamente por los valores 2, 3 y 4, mientras que el grado de heterogeneidad se mantiene igual en todas las metáforas consideradas. Ahora bien, los resultados obtenidos también dependen, hasta cierto punto, del modo en que construimos el árbol semántico de referencia. Por desgracia, el actual estadio de los conocimientos relativos a la organización semántica de las lenguas naturales todavía no permite seleccionar una representación semántica óptima, capaz de afrontar, pues, todas las demás representaciones posibles.

Consideremos algunas otras aplicaciones simples, pero significativas, de la poética matemática.

Es bien conocido aquel verso de Eminescu que dice: *Iambii suitori, troheii, săltăreșele dactile*, ‘Los yambos subientes, los troqueos, los saltarines dáclicos’⁵⁸. Șerban Cioculescu observa en alguna parte que el yambo es más grave, más len-

58. [Del poema *Scrisoarea II*, ‘*Segunda carta*’].

to; que el dinamismo requiere dáctilos, mientras que los sentimientos de amargura se manifiestan en los pesados repliegues del troqueo⁵⁹. He aquí, entonces, dos intuiciones que concuerdan en lo que se refiere a las preferencias semánticas que acusan los distintos tipos de metro en poesía. ¿Cómo podríamos confirmar estas intuiciones, estas hipótesis sobre la función que desempeña un esquema rítmico?

En este problema, representa un instrumento eficaz la energía informacional, es decir una analogía informacional de la energía cinética tal como la ha presentado Octav Onicescu⁶⁰. En un sistema informacional dado, formado por elementos a_1, \dots, a_n con probabilidades de aparición p_1, \dots, p_n , la energía informacional del sistema equivale a $p_1^2 + \dots + p_n^2$. En un esquema rítmico, se opera con dos elementos: la sílaba acentuada (larga), representada por una línea (—), y la sílaba no acentuada (corta), representada por un semicírculo (◡). El troqueo consiste en la sucesión de una sílaba acentuada y una sílaba no acentuada y se indica, pues, a base de una línea seguida de un semicírculo: —◡. La aparición, en un troqueo, de cada una de estas dos sílabas es de 1/2, lo que implica, para la energía informacional del troqueo, el valor $(1/2)^2 + (1/2)^2 = 1/2$. El mismo valor energético resulta también para el yambo, que consta de una sílaba no acentuada seguida de una sílaba acentuada: ◡—. Por lo que se refiere al dáctilo, formado por una sílaba acentuada y a continuación dos sílabas no acentuadas, —◡◡, la energía vendrá dada por el valor $(1/3)^2 + (2/3)^2 = 5/9$ y, por lo tanto, será mayor que la del troqueo y del yambo. Esta situación confirma la capacidad de que la energía informacional sea una medida del grado de vivacidad de un esquema rítmico. La intuición de Eminescu sobre el carácter saltarín del dáctilo tiene una base objetiva, expresada por su superior energía informacional. Al mismo tiempo, el troqueo, tan característico de la “doina”⁶¹, y el yambo grave, apropiado a unas meditaciones filosóficas como en *Se bate miezul nopții*, ‘*Suena la media noche*’, del mismo Eminescu, ofrecen una energía informacional inferior, lo que confiere asimismo una base objetiva al fenómeno en cuestión⁶².

Otro ejemplo que quisiéramos considerar se refiere a la sintaxis poética. Es sabido que a menudo la poesía adopta una sintaxis contorsionada y una tópica fuera de lo habitual. ¿Cabe dar una expresión más precisa también a este fenómeno? Existen diversas respuestas a esta pregunta. Nos referiremos tan sólo a dos de ellas. Una alude a la propiedad de proyectividad sintáctica, ya estudiada

59. Cf. CIOCULESCU, “Valori muzicale”.

60. Cf. ONICESCU, “Énergie”.

61. [Composición lírica popular rumana de carácter elegíaco, típicamente con temas de amor, añoranza, tribulación, etc.].

62. Cf. MARCUS, “Types of meters”.

en un párrafo anterior (cf. 3.16), y la otra es la que podemos llamar la propiedad de Tesnière.

Observemos la proposición

Citesc o carte foarte frumoasă, ‘Leo un libro muy bonito’⁶³.

En ella notamos una serie de dependencias sintácticas (directas o indirectas): *o*, ‘un’, depende de *carte*, ‘libro’; *carte*, de *citesc*, ‘leo’; *frumoasă*, ‘bonito’, de *carte* (e, indirectamente, también de *citesc*) *foarte*, ‘muy’, de *frumoasă* (e, indirectamente, también de *carte* y de *citesc*). Se produce, entonces, la siguiente situación: tantas veces como un término *a* de la proposición dependa (directa o indirectamente) de un término *b*, todo término comprendido entre *a* y *b* depende igualmente (directa o indirectamente) de *b*. Así, p. ej., si nos referimos a la dependencia de *frumoasă* respecto de *citesc*, observamos que cada término intermedio —*o*, *carte* y *foarte*— depende asimismo (directa o indirectamente) de *citesc*. A esta propiedad de la frase en cuestión se le denomina **proy ecti v i d a d s i n t á c t i c a**. (El empleo, aquí, de un término geométrico no es casual, pero no hay lugar ahora para justificarlo.) Lo curioso es que en lenguas naturales como en castellano, rumano, francés, ruso, inglés y otras, la mayoría absoluta de las proposiciones poseen esta propiedad, que aparece, así, como un fenómeno de regularidad que gobierna estadísticamente la estructura de dependencia de las proposiciones.

En estas circunstancias, interesa indagar qué significado tienen las proposiciones no proyectivas en tanto que constituyen un fenómeno de excepción. Un examen a la no-proyectividad sintáctica demuestra que una de sus principales fuentes es el efecto poético⁶⁴. Nos contentaremos con un solo ejemplo. En la novela *Întunecare*, ‘Oscurecimiento’, Cezar Petrescu escribe en un momento dado:

O piscă trecu albă pe linia gardului
‘Un gato pasó blanco por la línea del cercado’⁶⁵

Lo que sorprende en esta frase es la situación ambigua de la palabra *albă*, ‘blanco’, puesto que acusa una doble dependencia, una respecto del nombre *piscă*, ‘gato’, y otra respecto del verbo *trecu*, ‘pasó’. En otras palabras, *albă* cumple al mismo tiempo una función adjetiva y una función adverbial. Sin embargo, el verbo *trecu* no depende (ni directa ni indirectamente) del nombre *piscă*, a pesar de hallarse situado entre *piscă* y *albă*. Ocurre, entonces, que esta proposición no es proyectiva.

63. [*Carte frumoasă*, ‘libro bonito’, es femenino en rumano].

64. Cf. MARCUS. “Trois types”.

65. [*Piscă albă*, ‘gato blanco’, es femenino en rumano].

Los estudios estadísticos realizados sobre un gran número de textos han llevado a la conclusión de que la dependencia de un mismo término respecto de por lo menos dos otros términos distintos es, en el ámbito de una proposición, un fenómeno excepcional. La mayor parte de las veces, cada término de una proposición depende a lo sumo de un solo término de la misma proposición. En esto precisamente consiste la llamada propiedad de Tesnière, según el nombre de quien la puso de manifiesto. Observamos, así, que la proposición de Cezar Petrescu viola no sólo la propiedad de proyectividad, sino también la de Tesnière. El incumplimiento de la propiedad de Tesnière desempeña a menudo una función expresiva.

El folklore es, a su vez, otro dominio de gran actualidad para el tratamiento de los métodos matemáticos y lingüístico-matemáticos.

El perfeccionamiento de los medios modernos para la colección de material folklórico mediante registros magnetofónicos así como la proliferación de especialistas, cada vez mejor preparados, en la recopilación de estos materiales ha promovido, durante los últimos años, un mejor conocimiento sobre el creciente número de variantes con que se encuentran diversos cuentos y baladas populares. *Miorița*, p. ej., es conocida en muchos cientos de variantes, y *Balada Meșterului Manole*, 'La balada del maestro Manole', en varias decenas, sin que ello represente ninguna excepción frente a otras piezas populares. Hoy existen mejores posibilidades que en el pasado para conservar la autenticidad del folklore y para preservarlo de la corrupción que le puede afectar en el proceso de registro. En tales condiciones, es cada vez más importante perfeccionar los métodos de clasificación tipológica a fin de comparar y jerarquizar las variantes respectivas. La cantidad de parámetros que interviene en la descripción de un cuento fantástico o de una balada popular es bastante grande. Algunos de ellos se refieren a las condiciones exteriores, tales como el lugar y la fecha del registro, y otros, a la estructura lingüístico-literaria. Para clasificar las producciones folklóricas, en función de estos parámetros, deben realizarse operaciones extremadamente complejas, pero necesarias, ya que es indispensable comprender sus tendencias evolutivas extrayendo la significación de las más características. En Francia, Estados Unidos y en otros países trabajan, en estos problemas, equipos mixtos de folkloristas y matemáticos, que programan en ordenadores electrónicos particularmente perfeccionados distintos procesos algorítmicos para el análisis de las producciones folklóricas⁶⁶.

Pero no debe pensarse que el uso de los ordenadores electrónicos resuelva por sí mismo la gran problemática de la interpretación folklórica y de la definición de los diversos géneros que allí actúan. A este propósito, es imprescindible desarrollar una intensa actividad en el empleo de métodos poderosos elabora-

66. Cf. RICHARD-LÉVY-VIRVILLE, "Essai de description".

dos por la lingüística y la estética moderna y corroborados con procedimientos de formalización obtenidos por la modelación matemática.

En el marco de la escuela rumana de lingüística matemática se ha iniciado una serie de preocupaciones con respecto a la modelación matemática del folklore. Los procedimientos que hemos ensayado en poética matemática para el estudio comparativo de diversos textos han sido utilizados con éxito por el matemático Tudor Bălănescu en la comparación de algunas variantes de la balada *Miorița* y por Ion Rădoi en una comparación de variantes sobre la balada *Meșterul Manole*, 'El maestro Manole'⁶⁷. En ambas investigaciones se ha confirmado útil el método, recientemente elaborado por O. Onicescu, del llamado problema del "leader", y que consiste en encontrar la jerarquía total, de síntesis, que resulta de otras jerarquías sobre los mismos objetos, las cuales se habían obtenido previamente a partir de distintos criterios de comparación aplicados a aquellos objetos⁶⁸.

Otro método, últimamente ensayado por el matemático alemán Walther Fischer, se refiere al empleo del álgebra homológica en el estudio de las estructuras repetitivas de los textos poéticos⁶⁹. Por nuestra parte, hemos sugerido la aplicación de este método al estudio de unas baladas populares rumanas, que se ha concretado ya sobre la *Balada Jianului*, 'La balada de Jianu'⁷⁰.

Una investigación sobre la que quisiéramos insistir un poco más es la que hemos realizado en colaboración con Stanca Fotino⁷¹. Se dice con insistencia, últimamente, que la obra de arte no debe contemplarse como un objeto, sino como un proceso. Esto resulta especialmente cierto en lo que atañe a las obras folklóricas. En efecto, esta clase de obra es un continuo proceso en el que se pasa de una variante a otra, puesto que cada oyente se convierte en un posible autor al modificarla y propalarla nuevamente. En tales condiciones, conviene entender qué tipo de transformaciones explica esta metamorfosis constante, a qué tipo de estructuras, en definitiva, se somete. Es como si existiera una máquina que produce indefinidamente nuevas y nuevas variantes del mismo texto, pero de tal modo, que esta continua novedad tampoco puede sobrepasar los tipos de transformación que la máquina tiene inscritas en sus posibilidades de funcionamiento. Una hipótesis similar se había hecho ya, hace unos 15 años, acerca de la capacidad del cerebro humano para comprender y producir las frases de la lengua natural. Esta hipótesis condujo entonces a la creación, por parte de Noam Chomsky, de la teoría de las gramáticas generativas, definidas preci-

67. Cf. BĂLĂNESCU, "Studiu matematic", y RĂDOI, "Studiu matematic".

68. Cf. ONICESCU, "Procedee de estimare".

69. Cf. FISCHER, "Topologische Stilcharakteristiken" y "Beispiele".

70. Cf. GORUN, "Une application".

71. Cf. FOTINO-MARCUS, "Gramatica basmului" I y II.

samente como una máquina de esta naturaleza capaz de producir frases (cf. 3.12)⁷².

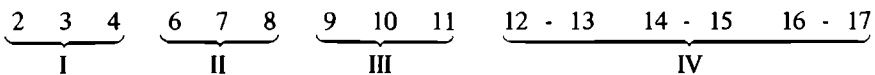
Ello nos ha dado la posibilidad de ensayar la misma teoría generativa en el estudio de varios cuentos populares rumanos.

Prestemos atención al cuento-jácara del tipo "Păcală", *Poveste din lume*, 'Cuento del mundo'⁷³. En él se contiene una serie de episodios, que numeraremos del siguiente modo:

(1) Un muchacho entra al servicio de un pope con la condición de que ambos respetarán un determinado acuerdo. (2) El primer día, el pope trata de inducir al chico a transgredir el acuerdo, pero éste se resiste, (3) el segundo día ocurre lo mismo, (4) pero el tercer día el muchacho sufre un castigo por haber incumplido el acuerdo. (5) El hermano del chico —Păcală— se hace criado del pope con la misma condición. (6) El amo reanuda sus intentos, al primero de los cuales Păcală resiste, (7) al segundo también (8) y también al tercero. (9) Ahora es el amo el que empieza a resistir los intentos de Păcală a propósito de la condición recíproca; (10) y (11) el pope sigue resistiendo dos veces sucesivas. (12) La situación ha cambiado, ya que el pope empieza a ceder ante la constante paciencia de Păcală, hasta que decide marcharse con su hijo para huir de Păcală. (13) Pero éste los engaña y se va tras ellos. (14) Deciden entonces aplicar una nueva táctica para escapar de él, (15) pero Păcală los engaña de nuevo, (16) lo que agota la paciencia del pope y le hace a éste transgredir el acuerdo inicial, (17) con lo que recibe el castigo merecido.

Hay, por tanto, 17 segmentos componentes en el cuento, que llamaremos e v e n t o s. Los eventos 1 y 5 cumplen la misión de introducir un personaje, respectivamente, por lo que pueden eludirse. Los 15 eventos restantes se distribuyen en cuatro grupos, como sigue:

Un primer grupo está formado por los eventos 2, 3 y 4; el segundo, por 6, 7 y 8; el tercero, por 9, 10 y 11, y el cuarto, por las tres parejas 12-13, 14-15 y 16-17.



Cada grupo consta de tres componentes, que denominaremos s e g m e n t o s n a r r a t i v o s. Cada uno de estos segmentos narrativos contiene un intento de transgresión al cual se somete un héroe. Hay una analogía entre 2, 6, 9 y 12-13, referidos a un primer intento, entre 3, 7, 10 y 14-15, referidos a un segundo intento, y entre 4, 8, 11 y 16-17, referidos a un tercer intento. Ahora bien,

72. Cf. CHOMSKY. *Syntactic Structures*.

73. [Păcală es un personaje pícaro, bromista e ingenioso muy común en los cuentos populares rumanos].

esta analogía resulta bastante débil, puesto que, p. ej., 4 y 8 difieren tanto por la identidad del héroe como por el desenlace, positivo o negativo, del intento: 8 se distingue de 11 por la naturaleza del intento, mientras que 11 se distingue de 16-17 por la inversión de papeles. Efectivamente, si la secuencia I presenta como rivales al pope y al hermano de Păcală, las secuencias II y III colocan a éste en lugar de su hermano y la secuencia IV, que mantiene a los mismos rivales, invierte la situación poniendo a prueba al pope y no a Păcală. Surge, entonces, una simetría del proceso narrativo en el sentido de que las fases I y IV permanecen bajo el signo de la oposición entre un personaje dominador (el pope en I, Păcală en IV) y un personaje dominado (el hermano de Păcală en I, el pope en IV), frente a las fases II y III, cuya característica consiste en un equilibrio entre los personajes rivales (Păcală y su oponente).

Simolicemos por medio de *a* la común significación de los segmentos narrativos 2 y 12-13 (primer intento en la oposición entre un dominador y un dominado), por medio de *b* la común significación de los segmentos 3 y 14-15 (segundo intento en la oposición entre un dominador y un dominado) y por medio de *c* la común significación de los segmentos 4 y 16-17 (tercer intento en la oposición entre un dominador y un dominado). Añadamos los símbolos a_1 , b_1 y c_1 para indicar la común significación de los segmentos 6 y 9, 7 y 10, y 8 y 11, respectivamente (primero, segundo y tercer intento en la confrontación equilibrada). El desarrollo del cuento recibe, así, la siguiente representación simbólica:

$$a \ b \ c \ a_1 \ b_1 \ c_1 \ a_1 \ b_1 \ c_1 \ a \ b \ c$$

Aquí, las únicas oposiciones que nos interesan son: Păcală / no-Păcală y dominado / no-dominado.

La primera fuente de transformación se refiere a la repetición, el mismo número de veces, de las secuencias en que se enfrenta un dominador y un dominado. Repitiendo *n* veces estas secuencias, obtenemos una representación de forma:

$$(a \ b \ c)^n \ a_1 \ b_1 \ c_1 \ a_1 \ b_1 \ c_1 \ (a \ b \ c)^n \ (n = 1, 2, \dots)$$

Evidentemente, existen también otras fuentes, que no consideramos aquí. Pero las dos representaciones aducidas resumen, a un determinado nivel de abstracción, dos posibilidades de evolución del cuento respectivo. Otras particularidades —demasiado técnicas para describirlas ahora— ofrecen la posibilidad de construir “máquinas generativas” de las representaciones que acabamos de hacer, máquinas que proporcionan el proceso de una virtual metamorfosis en el cuento.

Entre las muchas consecuencias que cabe extraer a partir de esta perspectiva y que conducen a una verdadera tipología generativa de los cuentos populares rumanos, quisiéramos poner de manifiesto al menos una. Se sabe que la estructura de estos cuentos está dominada por una tendencia a la simetría (la cual, en

nuestro cuento, ha quedado bien demostrada y, por cierto, en relación con los números mágicos que rigen la acción: en el caso presente, el número 3) y a la repetición de unos mismos tipos de episodios. Podría pensarse que ambas tendencias son convergentes, con lo cual la simetría sería un caso particular de repetición. La investigación que hemos emprendido ha puesto de manifiesto que cuanto mayores son las coerciones de simetría (en el sentido de unos condicionamientos a distancia entre los elementos del cuento), tanto más complicada es la máquina generativa que explica la evolución del cuento.

Claro que estas investigaciones se hallan todavía al principio y precisan de progresos ulteriores. Pero los ejemplos analizados manifiestan que la literatura folklórica en especial y el arte poética en general consisten en una alternancia equilibrada de singularidades y repeticiones.

Decir que la poesía sólo se caracteriza por la presencia de elementos únicos, y que los elementos repetibles pertenecen a sus extramuros, equivale a entender de una manera simplista la creación poética. La singularidad no puede valorizarse si no es en el contexto de unos procesos recursivos y organizativos, del mismo modo que la reiteración de motivos y procedimientos no degenera en prosaísmo cuando se injerta a elementos de un valor único.

Observamos que los elementos efectivamente estudiados por los métodos matemáticos no son los irrepitibles, únicos, sino los elementos iterativos junto con sus relaciones respecto a los elementos singulares. En otros términos, el análisis estructural permanece indiferente ante las entidades en sí mismas de la poesía, por cuanto se concentra sobre las relaciones que estas entidades contraen.

Llegamos, así, a uno de los problemas fundamentales para la comprensión del lenguaje artístico. El núcleo de este lenguaje resulta inaccesible al análisis. Pero esta situación no es típica de las investigaciones sobre el lenguaje poético. En muchas de las ciencias contemporáneas, empezando por la física nuclear y terminando por el estudio de las actividades nerviosas superiores, encontramos situaciones análogas, que ponen a los investigadores en la tesitura de imaginar vías de acceso al objeto investigado muy indirectas. Los cibernéticos han sugerido una metáfora, la de la "caja negra", es decir un dispositivo en cuyo interior no sabemos qué sucede, pero que se halla al propio tiempo sometido a la acción de unos estímulos exteriores conocidos (las llamadas "señales de entrada") que determinan en él un comportamiento concretado en unas ciertas "señales de salida", igualmente conocidas. Investigando las señales de entrada y de salida, cabe simular, con la ayuda de unos modelos, la actividad interna del dispositivo. El cerebro humano es un perfecto ejemplo de "caja negra". Y también el lenguaje artístico lo es, en su estructura profunda. Pero ello no debe descorazonarnos, ni inducirnos a renunciar al análisis, ya que, a lo sumo, nos advierte que hemos de imaginar determinados caminos indirectos, acaso tan tortuosos, que a lo largo de su trayectoria lleguen a alejarnos del objetivo buscado.

Al hombre de ciencia se le exige cada vez más mucha “astucia” para poder “engañar” a una naturaleza que muy difícilmente se deja penetrar.

A este propósito, saber evitar ciertas preguntas demasiado arduas (del tipo “¿qué es la poesía?”), saber plantear aquellos problemas para los cuales la ciencia de nuestro tiempo está capacitada y saber aprovechar de la mejor manera los resultados de las otras ciencias y los métodos que ellas ponen a disposición de todos son otras tantas manifestaciones de sabiduría particularmente favorables incluso para la investigación del lenguaje artístico. Una tal investigación no sólo requiere respuestas nuevas a viejas preguntas, sino también una forma nueva de plantear los problemas, es decir de preguntar.

3.29. Algunos problemas especiales de la modelación matemática en lingüística

Para facilitar la argumentación que sigue, trataremos de esquematizar algunos aspectos de la modelación. Sea A el original y B el modelo de A . Sea α el conjunto de rasgos de A que pueden ponerse en correspondencia con determinados rasgos de B , los cuales, a su vez, formarán un conjunto que llamaremos α' . Este conjunto está contenido en el conjunto β de todos los rasgos de B . Si B es un objeto matemático, entonces es posible estudiar B por métodos deductivos. Y si este estudio utiliza sólo propiedades de α' , es, entonces, evidente que los resultados obtenidos proporcionarán, en base a la correspondencia entre α' y α , algunas informaciones sobre A . Es normal suponer que, en la medida en que la investigación deductiva de B aprovecha también propiedades de $\beta - \alpha'$, los resultados de esta investigación tendrán un significado reducido en lo que atañe al original A . Por otra parte, los resultados son tanto más significativos para A cuanto más pequeña sea la cantidad de rasgos esenciales de A que no pertenezcan al conjunto α . De hecho, el proceso de modelación matemática, al menos en las disciplinas humanísticas, es mucho más complicado de lo que sugiere este esquema.

Tomemos, p. ej., el problema de modelar el concepto de parte del discurso. Aquí es difícil precisar qué es lo que debe modelarse, en qué consiste el original A . Así, uno de los rasgos de las partes del discurso en casi todas las lenguas —en castellano y en rumano, relativamente menos; pero en inglés, en gran medida— consiste en la posibilidad de que una misma palabra corresponda a más de una parte del discurso. Esto, sin embargo, se refiere a un tratamiento intuitivo o empírico respecto a la investigación de las partes del discurso. Pero es evidente que una modelación de las partes del discurso que se proponga registrar incluso la pertenencia de una palabra a más de una parte requerirá, por este simple hecho, un modelo bastante complicado (cf. 7.2). Por otro lado, no está claro si la propiedad de que una palabra pertenezca a más de una parte no deba conside-

rarse inherente al concepto de parte del discurso, ya que este mismo concepto es oscuro y, a su vez, el objetivo de la modelación es contribuir a que quede delimitado. Es problemático, entonces, decidir si la pertenencia de una palabra a más de una parte debe o no ponerse en correspondencia con un rasgo de α' , en una modelación más adecuada sobre las partes del discurso. Aquí se produce un proceso de alguna manera inverso al proceso de aproximación al original por medio del modelo: el concepto que modelamos debe revisarse justamente a partir de los resultados obtenidos en la modelación. Algo similar ocurre en la modelación de la categoría gramatical de género. Partiendo de una concepción intuitiva sobre esta categoría, se construye un modelo matemático que empieza del siguiente modo (damos una presentación no formalizada y muy simplificada de los hechos); un nombre x es gramaticalmente masculino si para toda forma flexiva y de x y para toda forma flexiva z de *bărbat*, 'hombre', existe una forma de *bărbat* que admita los mismos contextos de y o bien que exista una forma de x que admita los mismos contextos de z . Sustituyendo, en esta formulación, la palabra *bărbat* por la palabra *femeie*, 'mujer', se obtiene la definición de los nombres de género gramatical femenino. Por fin, los nombres que no son masculinos ni femeninos, son neutros⁷⁴. Pese a que este modelo se ha servido del aspecto intuitivo de la categoría del género, él mismo permite desprendernos de este aspecto, con lo que la categoría del género queda replanteada a partir de esta modelación. En efecto, del desarrollo ulterior del modelo se infiere, mediante la definición de nuevos géneros, que los géneros gramaticales dependen del conjunto de los contextos con que se opera y que sólo en el caso particular en que este conjunto esté formado por los contextos donde un nombre vaya precedido o seguido por un adjetivo, los géneros gramaticales obtenidos por modelación coinciden con los habituales.

Aquí se manifiesta una nueva faceta acerca de la contradicción dialéctica del proceso de modelación. Por una parte, el modelo debe tener en cuenta la base empírica del fenómeno modelado; por otra, la eficacia del modelo se muestra precisamente en su capacidad de superar esta base empírica.

3.30. Perspectivas y significación

Hasta la actualidad, la modelación matemática se ha aplicado más bien a los aspectos sincrónicos de la lengua. Pero también han empezado a aparecer trabajos

74. [En rumano, los nombres neutros concuerdan como masculinos en singular y como femeninos en plural:

	Sing.	Plur.
Masc. (hombre):	<i>bărbat frumos</i>	<i>bărbați frumoși</i>
Fem. (mujer):	<i>femeie frumoasă</i>	<i>femei frumoase</i>
Neut. (rio):	<i>riu frumos</i>	<i>riuri frumoase</i>].

de modelación probabilística sobre ciertos fenómenos diacrónicos; entre ellos, p. ej., para establecer la etimología o el parentesco entre lenguas (cf. 3.22) En principio, no cabe fijar ningún tipo de límite apriorístico a la modelación matemática. Pero es esencial observar que toda dogmatización absoluta del método de modelación resulta perjudicial y se convierte en una fuente de interpretaciones y pervertimientos idealistas. El método de la modelación es sólo un auxiliar de la lingüística y, como tal, ha de completarse con los demás métodos probadamente fecundos en el estudio de la lengua. En particular, la modelación matemática de la lengua no impide, sino que facilita la averiguación de las relaciones entre la lengua y la sociedad, entre la lingüística y otras ciencias sociales, desde el momento en que posibilita un conocimiento más profundo sobre las estructuras lingüísticas y, por este medio, una visión más clara sobre las semejanzas y las diferencias que hay entre la lengua y los demás fenómenos sociales. A este tenor, nos parecen injustificados los temores de Francis J. Whitfield al referirse al “abismo entre la lingüística y la humanidad” y al “divorcio entre la lingüística y la investigación literaria”⁷⁵.

75. Cf. *WHITFIELD*, “Criteria”, pp. 581-2.

4. La modelación algorítmica de la lengua

4.1. Introducción

El concepto de *algoritmo* es una de las nociones fundamentales de la matemática. Por el momento, diremos tan sólo que un algoritmo es una sucesión determinada de operaciones que, partiendo de unos ciertos datos, promueve la resolución de un problema dado. Bien es verdad que este modo de expresarse no corresponde al espíritu riguroso de la matemática, pero hasta el siglo presente no ha existido una definición más precisa en torno al concepto de algoritmo. En la Edad Media se atribuía este nombre a la regla con cuya ayuda se efectuaba una de las cuatro operaciones aritméticas en el sistema de numeración decimal. Este tipo de reglas fue indicado, en el siglo ix, por el matemático usbeo Horezmi (en árabe Al-ḥwārizm, sobrenombre de Abu Yáfar Abenmusa), por cuyo nombre se designaron en Europa las reglas de ejecución de los cálculos aritméticos. Más tarde, la forma primitiva “algorismo”, “alguarismo” (de donde “guarismo”), se ha convertido en “algoritmo” por una probable influencia de la palabra griega *arizmós*, ‘número’¹.

Durante los últimos veinte años, el concepto matemático de algoritmo ha quedado gradualmente elaborado en los trabajos de lógica matemática de S. Kleene, A. Church, A. Turing, E. Post y sobre todo en la obra de A. A. Markov, donde se emprende una investigación directa y sistemática acerca de la noción de algoritmo en toda su extensión². Otras contribuciones más recientes a la definición de este concepto y a la construcción de una teoría pertenecen a A. N. Kolmogorov y a otros³.

1. Cf. *Bolshaia Sovetskaia Entsiklopedia*, vol. I, palabra *algoritmo*.

2. Cf. KLEENE, *Introduction*; CHURCH, *The Calculi*; TURING, “Computable numbers”; POST, “Enumerable sets”, y MARKOV, *Teoria algoritmov*.

3. Cf. KOLMOGOROV, “O poniatii algoritma”.

La importancia creciente de la noción de algoritmo se vincula, por una parte, al método constructivo en matemática y, por otra, al desarrollo de la cibernética⁴. El estudio cibernético de los procesos de ordenación es imposible sin una descripción algorítmica de dichos procesos⁵. Desde este punto de vista, la noción de algoritmo interesa no sólo a matemáticos, sino también a físicos, ingenieros, biólogos, psicólogos, lingüistas y economistas.

4.2. El algoritmo de la sustracción

Como los algoritmos más frecuentes se relacionan con las operaciones aritméticas elementales, empezaremos con un análisis lógico de una de estas operaciones. Tomemos, p. ej., la operación de restar dos números más pequeños que 100. Si se trata de una resta de tipo $53 - 48$, la operación se realiza de un modo diferente a cuando se trata de una resta de tipo $53 - 21$, puesto que en el primer caso la cifra de las unidades del minuendo es inferior a la correspondiente del sustraendo. Más en general, consideremos dos números ab y cd (donde a, b, c y d representan cifras del 0 al 9) y supongamos que ab es mayor que cd . Para analizar la operación de la resta de cd a partir de ab , la descompondremos en las fases elementales que la conforman. El resultado será, evidentemente, inferior a 100 y, por tanto, compuesto por un máximo de dos cifras. Si notamos por medio de f la cifra de las unidades y por e la cifra de las decenas, la operación de resta se descompone en las siguientes “fases elementales”:

- 1) verificar si d es mayor que b ; en caso negativo,
- 2) efectuar la resta $b - d$;
- 3) escribir $f = b - d$;
- 4) efectuar la resta $a - c$;
- 5) escribir $e = a - c$ y concluir la operación.

Si d es mayor que b , entonces las fases elementales 2), 3), 4) y 5) se sustituyen por:

- 2') efectuar la suma $b + 10$;
- 3') efectuar la resta $(b + 10) - d$;
- 4') escribir $f = (b + 10) - d$;
- 5') efectuar la resta $a - 1$;

4. Se trata de la matemática constructiva que han desarrollado intuicionistas como Heyting y matemáticos como A. A. Markov.

5. La descripción algorítmica de un proceso de ordenación permite que su ejecución sea tomada por la máquina.

- 6') efectuar la resta $(a - 1) - c$;
 7') escribir $e = (a - 1) - c$ y concluir la operación.

Hemos obtenido, en total, 11 fases elementales. La número 1) consiste en la verificación de una condición lógica dada; en este caso concreto, la de que d sea mayor que b . Esta condición, que indicamos por medio de C_1 , admite dos valores lógicos, el de "verdadero" y de "falso". Según que se cumpla el primero o el segundo de estos valores, se aplican en adelante las fases elementales 2'), 3'), 4'), 5') 6') y 7') o bien las fases 2), 3), 4) y 5). Todas estas fases elementales son de una índole muy distinta respecto de la fase 1): las llamaremos operadores y las representaremos por medio de $O_2, O_3, O_4, O_5, O_6, O_7$ y O_2, O_3, O_4, O_5 , respectivamente. Con estas notaciones, el esquema lógico para la operación de sustraer cd de ab se representa así:

$$C_1 \begin{array}{c} \uparrow \\ O_2, O_3, O_4, O_5, O_6, O_7, \\ \downarrow \\ O_2, O_3, O_4, O_5. \end{array}$$

Si C_1 se cumple, entonces se aplica el operador situado inmediatamente a la derecha, es decir O_2 , y a continuación los operadores O_3, O_4, O_5, O_6 y O_7 , de izquierda a derecha. El operador O_7 contiene la indicación de concluir la operación. Si C_1 no se cumple, entonces el vector de índice 1 hacia arriba de la derecha de C_1 anuncia que no se aplica el operador que sigue a continuación, es decir O_2 , sino que se pasa a la aplicación del operador colocado inmediatamente a la derecha del vector de índice 1 hacia abajo, es decir O_2 . Luego se aplica O_3, O_4 y O_5 . Este último contiene la indicación de concluir la operación.

4.3. El algoritmo de Euclides

Consideremos dos números, p. ej., 360 y 252. Si nos proponemos hallar el máximo común divisor (m. c. d.) de ambos números, es decir el número más grande posible que divida exactamente tanto a 360 como a 252, utilizaremos la siguiente regla: se divide 360 por 252; si el resultado fuese exacto, entonces el m. c. d. sería justamente 252. En nuestro caso, sin embargo, no ocurre así, ya que se obtiene un resto de 108. Se divide luego 252 por 108; también aquí, si el resultado fuese exacto, el m. c. d. sería 108, pero en esta operación aparece un nuevo resto de 36. Dividiendo ahora 108 por 36, comprobamos que ya no hay resto ninguno. Resulta, pues, que 36 es el m. c. d. de los números 360 y 252.

Generalicemos ahora estas consideraciones. Sean A y B dos números enteros positivos ($A > B$). Sea R_1 el resto de la división de A por B ; R_2 , el resto de la división de B por R_1 ; R_3 , el resto de la división de R_1 por R_2 ; ..., R_{n+1} , el resto de la división de R_{n-1} por R_n , y así sucesivamente hasta que obtengamos un resto igual a cero. Si $R_1 = 0$, entonces el m. c. d. de los números A y B es

precisamente B ; si no, el último resto distinto de 0 es el m. c. d. de los números A y B . Podemos emplear, para describir estas reglas, el siguiente esquema:

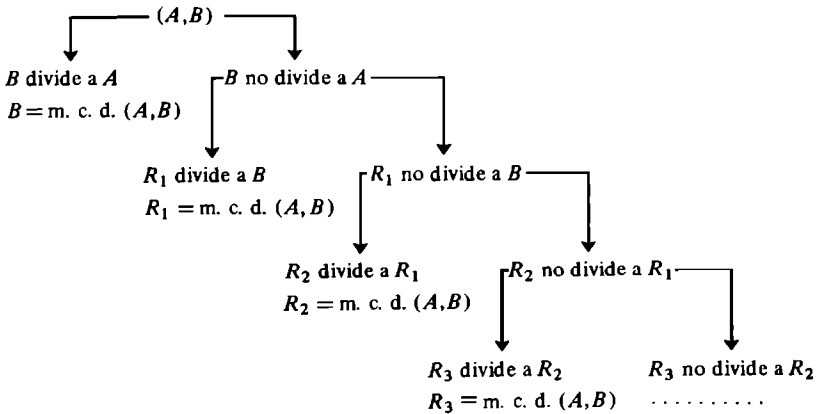


Figura 9

También este esquema se compone de un determinado número de fases elementales. Intervienen, por una parte, diversas operaciones (operadores): como el operador O_0 , que consiste en dividir A por B ; el operador O_1 , que divide B por R_1 ; el operador O_2 , que divide R_1 por R_2 , etc., y el operador O_α que consiste en igualar el m. c. d. con el divisor de la última división y en concluir el algoritmo. Intervienen, por otra parte, ciertas opciones elementales (opciones entre “si” y “no”), a las que hemos llamado condiciones lógicas. En efecto, hay condiciones lógicas como C_0 que consiste en elegir entre la situación $R_1 = 0$ y la situación $R_1 \neq 0$; la condición lógica C_1 , que consiste en elegir entre la situación $R_2 = 0$ y la situación $R_2 \neq 0$; la condición lógica C_2 , que consiste en elegir entre la situación $R_3 = 0$ y la situación $R_3 \neq 0$, y así sucesivamente. Para simplificar la exposición, diremos que C_i se verifica siempre que $R_{i+1} = 0$; en caso contrario, diremos que no se verifica. El algoritmo de Euclides puede representarse ahora del siguiente modo:

$$O_0 C_0 \uparrow O_\alpha \downarrow_1 O_1 C_1 \uparrow O_\alpha \downarrow_2 O_2 C_2 \uparrow O_\alpha \downarrow_3 O_3 \dots O_n C_n \uparrow O_\alpha \downarrow_{n+1} O_{n+1} \dots$$

El algoritmo empieza al poner en marcha el operador O_0 . Si C_0 se verifica, entonces se aplica el operador O_α y el algoritmo se concluye. En caso contrario, los vectores de índice 1 envían al operador O_1 y luego a la comprobación de C_1 . Si C_1 se verifica, entonces se aplica el operador O_α y el algoritmo se concluye; en caso contrario, siguiendo los vectores de índice 2, se pone en marcha el

operador O_2 , y así sucesivamente. En resumen, tanto aquí como en el algoritmo de la sustracción, el vector de índice i hacia arriba muestra que si C_{i-1} no se verifica hay que continuar poniendo en marcha el operador situado inmediatamente a la derecha del vector de índice i hacia abajo.

4.4. Otros algoritmos numéricos⁶

Los algoritmos que, como los examinados hasta ahora, se basan en las cuatro operaciones aritméticas fundamentales se denominan **algoritmos numéricos**. El lector mismo puede tratar de efectuar la descomposición en operadores y en condiciones lógicas de los algoritmos relativos a las operaciones de adición, multiplicación y división. Otro ejemplo de algoritmo numérico puede aplicarse a resolver un sistema de dos ecuaciones de primer grado con dos incógnitas, en que, por tanto, partiendo del sistema

$$\begin{cases} a_1x + b_1y = c_1 \\ a_2x + b_2y = c_2 \end{cases}$$

se llegue a la solución

$$x = \frac{c_1b_2 - c_2b_1}{a_1b_2 - a_2b_1}, \quad y = \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1}$$

También la operación de extraer una raíz cuadrada constituye un algoritmo, que el lector mismo puede descomponer en operadores y en condiciones lógicas.

Una clase de problemas matemáticos de un tipo determinado se considera **algorítmicamente resoluble** siempre y cuando quepa aplicar un algoritmo que conduzca a la solución de cualquier problema de la clase dada. Todos los ejemplos examinados antes constituyen ejemplos de clases de problemas algorítmicamente resolubles.

Pero también existen situaciones en que, dentro de una clase de problemas del mismo tipo, hay que utilizar, para cada problema concreto, un procedimiento especial y característico del problema dado. En este caso, decimos que se trata de una clase de problemas **algorítmicamente irresoluble**. Como ilustración, consideremos el conjunto de las ecuaciones diofánticas, es decir de las ecuaciones de forma

$$P = 0$$

donde P es un polinomio con coeficientes enteros. Aquí entran ecuaciones como

$$x^2 + y^2 - z^2 = 0, \quad 6x^{18} - x + 3 = 0,$$

6. Cf. TRAJTENBROT. *Algoritmy i mashinnoe*.

en las que cabe un número arbitrario —pero finito— de incógnitas. La primera de estas ecuaciones admite las soluciones enteras

$$x = 3, \quad y = 4, \quad z = 5;$$

la segunda, en cambio, no, ya que

$$6x^{18} > x - 3$$

para todo valor de x^7 .

En el año 1901, en el Congreso internacional de los matemáticos, que tuvo lugar en París, el gran matemático David Hilbert presentó 20 problemas no resueltos hasta aquel entonces y cuya solución revestía la mayor importancia para el desarrollo ulterior de la matemática. Uno de los problemas, denominado desde aquella ocasión el “décimo problema de Hilbert”, se enuncia así:

“Elaborar un logaritmo por el cual sea decidible, para toda ecuación diofántica, si tiene o no soluciones enteras.”

Para el caso concreto de las ecuaciones diofánticas con una sola incógnita se conoce, efectivamente, un logaritmo de este tipo. Es decir, se establece que si la ecuación

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0$$

con coeficientes enteros, admite como raíz entera a x_0 , entonces a_0 se divide exactamente por x . Para elaborar el algoritmo correspondiente, habrá que tener en cuenta, pues, las indicaciones siguientes: al principio, se establecen todos los divisores de a_0 . Para cada uno de estos divisores se establece el valor numérico correspondiente del primer miembro de la ecuación de más arriba; si este valor numérico equivale a 0, entonces el divisor respectivo es una raíz (entera) de la ecuación.

Sin embargo, el décimo problema de Hilbert no ha recibido todavía una resolución. Más aún, es lícito sospechar que no existe un algoritmo del tipo que requería Hilbert. La explicación cumplida del sentido de esta afirmación sobrepasa, empero, el marco de nuestros propósitos⁸.

7. [Nótese que la primera ecuación es el teorema de Pitágoras. Las ecuaciones diofánticas forman un grupo especial dentro de las ecuaciones indeterminadas y se caracterizan porque sus soluciones sólo admiten números enteros (hay algunas otras restricciones estructurales que aquí omitimos)].

8. [Como se ha dicho, se conoce un algoritmo de resolución para las ecuaciones diofánticas de tipo $ax + by = c$. Una ecuación diofántica muy sencilla es, p. ej.:

$$x + 2y = 3.$$

Como se ve, en ella se trata de hallar dos números (x, y) que cumplan la condición de que la suma del primero con el doble del segundo sea equivalente a 3, siendo x e y números enteros. Son infinitos, en efecto, los pares de números enteros que satisfacen esta condición (1 y 1, 9 y -3, etc.). Pero ello no ocurre, p. ej., entre 2 y 4, ya que $2 + 2 \cdot 4 \neq 3$ y, en todo caso, para $x = 2$, y debería ser un número no entero ($y = 0,5$), lo

4.5. La resolución de la homonimia morfológica

Consideremos el sustantivo rumano *omului*, ‘del/al hombre’. Nos proponemos, utilizando las indicaciones de los distintos tipos de contexto en que puede aparecer esta palabra, dar una regla que determine en qué caso gramatical se emplea (se trata de resolver, pues, la alternativa genitivo / dativo)⁹.

Si *omului* aparece situado detrás de un nombre con artículo determinado, de una preposición distinta de *mulțumită*, *datorită* y *grație*, ‘gracias a, debido a’, o de un artículo genitival (*al*, *a*, *ai*, *ale*), entonces se halla en caso genitivo; en todos los demás contextos, se halla en caso dativo. Cabe generalizar estos asertos tomando, en lugar de *omului*, cualquier otro nombre que presente homonimia entre genitivo y dativo singular determinados. Para todo nombre que se encuentre en la misma situación —situación, por lo demás, general a todos los nombres del rumano—, la resolución de la homonimia sobre la que tratamos se descompone en las siguientes fases elementales:

O_1) leer la palabra anterior;

C_1) verificar si es una preposición del tipo indicado;

en caso afirmativo,

O_2) conceder a *omului* el valor de “genitivo” y concluir la operación;

que colocaría el resultado fuera del ámbito de las ecuaciones diofánticas. En otras palabras, x y y deben mantener una relación no arbitraria, precisa y constante para que $x + 2y = 3$. Para encontrar esta relación, basta aplicar el algoritmo de las “sustituciones”, propuesto por Euler. Así, $x + 2y = 3$ puede expresarse a través de y (el coeficiente mayor) mediante $2y = 3 - x$, de donde

$$y = \frac{3-x}{2} = 1 + \frac{1}{2} - \frac{x}{2} = 1 - \frac{x-1}{2}.$$

Sustituyendo ahora la expresión $\frac{x-1}{2}$ por n , es decir concediendo que $\frac{x-1}{2} = n$, se obtiene $2n = x - 1$. Expresando esta igualdad por medio de x , ocurre que $x = 2n + 1$, para todo valor de n .

Sustituyendo luego $\frac{x-1}{2}$ por n también en la última expresión de y donde $y = 1 - \frac{x-1}{2}$, se obtiene $y = 1 - n$, para todo valor de n .

Con ello, las nuevas expresiones

$$x = 2n + 1$$

$$y = 1 - n$$

indican justamente la relación buscada entre los números enteros x y y en la ecuación $x + 2y = 3$, para cualquier valor entero atribuible a n . En efecto, para $n = 3$:

$$x = 2 \cdot 3 + 1 = 6 + 1 = 7$$

$$y = 1 - 3 = -2$$

Aplicando estos valores a la expresión $x + 2y = 3$, se comprueba que, efectivamente, $7 + 2(-2) = 7 - 4 = 3$. Y así sucesivamente para todo valor entero, positivo o negativo, de n .

9. No disponemos de un inventario exhaustivo de todos los tipos de contexto en que pueda aparecer un nombre en genitivo o en dativo singular determinado [es decir, con artículo determinado]. El análisis algorítmico

en caso negativo,

O_1) leer la palabra anterior;

C_2) verificar si es un artículo genitival;

en caso afirmativo,

O_2) conceder a *omului* el valor de “genitivo” y concluir la operación;

en caso negativo,

O_1) leer la palabra anterior;

C_3) verificar si es un nombre con artículo determinado;

en caso afirmativo,

O_2) conceder a *omului* el valor de “genitivo” y concluir la operación;

en caso negativo,

O_3) conceder a *omului* el valor de “dativo” y concluir la operación¹⁰.

En este análisis han intervenido dos tipos de fases elementales: los operadores (O_1 , O_2 y O_3) y las condiciones lógicas u opciones elementales, es decir las opciones entre “sí” y “no” (C_1 , C_2 y C_3). El esquema lógico de este análisis sobre la homonimia entre genitivo y dativo es como sigue:

$$O_1 C_1 \uparrow O_2 \downarrow O_1 C_2 \uparrow O_2 \downarrow O_1 C_3 \uparrow O_2 \downarrow O_3.$$

La resolución de la homonimia empieza por poner en marcha el operador O_1 para luego plantear la condición C_1 ; si C_1 se verifica, entonces se aplica el operador O_2 y la homonimia queda resuelta; si C_1 no se verifica, entonces el vector de índice 1 hacia arriba que está situado a la derecha de C_1 indica que no debe continuarse con el siguiente operador, O_2 , sino con el que está situado

mico de la homonimia entre genitivo y dativo debería empezar precisamente por componer este inventario, al menos para textos de un determinado dominio científico. Sólo de este modo cabría determinar los límites de las posibilidades resolutivas —a base de consideraciones exclusivamente posicionales— de esta homonimia. Este tipo de análisis algorítmico se ha llevado a cabo con éxito en muchas lenguas (especialmente, en ruso y en inglés) y no hay motivo para pensar que el rumano vaya a ser una excepción. La presente discusión, pues, se propone indicar tan sólo el espíritu con que debe emprenderse esta tarea para un tratamiento mecánico de la lengua. Operamos, entonces, bajo el supuesto de unas cantidades exiguas de textos que impiden la formulación de unas reglas de cobertura total.

10. En este análisis se ha excluido el caso en que *omului*, aun precedido de nombre con artículo determinado esté pensado en dativo: *Repar haina omului*, [‘Remiendo la chaqueta al hombre’, pero también ‘Remiendo la chaqueta del hombre’ (dativo o genitivo, pues)]. Esta situación ambigua se deshace alterando el orden de las palabras: *Repar omului haina*, ‘Remiendo al hombre la chaqueta’.

inmediatamente después del vector de índice 1 hacia abajo, es decir con O_1 . Se plantea luego la condición C_2 ; si se verifica, entonces se pone en marcha el operador O_2 y se resuelve la homonimia; si no se verifica C_2 , entonces, siguiendo la indicación de los vectores de índice 2, se pone en marcha el operador O_1 ; acto seguido se comprueba la condición C_3 ; si se verifica, la homonimia se resuelve como hasta ahora, mediante el operador O_2 , es decir concediendo al nombre el valor de caso genitivo; si C_3 no se verifica, entonces, siguiendo las indicaciones de los vectores de índice 3, se pone en marcha el operador O_3 y de este modo se resuelve la homonimia atribuyendo al nombre el valor de caso dativo¹¹.

Observemos que en el proceso de este análisis ha intervenido un cierto grado de arbitrariedad. En efecto, hemos considerado en primer lugar todos los tipos de contexto en que el nombre *omului* está en genitivo, pero también se hubiera podido proceder a la inversa, examinando primero los tipos de contexto en que *omului* está en dativo. Veamos qué tipo de operadores y de condiciones lógicas aparecen si se utiliza este segundo procedimiento. Si a *omului* le precede un nombre sin artículo o un verbo, entonces tiene sin duda valor de “dativo”; pero si a *omului* le sigue un verbo cabe tanto la posibilidad de que sea genitivo como la de que sea dativo. Para dilucidar este dilema se actúa de la siguiente manera: en tanto que *omului* vaya seguido de un verbo se le atribuye valor de dativo si, y sólo si, no va precedido de un nombre con artículo determinado o de una preposición del tipo indicado antes o bien de un artículo genitival.

Comprobamos, pues, que aparecen los siguientes operadores: O_1 = leer la palabra anterior; O_2 = leer la palabra siguiente; O_3 = conceder a *omului* valor de “dativo” y concluir la operación; O_4 = conceder a *omului* valor de “genitivo” y concluir la operación. Asimismo, aparecen las siguientes condiciones lógicas: C_1 = es nombre sin artículo; C_2 = es verbo; C_3 = es nombre con artículo determinado; C_4 = es preposición del tipo indicado antes; C_5 = es artículo genitival. Por consiguiente, y aunque no hemos examinado todas las posibilidades que pueden surgir, es evidente que aparecen más operadores y más condiciones lógicas que en el caso anterior, cuando hemos procedido a partir de las situaciones que conducen al caso genitivo. El esquema lógico para resolver la homonimia según este último procedimiento sería

$$O_1 C_1 \uparrow O_3 \downarrow O_1 C_2 \uparrow O_3 \downarrow O_2 C_2 \uparrow O_1 C_3 \uparrow O_4 \downarrow O_1 C_4 \uparrow O_4 \downarrow O_1 C_5 \uparrow O_4 \downarrow O_3 \downarrow \dots$$

11. Naturalmente, la determinación del caso en un nombre es, en general, una cuestión sencilla para todo el que tiene una formación escolar básica [o media, en los países donde no se habla una lengua con declinación nominal]. Pero, en la traducción automática, hay que determinar el caso no a base de consideraciones relativas al significado (que la máquina ignoraría), sino a partir exclusivamente de las situaciones formales o posicionales. Cf., para ampliar este punto, Moisił, “Traducere automată”, y 6.1 y ss. Toda descripción algorítmica es incompatible con el significado. Esto hay que tenerlo en cuenta también para los demás ejemplos de más arriba.

No es difícil comprobar que el primer esquema empleado para la resolución de la homonimia de *omului* resulta mucho más simple y económico que este obtenido ahora. De ahí la importancia que adquiere la elección juiciosa de operadores y condiciones lógicas con que se trabaja.

[Confiamos en que el lector de habla española habrá podido seguir con facilidad la composición de este algoritmo (intraductible) dando simplemente por sentados los datos contextuales que intervienen en el análisis. No faltan, en nuestra lengua, situaciones parecidas de homonimia morfológica; v. gr. *cantaba* (él / yo), *cantamos* (ahora / entonces) o la más afín a la estudiada, *le* (*le veo / le digo*). En este último caso, y al margen de la doble referencia a *él-usted*, se trata de una homonimia entre acusativo y dativo: *Le presento a Carlos* es, por tal motivo, ambiguo (*Carlos es presentado por mí a él/ella/usted*, frente a *Él/usted es presentado por mí a Carlos*). Para elucidar esta homonimia, habría que recurrir, sin duda, a un algoritmo que tomara en consideración marcas de transitividad simple (acusativo) o doble (acusativo + dativo) en los verbos —junto a excepciones de tipo *Me gusta—*, de animación en los nombres y aun a procedimientos transformativos (capacidad de conversión en *se*, etc.)].

4.6. La resolución de la homonimia léxico-gramatical

Consideremos ahora un ejemplo de la lengua francesa¹². La palabra *pas* puede representar una partícula negativa o un nombre. Para describir la resolución de esta homonimia, introduzcamos los siguientes operadores: O_1 = leer la palabra anterior a *pas*; O_2 = leer la palabra anterior a la palabra anterior a *pas*; O_3 = leer la palabra situada después de *pas*; O_4 = conceder a *pas* valor de “partícula negativa” y concluir la operación; O_5 = conceder a *pas* valor de “nombre” y concluir la operación. Habrá que contar, además, con las siguientes condiciones lógicas: C_1 = es un verbo; C_2 = es un adverbio; C_3 = es la palabra *ne*; C_4 = es un verbo en infinitivo. El esquema lógico para resolver esta homonimia sería así:

$$O_1 C_1 \uparrow O_4 \downarrow O_1 C_2 \uparrow O_2 C_1 \uparrow O_4 \downarrow_{2,3} O_1 C_3 \uparrow O_3 C_4 \uparrow O_4 \downarrow_{4,5} O_5$$

Se empieza por leer la palabra anterior a *pas* (O_1). Se verifica si esta palabra es un verbo (C_1); en caso afirmativo, se concede a *pas* el valor de “partícula negativa” y se considera resuelta la homonimia (O_4). En caso negativo, se procede según las indicaciones de los vectores de índice 1 y se lee de nuevo la pa-

12. El siguiente análisis, extraído de DELAVENAY, *La machina à traduire*, no tiene un carácter exhaustivo [(Cf., también, SERRANO. “Resolución”)]. Desde luego, tampoco en el libro de Delavenay se ha tenido en cuenta el francés en su conjunto, sino una determinada categoría de textos.

labra anterior a *pas* (O_1). Se verifica si esta palabra es un adverbio (C_2); en caso afirmativo, se lee la palabra anterior a este adverbio (O_2), se verifica si es un verbo y, en caso afirmativo, se concede a *pas* el valor de “partícula negativa” (O_4). Si antes de *pas* no se encuentra un adverbio, o si se encuentra uno pero no precedido de un verbo, entonces, siguiendo los vectores de índice 2 y 3, se lee de nuevo la palabra anterior a *pas* (O_1) y se verifica si se trata de la palabra *ne* (C_3); en caso afirmativo, se lee la palabra que sigue a *pas* (O_3) y se verifica si se trata de un verbo en infinitivo (C_4). En caso afirmativo, se concede a *pas* el valor de “partícula negativa” y se considera resuelta la homonimia. Si antes de *pas* no se halla la palabra *ne* o si se halla *ne* pero después de *pas* no sigue un verbo en infinitivo, entonces, de acuerdo con los vectores de índice 4 y 5, se concede a *pas* ‘paso’ el valor de “sustantivo” y la homonimia queda completamente resuelta (O_5).

Frente al algoritmo para la resolución de la homonimia morfológica de la palabra *omului*, que se aplica a todos los nombres susceptibles de una forma ambigua entre genitivo y dativo, el algoritmo para resolver la homonimia de la palabra *pas* no contiene este rasgo de generalidad; sólo se aplica a la palabra *pas*. Se trata, entonces, de un algoritmo individual.

Existen, sin embargo, homonimias léxico-gramaticales cuya resolución conduce a un algoritmo general. Así, p. ej., en inglés [o en español] la homonimia entre un adjetivo demostrativo y un pronombre demostrativo admite un tratamiento por medio de un algoritmo general. En efecto, si introducimos los operadores: O_1 = leer la palabra siguiente; O_2 = atribuir el valor de “adjetivo demostrativo” y concluir la operación; O_3 = atribuir el valor de “pronombre demostrativo” y concluir la operación, junto con la condición lógica C_1 = es un nombre, la resolución viene dada por un esquema lógico a base de:

$$O_1 C_1 \uparrow O_2 \downarrow O_3.$$

Veamos un caso concreto de aplicación por medio de dos frases en inglés¹³:

- 1) *Under this assumption we may consider all integrals.*
- 2) *This is purely a matter of convenience.*

En la primera frase, leyendo la palabra *assumption* que sigue a *this* (O_1) y verificando si esta palabra es un nombre (C_1), comprobamos que la respuesta es afirmativa. Atribuimos, pues, a *this* el valor de “adjetivo demostrativo” (O_2)¹⁴.

En el segundo ejemplo, leyendo la palabra *is*, que sigue a *this* (O_1), y verificando si esta palabra es un nombre, se comprueba que la respuesta es negativa, por lo que atribuimos a *this* el valor de “pronombre demostrativo” (O_3).

13. Cf. MOLOSHNAIA, “Voprosy razlichenia omonimov”.

14. En el mencionado trabajo se han examinado sólo textos matemáticos, donde no suele aparecer nada entre *this* y el nombre subsiguiente.

Observemos, para terminar este apartado, que la resolución de la hominimia léxico-gramatical conduce, a veces, a algoritmos extremadamente complejos. Un ejemplo típico a este respecto lo constituye la homonimia de la palabra inglesa *that*, con posible valor de conjunción (v. gr., *We see that the solution is unique*), adjetivo demostrativo (v. gr., *We see that solution*) o pronombre demostrativo (v. gr., *We see that. That is the question*).

4.7. El hallazgo de una correspondencia castellana a una palabra inglesa

Consideremos de momento un problema simple, consistente en hallar una correspondencia en castellano a la palabra *example*¹⁵. Para ello basta determinar si aparece antes la palabra *for*. En caso afirmativo, el equivalente castellano es la locución *por ejemplo*; en caso negativo, el equivalente es el nombre *ejemplo*¹⁶. Han intervenido, por tanto, los siguientes operadores: O_1 = leer la palabra anterior; O_2 = elegir la locución castellana *por ejemplo* y concluir la operación; O_3 = elegir el nombre castellano *ejemplo* y concluir la operación. Sólo interviene una condición lógica: C_1 = ser la palabra *for*. Así podemos construir el esquema lógico que resuelve el problema:

$$O_1 C_1 \uparrow O_2 \downarrow O_3.$$

Consideremos ahora un caso algo más complicado¹⁷. Nos proponemos averiguar la correspondencia castellana a la palabra inglesa *many*. Para construir el algoritmo adecuado utilizaremos los siguientes operadores: O_1 = leer la palabra anterior a *many*; O_2 = leer la palabra posterior a *many*; O_3 = elegir el numeral castellano *cuántos* y concluir la operación; O_4 = elegir la locución castellana *tan(tos) como* y concluir la operación; O_5 = elegir la palabra castellana *muchos* y concluir la operación. Intervendrán, además, las siguientes condiciones lógicas: C_1 = que coincida la palabra con *how*; C_2 = que coincida la palabra con *as*; C_3 = que la palabra coincida con *very*; C_4 = que la palabra sea una preposición; C_5 = que la palabra sea un nombre. La resolución del problema se basa, entonces, en el esquema lógico

$$O_1 C_1 \uparrow O_3 \downarrow O_1 C_2 \uparrow O_4 \downarrow O_1 C_3 \uparrow O_5 \downarrow O_1 C_4 \uparrow O_2 O_5 \uparrow O_5 \downarrow \dots$$

Se empieza por leer la palabra anterior a *many* (O_1); se verifica si esta palabra es *how* (C_1) y, en caso afirmativo, se traduce por el numeral castellano

15. Cf. PÁNOV. "Avtomaticheski perevod".

16. [En el original, la traducción se hace al rumano, donde *example* equivale a *exemplu* y *for example*, a *de pildă*, peculiaridad que confiere, si cabe, una mayor elegancia a la ilustración. Sin embargo, en lugar de esta última locución, es muy frecuente también *de exemplu* y, mucho menos, *spre exemplu*].

17. Cf., también, PÁNOV. "Avtomaticheski perevod", donde sólo se han verificado textos matemáticos y no el inglés en conjunto. [En el original, se traduce asimismo al rumano].

cuántos (O_3). Si antes no se halla *how*, se sigue la indicación de los vectores de índice 1 y se pone de nuevo en marcha el operador O_1 , por el que se lee la palabra anterior a *many*. Se verifica si es la palabra *as* (C_2); en caso afirmativo, se traduce por la locución *tan(tos) como* (O_4), pero, en caso negativo, se sigue la indicación de los vectores de índice 2 y se lee de nuevo la palabra anterior (O_1) para verificar si se trata de la palabra *very* (C_3). En caso afirmativo, se traduce por *muchos* (O_5); en caso negativo, siguiendo la indicación de los vectores con índice 3, se lee de nuevo la palabra anterior (O_1), se verifica si es una preposición (C_4) y si la respuesta es afirmativa se lee la palabra posterior a *many* (O_2) y se verifica si se trata de un nombre (C_5); si ambas condiciones se cumplen, entonces traducimos de nuevo por medio de *muchos* (O_5). Pero si por lo menos una de las condiciones C_4 y C_5 no se cumple, entonces, siguiendo la indicación de los vectores con índice 4 y 5, ha de continuarse el algoritmo hasta un punto, que nosotros omitimos aquí.

Una situación especial ocurre cuando aparecen en un texto palabras cuya forma concreta no figura en el diccionario. Así, no encontraremos, p. ej., la palabra *books* en el diccionario, puesto que se trata de una forma de plural cuya característica no se recoge allí. En un caso de esta naturaleza se compara la palabra con cada una de las seis desinencias posibles del inglés: *-s* (plural o tercera persona del singular del presente verbal), *-ing* (gerundio), *-ed* (participio pasado), *-er* (comparativo), *-est* (superlativo) y *-th* (partícula ordinal en los numerales) (cf. 7.2). Si se comprueba que la terminación de la palabra coincide con una de estas desinencias, se separa ésta y se busca en el diccionario la palabra resultante sin desinencia. El lector mismo podrá sin duda componer fácilmente el esquema lógico del proceso para separar las desinencias inglesas y para buscar las formas resultantes en el diccionario. Se puede partir, para ello, del esquema de la figura 10 en Panov, "Avtomaticheski". También cabe emplear consideraciones análogas para resolver algorítmicamente algunas homonimias léxico-gramaticales a base de criterios total o parcialmente morfológicos¹⁸.

4.8. Algunos problemas no lingüísticos

Interesa examinar también algunos problemas no lingüísticos —cuya resolución sigue un camino semejante al que hemos descrito antes— con el fin de contemplar la generalidad de los esquemas lógicos que hemos venido utilizando. En este sentido, muchos problemas lingüísticos se identifican con problemas de otra naturaleza, de forma que todos se acogen a un determinado modo de planteamiento como es la descripción algorítmica de los procesos de ordenación, rama fundamental de la cibernética.

18. Para ello, pueden tomarse algunos ejemplos de MOLOSHNAIA, "Voprosy razlichenia omonimov".

Un problema de esta clase lo hallamos en la labor del jefe de zona en el transporte ferroviario. El jefe de zona recibe informaciones sobre el estado de las unidades ferroviarias en una porción de línea dada. Elabora la información recibida y dispone las decisiones pertinentes para satisfacer el gráfico de movimiento de los trenes. Con este propósito, da instrucciones a los mecánicos o a los jefes de estación para todo lo que se refiere al orden de movimiento de las unidades ferroviarias. En la descripción de un proceso como éste aparecen operadores del siguiente tipo: O_k = orden a la estación K para que salga el tren por la línea principal; O'_k = orden a la estación K de retener el tren en la línea principal; O''_k = orden a la estación K de pasar el tren a la línea de reserva; O'''_k = orden a la estación K para que salga el primer tren que se halle en la línea de reserva. Intervienen, además, condiciones lógicas como C_k = que se halle un tren en la estación k ; C'_k = que llegue un tren, de acuerdo con el horario, a la estación k ; C''_k = que entre la estación $k - 1$ y la estación k no se encuentre ningún tren, etcétera. Con ayuda de estos operadores y de algunas condiciones lógicas se puede componer el esquema de funcionamiento con que opera el despachador de vía férrea¹⁹.

Otro problema que comporta una descripción algorítmica es el funcionamiento de los reflejos condicionados (cf. 5.4). En un reflejo condicionado se producen dos estímulos, α y β , de los cuales α determina la reacción A , mientras que β no determina esta reacción. Pero si el estímulo α va acompañado muchas veces del estímulo β , este último llega a determinar la reacción A incluso en ausencia del estímulo α . En la conocida experiencia de Pavlov, el estímulo α era la aparición del alimento, β era el sonido de una campana y A , la secreción de saliva en el perro. He aquí cómo se establece y funciona el reflejo condicionado (el signo \rightarrow indica implicación lógica):

$$\alpha \rightarrow A, \beta \not\rightarrow A, \alpha\beta \rightarrow A, \alpha\beta \rightarrow A, \dots, \alpha\beta \rightarrow A, \beta \rightarrow A,$$

$$\beta \rightarrow A, \dots, \beta \rightarrow A, \beta \not\rightarrow A, \alpha\beta \rightarrow A, \dots, \beta \rightarrow A, \dots$$

Por $\beta \not\rightarrow A$ se entiende que β no promueve la reacción A . El grupo $\alpha\beta$ indica el conjunto de los estímulos α y β , el cual produce, evidentemente, la reacción A ($\alpha\beta \rightarrow A$). Después de que aparezca el estímulo α acompañado del estímulo β en diversas ocasiones, surge el reflejo condicionado: β produce la reacción A aunque ya no aparezca acompañado de α ($\beta \rightarrow A$). Pero si β aparece demasiadas veces sin el acompañamiento de α , entonces se extingue el reflejo ($\beta \not\rightarrow A$) y se hace necesaria una nueva acción simultánea por parte de los estímulos α y β , aunque en esta ocasión basta un número más reducido de experiencias $\alpha\beta \rightarrow A$ para obtener la reaparición del reflejo.

19. Para más detalles, cf. LIAPUNOV, "O nekotoryj obshchij voprosaj", y LIAPUNOV-SHESTOPAL, "Ob algoritmicheskom opisani".

En la descripción algorítmica de este proceso interviene una dificultad por el hecho de que no queda perfectamente determinado el número de veces en que el estímulo β debe acompañar al estímulo α para que se establezca el reflejo condicionado. Ello reclama el uso de ciertas nociones de cálculo de probabilidades, que aquí omitimos y sin las cuales no se pueden definir los operadores resultantes. Nos conformaremos, sin embargo, con una descripción menos general mediante el empleo de los siguientes operadores: O_1 = poner en marcha la reacción A ; O_2 = no poner en marcha la reacción A . Como condiciones lógicas: C_1 = que α actúe y β no actúe; C_2 = que β actúe y α no actúe; C_3 = que α y β actúen. La aparición y el funcionamiento del reflejo condicionado, tal como han sido descritos antes, se expresan con ayuda de un algoritmo que adquiere el siguiente esquema lógico (cf. 5.5 y fig. 19):

$$C_1 \uparrow O_1 C_2 \uparrow O_2 C_3 \uparrow O_1 C_3 \uparrow O_1 \dots C_3 \uparrow O_1 C_3 \uparrow O_1 C_3 \uparrow O_1 C_2 \uparrow O_1 \dots \\ \dots C_2 \uparrow O_1 C_2 \uparrow O_1 C_2 \uparrow O_2 \dots C_3 \uparrow O_1 \dots C_3 \uparrow O_1 C_2 \uparrow O_1 \dots$$

Nótese que faltan vectores hacia abajo; ello significa que si una de las condiciones no se cumple, entonces se para el funcionamiento del algoritmo. Frente a los otros algoritmos examinados hasta ahora, el presente ofrece una particularidad por el carácter cíclico con que se ponen en marcha los operadores y con que se verifican las condiciones lógicas²⁰.

Para comprender mejor el carácter cíclico de ciertos algoritmos, aludiremos a un nuevo problema. Queremos construir una descripción algorítmica para la operación por la cual tratamos de establecer conexión, desde un teléfono público, con un número determinado. Intervienen los siguientes operadores: O_1 = = descolgar el receptor; O_2 = introducir la moneda; O_3 = marcar el número; O_4 = respuesta del interlocutor y conclusión del algoritmo; O_5 = colgar el receptor; O_6 = sacar la moneda. Intervienen, además, las siguientes condiciones lógicas: C_1 = que llegue el tono; C_2 = que responda el número marcado. El esquema lógico de la operación indicada es como sigue:

$$O_1 O_2 C_1 \uparrow O_3 C_2 \uparrow O_4 \downarrow O_5 O_6 O_1 O_2 C_1 \uparrow O_3 C_2 \uparrow O_4 \downarrow O_5 O_6 O_1 O_2 \dots$$

La ausencia de vectores de índice 1 hacia abajo significa que el funcionamiento del algoritmo se detiene en cuanto la condición C_1 no se cumpla (en otras palabras, no se puede establecer contacto telefónico si no llega el tono). Después de recuperar la moneda (O_6) repetimos la secuencia de actos elementales en el mismo orden que la primera vez hasta el momento en que se cumpla la

20. Para una explicación más precisa y detallada del algoritmo utilizado en el funcionamiento del reflejo condicionado. cf. la bibliografía de la nota anterior.

condición C_2 (cuando conteste el número marcado). Así, pues, el algoritmo para entablar contacto telefónico ofrece un carácter cíclico, en el cual el número de ciclos es, en teoría, indefinidamente grande.

4.9. El concepto de “algoritmo normal”

Este concepto fue introducido, hace unos veinte años, por el matemático soviético A. A. Markov, hijo del gran matemático del mismo nombre. Para no confundirlos, a menudo se distingue entre A. A. Markov-padre y A. A. Markov-hijo. El primero fue un gran especialista en la teoría de las probabilidades y se hizo célebre por los procesos que adujo y estudió, y que hoy llevan justamente el nombre de “procesos de Markov”. En cuanto a A. A. Markov-hijo, hay que destacar su obra capital, *Teoría de los algoritmos*, aparecida en 1954, de la cual trataremos de comentar algunos conceptos en lo que sigue²¹.

Consideremos un conjunto A de elementos llamados *l e t r a s*. Así, A constituye un *a l f a b e t o*. Cualquier sucesión finita de letras, de las cuales algunas pueden repetirse, se denominará *p a l a b r a s o b r e e l a l f a b e t o* A o, si no existe peligro de confusión, simplemente *p a l a b r a*. El conjunto de las palabras sobre A se notará mediante $\mathcal{P}(A)$. Un elemento de $\mathcal{P}(A)$ es, pues, de forma:

$$a_1 a_2 \dots a_i \dots a_n,$$

donde $a_i \in A$ ($i = 1, \dots, n$). La *extensión de una palabra* es, por definición, el número de letras (distintas o no) que componen esta palabra.

A partir de dos palabras dadas, $\alpha = a_1 \dots a_i \dots a_n$ y $\beta = b_1 \dots b_j \dots b_m$, definiremos la palabra $\alpha\beta$ como igual a

$$a_1 a_2 \dots a_i \dots a_n b_1 \dots b_j \dots b_m$$

De un modo análogo, a partir de tres palabras dadas, α, β, γ , podemos definir la palabra $\alpha\beta\gamma$ como formada por la agregación de las letras de α, β y γ en el mismo orden en que figuran en estas palabras.

Se considera también la *p a l a b r a v a c í a*, indicada por θ y que se define por la relación $\alpha = \alpha\theta = \theta\alpha$, cualquiera que sea la palabra α .

Sean dos palabras α y β . Diremos que β está incluido en α , si existen dos palabras γ y δ tales que

$$\alpha = \gamma\beta\delta \tag{1}$$

No se excluye la posibilidad de que una de las palabras γ, δ o ambas sea la

21. Cf., en la Bibliografía, MARKOV, *Teoría algoritmov*.

palabra vacía. Resulta, pues, que cada palabra está incluida en sí misma y que la palabra vacía está incluida en toda palabra.

Si β está incluida en α , elegimos, entre las distintas representaciones de α de forma (1), aquella en que γ tenga la extensión menor. (Se puede demostrar que existe una representación de este tipo y sólo una.) La representación de α de forma (1), con γ en su extensión mínima, se denomina *primera inclusión de β en α* .

Sean ahora $\alpha_1, \dots, \alpha_n$ y β_1, \dots, β_n palabras de $\mathcal{P}(B)$, donde B es un alfabeto que contiene al alfabeto A .

Consideremos el esquema

$$(E) \alpha_1 \rightarrow \beta_1, \alpha_2 \rightarrow \beta_2, \dots, \alpha_i \rightarrow \beta_i, \dots, \alpha_j \rightarrow \beta_j, \dots, \alpha_n \rightarrow \beta_n.$$

Con ayuda del esquema E , definiremos una función \mathcal{A} que asocie a cada una de diversas palabras de $\mathcal{P}(B)$ una palabra bien determinada de $\mathcal{P}(B)$.

Sea π una palabra de $\mathcal{P}(B)$. Si ninguna de las palabras $\alpha_1, \alpha_2, \dots, \alpha_n$ está incluida en π , entonces ponemos por definición $\mathcal{A}(\pi) = \pi$. Si existe un α_j incluido en π , entonces, escogiendo a j de modo que α_k no esté incluido en π para ningún valor $k < j$, consideramos la primera inclusión de α_j en π : $\pi = \gamma\alpha_j\delta$. Tal como indica el esquema E , sustituimos α_j por β_j y obtenemos la palabra $\pi_1 = \gamma\beta_j\delta$. Si el vector de rango j fuera seguido de un punto (como ocurre con el vector de rango i), entonces hubiéramos puesto $\mathcal{A}(\pi) = \pi_1$. En caso contrario, la palabra π_1 se somete al mismo tratamiento que la palabra π . Si ninguna palabra α_i ($1 \leq i \leq n$) está incluida en π_1 , entonces ponemos $\mathcal{A}(\pi) = \pi_1$. En caso contrario, α_k debe incluirse en π_1 de tal modo que α_m no esté incluida en π_1 para ningún $m < k$. Sea $\pi_1 = \varepsilon\alpha_k\omega$ la primera inclusión de α_k en π_1 . Siguiendo la indicación del vector de rango k del esquema E , sustituimos α_k por β_k y obtenemos la palabra $\pi_2 = \varepsilon\beta_k\omega$. Si al vector de rango k sigue un punto, ponemos $\mathcal{A}(\pi) = \pi_2$. En caso contrario, aplicamos a π_2 el mismo tratamiento que hemos aplicado anteriormente a π y a π_1 , y así sucesivamente.

Si, procediendo tal como indicamos, no llegamos nunca a un vector seguido de un punto o a una palabra π_i en la cual no esté incluida ninguna de las palabras $\alpha_1, \dots, \alpha_n$, entonces la función \mathcal{A} no se define para la palabra π o, en otros términos, no tiene sentido hablar de $\mathcal{A}(\pi)$. Pero si llegamos a un vector seguido de un punto o a una palabra π en la que no esté incluida ninguna de las palabras $\alpha_1, \dots, \alpha_n$, entonces la palabra $\mathcal{A}(\pi)$ existe y es única.

La función \mathcal{A} , tal como se ha descrito, es, por definición, un *algoritmo normal* sobre el alfabeto B . El esquema E es el *esquema del algoritmo \mathcal{A}* . Un algoritmo normal está determinado por el esquema E .

Para ilustrar la noción de algoritmo normal utilizaremos el ejemplo de la operación de suma aplicada a los números naturales. A este propósito, se comprenderá también en qué sentido se considera que un alfabeto B es más rico que el alfabeto inicial A .

Sea un alfabeto A formado por una sola letra, que representaremos por la cifra "1". Los números naturales serán descritos entonces sobre la base 1, de modo que para el número 7 se escribirá

$$1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1.$$

Consideremos P números naturales N_1, N_2, \dots, N_p . Cada uno de los números N_1, \dots, N_p admite una representación como palabras sobre el alfabeto A , lo que conduce a su representación sobre la base 1. Sin embargo, para representar el sistema de todos estos P números naturales, hay que utilizar todavía otro símbolo auxiliar, que será notado por medio de un asterisco: *. De esta forma, obtenemos un alfabeto B que contiene al alfabeto A : $B = \{1, *\}$.

El sistema de estos P números naturales se representa, en el alfabeto considerado, por medio de la palabra

$$\alpha = \underbrace{1 \ 1 \ \dots \ 1}_{N_1 \text{ veces}} * \underbrace{1 \ 1 \ \dots \ 1}_{N_2 \text{ veces}} * \dots * \underbrace{1 \ 1 \ \dots \ 1}_{N_p \text{ veces}} \quad (2)$$

Hay que encontrar, entonces, un algoritmo \mathcal{A} que transforme la palabra α en la palabra

$$\mathcal{A}(\alpha) = \underbrace{1 \ 1 \ \dots \dots \dots \ 1}_{N_1 + N_2 + \dots + N_p \text{ veces}},$$

palabra que constituye la representación sobre la base 1 de la suma $N_1 + \dots + N_p$.

Consideremos el esquema E siguiente:

$$* \longrightarrow \theta.$$

(Recordemos que θ es la palabra vacía). Demostraremos que la función \mathcal{A} definida por este esquema es justamente el algoritmo normal buscado.

Sea una palabra β de $\mathcal{P}(A)$, es decir que no contenga la palabra *. Resulta que $\mathcal{A}(\beta) = \beta$. Esto es natural, puesto que se refiere al caso en que los p números que deben sumarse se reducen a uno solo, y éste será, evidentemente, el resultado de la suma.

Sea ahora una palabra α que contenga tanto la letra 1 como la letra *. Así, pues, α es de forma (2). La primera inclusión de * en α es

$$\alpha = \gamma * \delta,$$

donde

$$\gamma = \underbrace{1 \ \dots \ 1}_{N_1 \text{ veces}} \quad \delta = \underbrace{1 \ \dots \ 1}_{N_2 \text{ veces}} * \dots * \underbrace{1 \ \dots \ 1}_{N_p \text{ veces}}.$$

Aplicando el esquema E , obtenemos la palabra

$$\alpha_1 = \gamma \theta \delta = \underbrace{\gamma}_{N_1 + N_2 \text{ veces}} * \underbrace{\delta}_{N_3 \text{ veces}} * \dots * \underbrace{\theta}_{N_p \text{ veces}} .$$

La primera inclusión de $*$ en α_1 es

$$\alpha_1 = \gamma_1 * \delta_1,$$

donde

$$\gamma_1 = \underbrace{\gamma}_{N_1 + N_2 \text{ veces}}, \quad \delta_1 = \underbrace{\delta}_{N_3 \text{ veces}} * \dots * \underbrace{\theta}_{N_p \text{ veces}} .$$

Aplicando el esquema E , obtenemos la palabra

$$\alpha_2 = \gamma_1 \theta \delta_1 = \gamma_1 \delta_1 = \underbrace{\gamma_1}_{N_1 + N_2 + N_3 \text{ veces}} * \underbrace{\delta_1}_{N_4 \text{ veces}} * \dots * \underbrace{\theta}_{N_p \text{ veces}} .$$

Continuando de este modo, llegamos, al cabo de un número finito de pasos, a la palabra

$$\alpha_{p-1} = \underbrace{\gamma_1 \delta_1 \dots \theta}_{N_1 + \dots + N_p \text{ veces}}$$

y, como $*$ no está incluido en α_{p-1} , resulta que

$$\mathcal{A}(\alpha) = \alpha_{p-1}.$$

4.10. Representación de un algoritmo normal con ayuda de operadores y de condiciones lógicas

Consideremos un algoritmo normal \mathcal{A} definido por el esquema E del párrafo precedente. Para fijar las ideas, supongamos que el único vector seguido de punto ("vector final") es el vector de índice j . Introduzcamos, para cada número natural i ($1 \leq i \leq n$), el operador O_i , que consiste en sustituir, en la primera inclusión de la palabra α_i en la palabra transformada, α_i por β_i . Introduzcamos, además, para cada número natural i , $1 \leq i \leq n$, la condición lógica $C_i =$ que la palabra α_i esté incluida en la palabra transformada. E introduzcamos, por fin, la condición lógica idénticamente falsa, que indicaremos por medio de C_ω .

Consideremos ahora la siguiente sucesión de operadores y de condiciones lógicas²²:

$$\begin{array}{cccccccccccccccc} \downarrow & C_1 & \dots & \downarrow & & \downarrow & \dots & \downarrow & C_1 & \uparrow & O_1 & C_\omega & \uparrow & \downarrow & C_2 & \uparrow & O_2 & C_\omega & \uparrow & \dots \\ n+1 & & & n+j-1 & & n+j+1 & & 2n & & & & & & & & & & & & & \\ \dots & \downarrow & C_j & \uparrow & O_j & C_\omega & \uparrow & \dots & \downarrow & C_n & \uparrow & O_n & C_\omega & \uparrow & \downarrow & & \downarrow & & & & \\ j-1 & & & & & & & & & n-1 & & n & & & n & & n+j & & & & \end{array}$$

Este esquema transforma cada palabra α en la palabra $\mathcal{A}(\alpha)$, donde \mathcal{A} es el algoritmo normal definido por el esquema E del párrafo anterior. Si C_1 se verifica o, dicho de otro modo, si α_1 está incluida en la palabra α , entonces se aplica el operador O_1 , situado inmediatamente a la derecha de C_1 , y por tanto se sustituye α_1 por β_1 en la primera inclusión de α_1 en α . Viene luego la condición C_ω , la cual, siendo idénticamente falsa, no se verifica y, por tanto, siguiendo la indicación de los vectores de índice $n+1$, se empieza de nuevo por verificar la condición C_1 , etcétera. Si C_1 no se verifica, es decir si α_1 no está incluida en α , entonces, siguiendo la indicación de los vectores de índice 1, se verifica la condición C_2 . Si α_2 está incluida en α , entonces se cumple C_2 y se aplica el operador O_2 , situado inmediatamente a la derecha; sigue luego la verificación de la condición C_ω , la cual, siendo idénticamente falsa, no queda satisfecha y, por lo tanto, siguiendo la indicación de los vectores de índice $n+2$, se recomienza por verificar la condición C_1 , etc.

C_j y O_j presentan un caso especial. Si C_j se verifica, entonces se aplica el operador O_j y luego sigue la condición idénticamente falsa C_ω , de manera que, siguiendo los vectores de índice $n+j$, llegamos a la extremidad derecha del esquema, por lo que $\mathcal{A}(\alpha)$ ha sido hallado.

4.11. Conclusiones sobre el concepto general de algoritmo. Las propiedades algorítmicas. Implicaciones lingüísticas

Tratemos ahora de destacar determinados rasgos de los procesos examinados hasta aquí. Todos los ejemplos han presentado, pese a su variedad, la siguiente coincidencia: aparece una cierta información inicial y hay que elaborar otra información, de respuesta. La información inicial ha consistido, sucesivamente, en un par de números menores que 100, una homonimia morfológica, algunas homonimias léxico-gramaticales, las palabras inglesas *example* y *many*, la situación de unas unidades ferroviarias, la aparición de dos estímulos y la operación

22. Cf. IANOV. "O logicheski skemaj".

de marcar un número telefónico. La información de respuesta, a su vez, ha consistido, respectivamente, en el resultado de una sustracción, la resolución de la alternativa entre genitivo y dativo, entre nombre y partícula negativa, entre pronombre y adjetivo demostrativos, el hallazgo de la correspondencia castellana a las palabras *example* y *many*, la determinación del desplazamiento de trenes en función de su situación anterior, la aparición, en determinadas condiciones, de ciertas reacciones y, en fin, la obtención de una conexión telefónica.

El proceso por el que se pasa de la información inicial a la información de respuesta, a pesar de ser distinto cada vez, ha permanecido inalterado desde el punto de vista de su estructura general. En efecto, en cada uno de los casos este proceso consta de una sucesión bien determinada de operadores y de condiciones lógicas. El funcionamiento en sí mismo podía variar tomando una u otra dirección a partir del cumplimiento o no cumplimiento de las condiciones lógicas. Pero una vez fijados los valores de estas condiciones lógicas, la sucesión de los operadores estaba unívocamente determinada (extremo que se indicaba mediante los vectores numerados).

Una sucesión de operadores y de condiciones lógicas de este tipo que, partiendo de una determinada información inicial, elabora una información de respuesta, constituye un algoritmo y la descripción correspondiente del fenómeno es una descripción algorítmica²³.

Una descripción algorítmica se distingue, en primer lugar, por su exactitud y por la exclusión de toda arbitrariedad. Los operadores y las condiciones lógicas que constituyen las fases elementales de un algoritmo nunca admiten más de una interpretación. La ausencia de cualquier ambigüedad en su definición repercute especialmente en la descripción algorítmica de los problemas lingüísticos. Una descripción algorítmica debe utilizar tan sólo casos formales o posicionales. Así, pues, en el algoritmo que resuelve la homonimia morfológica de la palabra *omului* (cf. 4.5), la definición de los operadores emplea situaciones de tipo: “palabra que precede a la palabra dada”, “palabra que sigue a la palabra dada”. Estas operaciones son totalmente formalizadas si se formaliza la definición del concepto de “palabra”. En la definición de las condiciones lógicas se han utilizado nociones como “verbo”, “nombre”, “artículo”, “artículo genitival”, “preposición”, etcétera. Hay, pues, que atender a la necesidad, entre otras, de definir formalmente las partes del discurso. Una situación semejante sobreviene al resolver algorítmicamente las homonimias léxico-gramaticales y al buscar la correspondencia castellana a una palabra inglesa.

Reconocemos, así, que toda descripción algorítmica debe ir precedida de

23. La variante que hemos escogido para esta exposición sobre el concepto de algoritmo se ha inspirado en LIAPUNOV, “O nekotoryj obshchij voprosaj” y en LIAPUNOV-SHESTOPAL, “Ob algoritmicheskom opisani”. La relación entre esta variante de algoritmo y la concepción algorítmica en el sentido de MARKOV, *Teoría algoritmov*, ha sido estudiada por IANOV, “O logicheski skemaj”, y resumida en el párrafo anterior.

un proceso de formalización de determinadas situaciones. Mientras la creación de los algoritmos de traducción tuvo un carácter más bien empírico, estas formalizaciones se efectuaban ad hoc. Las improvisaciones, sin embargo, no pudieron durar mucho, ya que pronto se sintió la necesidad de perfilar con mayor precisión las posibilidades de la descripción formal de la lengua, desde el momento en que de ello depende también la esfera de acción de las descripciones algorítmicas. Así empezó a reconocerse la importancia creciente de dos direcciones de investigación: la lingüística estructural y la modelación de la lengua sobre la base de la teoría de los conjuntos, de la lógica matemática, de la semántica lógica y del cálculo de probabilidades (cf. 3.1 y ss)²⁴.

La información inicial que elabora un algoritmo puede variar dentro de ciertos límites. Así, en el algoritmo de la sustracción cabe partir de cualquier pareja de números menores que 100; en el algoritmo para resolver la homonimia morfológica, la palabra *omului* puede ser sustituida por *vinului*, *calului*, *copilului*, 'del/al viento, caballo, niño', etcétera, y así sucesivamente. En el algoritmo para el funcionamiento ferroviario, la situación inicial de los trenes admite variaciones; en el algoritmo que describe el reflejo condicionado, el sonido de la campanilla puede permutarse por cualquier otro estímulo que no determine directamente la secreción salival, y, en fin, en el algoritmo que describe la conexión telefónica, la información inicial es representable por cualquier número de la guía.

Todas estas situaciones proclaman la amplitud del algoritmo, la capacidad con que se aplica a todos los objetos de una determinada clase.

Al mismo tiempo, conviene reconocer que algunos de los algoritmos descriptos hasta aquí tienen un carácter singular, individual. Así ocurre, p. ej., con el algoritmo que resuelve la homonimia de la palabra francesa *pas* y los algoritmos para el hallazgo de las correspondencias castellanas a las palabras inglesas *example* y *many*. Pero también estos algoritmos forman parte de otros algoritmos más amplios, que se aplican a la traducción de una lengua a otra. En este caso, la clase de objetos a que se aplica el algoritmo consta de distintos textos.

Un problema fundamental, con importantes repercusiones en la ciencia del lenguaje, consiste en la realización de algoritmos cada vez más vastos. Por ejemplo, la resolución de la homonimia puede estudiarse desde distintos niveles de generalidad. Cabe examinar la resolución de la homonimia de la palabra *omului*, la homonimia entre genitivo y dativo en el nombre, la homonimia entre genitivo y dativo en general, cualquiera que sea la parte del discurso a que pertenezca la palabra considerada, la homonimia casual o la homonimia morfológica en el género. A este propósito, es indispensable establecer el grado máximo de generalidad compatible a un algoritmo único para la correspondiente clase entera de objetos. A fin de alcanzar este objetivo, es menester construir una jerarquía

24. Para el concepto de modelación, cf. MARCUS, "Modelarea matematică", y 3.1 y ss.

completa de los objetos y estructuras lingüísticas. Para cada problema lingüístico, el conjunto de los objetos lingüísticos a los que se aplica el problema considerado se divide en clases de objetos, de tal modo que dentro de cada clase se aplique a cada objeto el mismo algoritmo de resolución y a objetos pertenecientes a clases distintas, también distintos algoritmos.

Un algoritmo siempre resuelve un problema determinado, nunca carece de finalidad. Pueden faltar a veces las condiciones lógicas —así sucede, p. ej., en el algoritmo para la suma de dos números o en el algoritmo para el funcionamiento del regulador de Watt en el mantenimiento de una velocidad constante²⁵—, pero los operadores no pueden faltar nunca.

En una descripción algorítmica existe la tendencia de descomponer la información en elementos irreducibles o informaciones elementales (“cuanta de información”). Este proceso de “cuantificar” la información también se refleja en la sucesión de opciones elementales (opción “sí-no”) que acompaña normalmente a un algoritmo en su carácter binario, dicotómico.

Los algoritmos aducidos en el presente capítulo no llevan hasta el final este proceso de cuantificación. Por ejemplo, el acto “verificar si es verbo” puede ser, a su vez, objeto de otro algoritmo y, por tanto, descompuesto en actos más simples.

La tendencia de analizar hasta llegar a los elementos primarios, cada vez más acentuada en la lingüística moderna, corresponde a las necesidades de una descripción algorítmica, mientras que estas descripciones orientan y estimulan, por su parte, el desarrollo de la lingüística.

Una descripción algorítmica tiende a tener un carácter exhaustivo o, en otras palabras, tiende a tener en cuenta todas las condiciones que influyen sobre el problema considerado. Por ejemplo, para resolver por completo la homonimia morfológica de la palabra *omului* es preciso tomar en consideración todos los tipos de contexto en que puede aparecer esta palabra²⁶. La tendencia a la exhaustividad se opone, en cierto modo, a la tendencia a la amplitud algorítmica.

En la introducción (cf. 4.1), hemos afirmado que la importancia del concepto de algoritmo está vinculada al desarrollo de la cibernética. Vamos a precisar ahora este aserto, puesto que con él se relaciona, en primer lugar, la importancia que asume la investigación de la lengua en el espíritu algorítmico.

La relación profunda entre “algoritmo” y “máquina de cálculo” se plantea en los siguientes términos: para transferir a la máquina un proceso de ordenación es necesario que ésta admita una descripción algorítmica. Más aún, si ima-

25. Cf. LIAPUNOV-SHESTOPAL. “Ob algoritmicheskom opisani”.

26. Precisamente por ello, la etapa inicial en la composición algorítmica de traducción contiene, entre otras operaciones, la de inventariar los tipos de contextos para las distintas situaciones gramaticales.

ginamos una máquina ideal (con capacidad infinita o con memoria ilimitada), se puede afirmar que todo proceso de ordenación que admita una descripción algorítmica es, por este mismo hecho, un proceso que puede ser transferido a la máquina. Este aserto es objeto de un teorema muy profundo de la lógica matemática. Desde luego, en este teorema se consideran los conceptos de “algoritmo” y de “máquina” en un grado superior de precisión, justamente la precisión que suele denominarse “precisión matemática”. Este punto de rigor sobrepasa el marco de la presente discusión.

Resulta, pues, que los problemas lingüísticos que admiten una resolución algorítmica definen la parte de la investigación lingüística y de la labor de los lingüistas susceptible de ser transferida a la máquina y, por consiguiente, de ser automatizada.

A pesar de que haya una relación tan estrecha entre el algoritmo y la máquina, el algoritmo no se comunica directamente con la máquina, sino por medio de una programación. El programa es la sucesión de instrucciones dadas a la máquina con el fin de realizar materialmente el algoritmo. Así, pues, frente a la noción de “algoritmo”, que pertenece a la lógica matemática, hay la noción de “programa”, que pertenece a la cibernética. La posibilidad de programar un algoritmo en una máquina depende tanto de la estructura del algoritmo como de las propiedades de la máquina.

La descripción algorítmica de un proceso lingüístico no es una reproducción exacta de este proceso, sino tan sólo una aproximación, puesto que presupone una modelación previa del proceso. Un problema fundamental reside en la calidad de esta aproximación y en los criterios por los cuales se valora esta calidad.

Otro problema deriva del carácter más o menos económico de un algoritmo. Esto se refiere, en primer lugar, a la posibilidad de programarlo de la forma más simple posible y, por tanto, con un programa que requiera un uso mínimo de la capacidad memorística de la máquina. También a este respecto aparece una verdadera problemática lingüística. Así, p. ej., no basta con dar una solución algorítmica a la homonimia morfológica, ya que conviene elegir, entre las distintas soluciones algorítmicas posibles, la óptima. Al tratar sobre la hominimia de la palabra *omului*, hemos tenido ocasión de ver que se obtiene un algoritmo más simple, más económico, cuando se empieza por verificar las condiciones en que la palabra aparece en caso genitivo. La negligencia de este factor determina, como se ha demostrado en 4,5, un algoritmo más complejo y más costoso desde el punto de vista de su programación.

Hemos de observar, al mismo tiempo, que debido al progreso rapidísimo de las máquinas de cálculo el problema de la posibilidad y de la dificultad de programación de un algoritmo va disminuyendo poco a poco. El problema principal se halla en la realización de algoritmos lo más comprensivos posible y lo más aproximados a los procesos que describan.

¿Acaso existen problemas lingüísticos que no admitan una resolución algorítmica? La respuesta es, evidentemente, afirmativa. ¿Cuáles son, entonces, los problemas lingüísticos susceptibles de un tratamiento algorítmico? Esta cuestión es muy compleja y depende del progreso general de la lingüística. Si bien es verdad que muchos de los problemas lingüísticos que aún esperan una resolución algorítmica la recibirán en el futuro, también es verdad, al mismo tiempo, que hay extensos dominios de la lingüística inaccesibles a una descripción algorítmica, incluso a los algoritmos individuales, puesto que en muchos problemas lingüísticos la referencia al significado es inevitable.

La noción de algoritmo se emplea a veces en un sentido más general que el utilizado en el presente capítulo. En algunos casos, la descripción algorítmica está concebida como un método inductivo, como un procedimiento formal por medio del cual se generan o construyen determinados objetos. En relación con esta acepción más amplia del término algoritmo, se considera que, en el estudio de las lenguas, pueden obtenerse los resultados más efectivos bien a base de describir algorítmicamente los objetos lingüísticos o bien a base de representar la lengua como un sistema axiomático-deductivo y, por tanto, como un cálculo lógico²⁷. En consecuencia, las nociones de “algoritmo” y de “cálculo” deben considerarse como conceptos básicos de la lingüística. La esfera de aplicación de los algoritmos no se limita, por lo demás, a la traducción automática. La investigación en lingüística teórica debe apoyarse, en la mayor medida posible, en la descripción algorítmica y en el cálculo lógico. La idea misma de “sistema”, ampliamente utilizada en todas las teorías lingüísticas contemporáneas, puede interpretarse como un procedimiento para la generación de determinados objetos y, por lo tanto, como un algoritmo en sentido lato²⁸.

27. Ésta es la concepción de ANDREEV-IVANOV-MELCHUK, “Nekotorye zamechania”.

28. Cf. la nota anterior.

5. Lenguas informacionales

5.1. Introducción

El desarrollo de la ciencia y la técnica contemporáneas ha promovido la aparición de unos tipos nuevos de máquinas. Hoy por hoy, se distinguen tres de estos tipos: el de las máquinas que modifican la forma de la sustancia, el de las que modifican la forma de la energía y el de las que manipulan la información. En la primera categoría entran el torno y las herramientas mecánicas, en la segunda figuran las máquinas eléctricas, las turbinas, los motores de explosión, etc., y a la última, pertenecen todos los sistemas de telecomunicación, así como los ordenadores electrónicos.

El progreso experimentado en los ordenadores electrónicos ha determinado, a su vez, la aparición de problemas lingüísticos nuevos, desde el momento en que el hombre entabla comunicación con estas máquinas, es decir se “dirige” a ellas y les “da instrucciones”.

En este capítulo estudiaremos algunos de los principales aspectos de las lenguas informacionales¹. Para una buena comprensión de la problemática, empezaremos por presentar los fundamentos de los ordenadores digitales.

5.2. Los ordenadores digitales

Los progresos de la cibernética están íntimamente ligados al desarrollo técnico de los ordenadores digitales, en tanto que medios modernos para el tratamiento de la información que se caracterizan por su gran velocidad operativa y por su funcionamiento automático.

En lo que sigue vamos a exponer brevemente, en primer lugar, una clasifi-

1. Cf. BOIANGIU-NICOLAU-NIȚĂ, *Instrumente*; GRENIEWSKI, *Elementy Cybernetiky*; NICOLAU, *Introducere*; POYEN-POYENNE, *Langage électronique*, y VAIDA, *Stroje*.

cación de los ordenadores electrónicos, después de lo cual veremos con cierto detalle el funcionamiento de los ordenadores digitales, única categoría de ordenadores electrónicos para la que puede disponerse un lenguaje especial. Presentaremos a continuación, en sus líneas principales, algunos de los lenguajes contruidos para expresar los programas de trabajo con los ordenadores digitales.

Los ordenadores son máquinas informacionales, que operan, por tanto, con la información: se les comunican determinados datos iniciales y un programa operacional, a partir de los cuales se obtienen los resultados de salida. Más exactamente, por ordenador se entiende un dispositivo capaz de realizar con su funcionamiento un determinado algoritmo. Los elementos funcionales son básicamente de naturaleza electrónica: tubos electrónicos, dispositivos semiconductores, etc.

En los ordenadores la información es llevada por ciertas magnitudes físicas, características del ordenador. En los electrónicos, la información viene representada por señales eléctricas. Puede hablarse, entonces, del lenguaje de la máquina por el que se entiende el código con que se representan las magnitudes en la máquina. En general, e incluso en una misma máquina, se emplean códigos distintos adecuados al funcionamiento de las distintas partes de la misma. Del mismo modo que el soporte físico de la información tiene una variación continua o discreta (discontinua), también los ordenadores electrónicos se dividen en ordenadores de funcionamiento continuo y ordenadores de funcionamiento discontinuo.

Si entre el valor de la magnitud característica del ordenador y el valor de la magnitud representada existe una correspondencia biunívoca y bicontinua, entonces el ordenador respectivo forma parte de la categoría de las máquinas analógicas, a veces también llamadas modelos. La regla de cálculo, por ejemplo, forma parte de esta categoría.

Si la magnitud física del ordenador ofrece una variación discontinua y si no puede tener más que un número finito de estados, entonces esta magnitud puede representarse por medio de un número de m cifras (con m finito). Las máquinas de esta índole se denominan digitales, numéricas o cíficas, y forman parte de la categoría de los autómatas con un número finito de estados (autómatas finitos).

Los ordenadores digitales funcionan a base de un algoritmo dado por el programador. Registran determinados datos iniciales, sobre los cuales efectúan luego operaciones en conformidad con el programa introducido. Todo ordenador, por su construcción, se caracteriza por el repertorio específico de operaciones que puede realizar. Algunas de estas operaciones son de naturaleza lógica y otras de naturaleza matemática. Examinemos en la figura 10 el esquema completo de un ordenador digital, en el que se distinguen más elementos.

El dispositivo de entrada tiene por finalidad la traducción de los datos iniciales y del programa operacional a señales adecuadas al lenguaje de la máquina. Como los ordenadores actuales funcionan con señales eléctricas, el dispositivo de entrada traduce las magnitudes de entrada a señales eléctricas con arreglo a un determinado código. En ciertas máquinas perfeccionadas, el dispositivo de entrada permite la lectura directa de datos dactilografiados. En la mayoría de ordenadores, sin embargo, es necesaria una etapa intermedia, en la que los datos iniciales y el programa se codifican en un lenguaje intermedio (perforaciones en una cinta o en una tarjeta de material electroaislante, o porciones transparentes y opacas en una película cinematográfica). El dispositivo de entrada traduce luego estas señales a señales eléctricas, interpretables por la máquina.

La parte aritmética efectúa los cálculos propiamente dichos.

La memoria de la máquina conserva los datos iniciales y el programa, así como algunos datos intermedios.

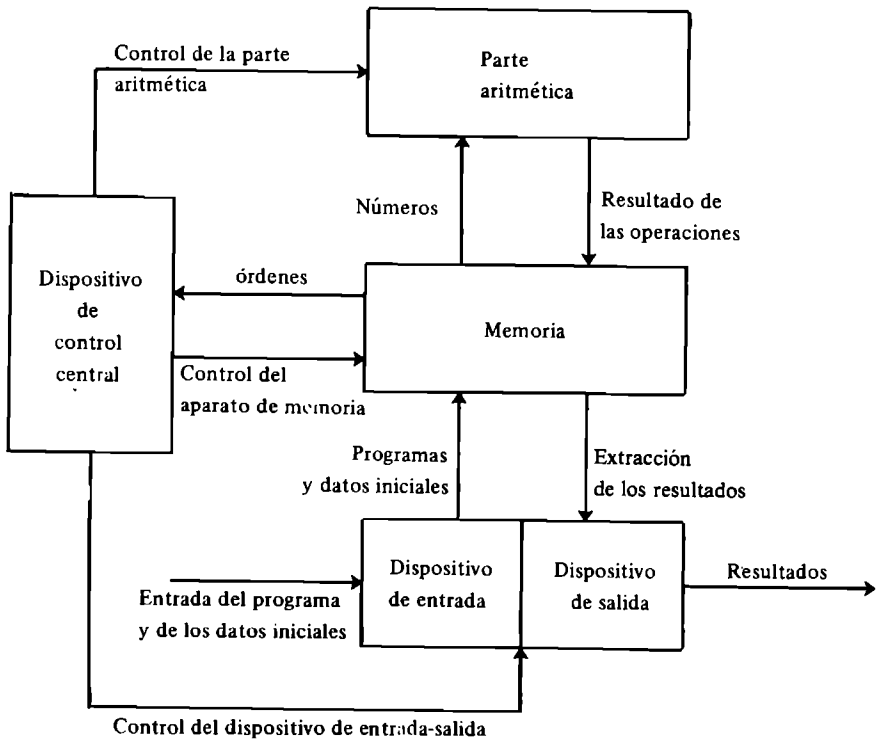


Figura 10

El control central es la parte del ordenador que dirige la sucesión de los cálculos efectuados por la máquina.

El dispositivo de salida tiene como misión traducir la información procedente del ordenador a una forma accesible al receptor de los datos (hombre o máquina). Este dispositivo de salida suele comprender una máquina de escribir eléctrica, que imprime los resultados a gran velocidad.

En los siguientes párrafos daremos algunos detalles sobre la organización propiamente dicha del ordenador. Para comprender su funcionamiento no es necesario, por lo demás, entrar en particularidades sobre el cometido concreto de cada uno de sus elementos; basta, en cambio, seguir los principios lógicos que están en la base de su concepción. Sin que ello constituya un hecho esencial, la mayoría de los ordenadores digitales existentes opera con señales binarias, que pueden traducirse en sucesiones formadas por las cifras 0 y 1.

El dispositivo de memoria se compone de una serie de células numeradas, denominadas direcciones, donde se conserva la información. Por ello se distinguen, por una parte, las direcciones —caracterizadas por su número de orden— y, por otra, el contenido de estas direcciones.

5.3. El programa del ordenador

Para resolver un determinado problema con ayuda de un ordenador digital, basta comprender que los datos iniciales se introducen en ciertas direcciones de la memoria y que el programa operativo de la máquina consiste en indicar el modo en que ésta toma la información contenida en distintas direcciones, la manipula y la comunica luego al exterior o bien la deposita en las direcciones especificadas en el programa mismo².

Consideremos, de momento, la llamada máquina “con cuatro direcciones”, cuyo programa operacional comprende una sucesión de instrucciones formadas, cada una de ellas, por los siguientes elementos: el número de orden de la instrucción (*Nr*), el código de una operación (*Op*) y cuatro direcciones:

$$Nr \quad Op \quad A_1 \quad A_2 \quad A_3 \quad A_4$$

Las operaciones se indican mediante números. Puede convenirse, p. ej., en que 01 signifique suma, 02 resta, 03 multiplicación, 04 división, etc.

De las cuatro direcciones de una instrucción, las dos primeras — A_1 y A_2 — corresponden a las direcciones en que se registran los dos números sobre los cuales se efectúa la operación indicada en la orden respectiva. La tercera, A_3 , indica la dirección en que se registrará el resultado de la operación. Y la cuarta,

2. Cf. VAIDA, *Avtomotizatsia programmirovania*.

A_4 , expresa el número de la orden siguiente a la que se pasa en el desarrollo del programa. Así, p. ej., la instrucción 25 puede ser:

25 02 125 213 626 62.

La interpretación de esta secuencia de números es como sigue: el primer grupo de cifras (25) representa el número de orden de la instrucción procedente del programa, es decir de la cadena de instrucciones también llamadas órdenes. El segundo grupo (02) indica la operación que debe efectuarse sobre los números de las direcciones A_1 y A_2 —en el presente caso, 125 y 213 (los dos siguientes grupos de cifras)—. Las dos últimas direcciones representan la dirección en que se introduce el resultado del cálculo y el número de orden que se ejecuta en la etapa sucesiva.

En resumen, la instrucción aducida como ejemplo significa que del contenido de la dirección 125 se reste el contenido de la dirección 213 y que el resultado se inscriba en la dirección 626. Se pasa luego a ejecutar la orden 62.

Para fijar mejor las ideas, vamos a componer un programa que efectúe el siguiente cálculo:

$$f(x) = \frac{ax + b}{cx - d}$$

Los números comprendidos en esta fracción se registran en la memoria tal como se indica en la siguiente tabla:

Número de la dirección	12	13	14	15	16
Contenido	a	b	c	d	x

Cuadro 1

Se utiliza el código utilizado anteriormente y se supone que la primera orden tiene el número 30. Se pide que el resultado final se inscriba en la dirección 89 (cuadro 2).

La orden 00 corresponde al cierre del cálculo.

Se observa fácilmente que, en rigor, la última columna, que indica la orden siguiente en cada instrucción, es inútil. Justamente por ello, en los ordenadores reales se adopta a menudo el sistema llamado “de tres direcciones”, donde (salvo las disposiciones especiales formuladas por el código de unas operaciones, denominadas “transferencias condicionadas”, sobre las que trataremos más adelante) detrás de cada orden se considera la siguiente.

De hecho, el programa completo debe comprender tanto la introducción

Número de la orden	Código de la operación	A_1	A_2	A_3	A_4	Cálculo
30	03	12	16	40	31	ax
31	01	40	13	41	32	$ax + b$
32	03	14	16	42	33	cx
33	02	42	15	43	34	$cx - d$
34	04	41	43	89	35	$f(x) = (ax + b)/(cx - d)$
35	00					

Cuadro 2

de los datos en el ordenador como también la orden de transformar los datos iniciales expresados en sistema decimal en datos binarios —en la entrada— y la operación inversa —en la salida de los resultados de la máquina—. El programa completo se escribirá, entonces, con el sistema “de tres direcciones”:

Número de la orden	Operación	<i>I</i>	<i>II</i>	<i>III</i>	
K	Comienzo	14	$K + 1$		} Código de órdenes
$K + 1$	$10 \rightarrow 2$	$K + 10$	4	$K + 10$	
$K + 2$	\times	$K + 10$	$K + 14$	$c + 1$	
$K + 3$	$+$	$c + 1$	$K + 11$	$c + 1$	
$K + 4$	\times	$K + 12$	$K + 14$	$c + 2$	
$K + 5$	$-$	$c + 2$	$K + 13$	$c + 2$	
$K + 6$:	$c + 1$	$c + 2$	$d + 1$	
$K + 7$	$2 \rightarrow 10$	$d + 1$	0000	$d + 1$	
$K + 8$	Impresión	0000	0000	$d + 1$	
$K + 9$	Cierre	0000	0000	0000	
$K + 10$			a		} Código numérico
$K + 11$			b		
$K + 12$			c		
$K + 13$			d		
$K + 14$			x		

Cuadro 3

La primera orden (K) fija la extensión del programa (14 órdenes) y la primera orden a la que se pasa ($K + 1$). Prestando atención a los sistemas “con tres direcciones” podemos comprender más en detalle el funcionamiento de un ordenador digital, del que ahora nos interesa el esquema del dispositivo de control central (cf. la figura 11). Se compone de un llamado “registro” o mecanismo de memorización, donde se registra, durante un periodo limitado y en forma de señales eléctricas, la orden que debe ejecutarse en una determinada etapa. Cada

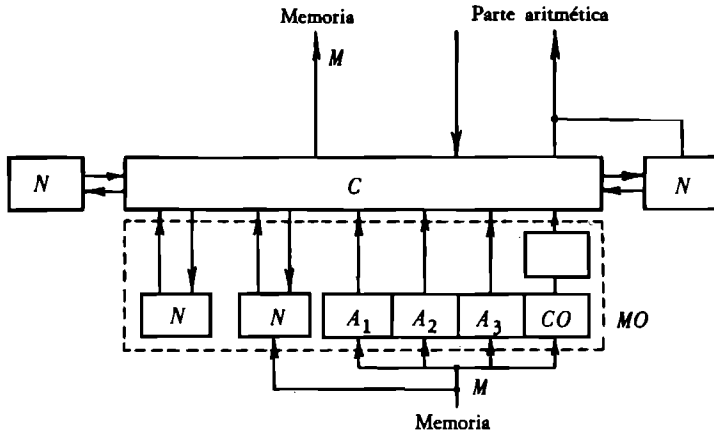


Figura 11.

grupo de señales eléctricas correspondiente al símbolo de la operación y también a las tres direcciones gobierna distintos conmutadores electrónicos, que son parte constitutiva del ordenador electrónico respectivo. De acuerdo con el símbolo de la operación, se modifica, mediante una orden electrónica, la posición de aquellos conmutadores cuya función consiste en establecer de qué manera debe tratarse la información. Así, ante la orden 01, los conmutadores se colocarán en la posición que corresponde a la suma de los números respectivos, etcétera. Las señales eléctricas correspondientes a los números que dan las tres direcciones de la orden a ejecutar establecen las posiciones de otros conmutadores gobernados electrónicamente que unen la parte aritmética a la memoria. Así, llega a la parte aritmética la información comprendida en las direcciones A_1 y A_2 , a fin de que, tras el cálculo, el resultado se dirija a la dirección A_3 .

En la técnica actual se conocen asimismo máquinas “de dos direcciones” e incluso “de una dirección”. En estos casos, las órdenes elementales comprenden la indicación de sólo dos y una dirección, respectivamente.

En las máquinas “de dos direcciones”, el resultado del cálculo puede permanecer en la parte aritmética o puede transferirse a una dirección determinada.

En estas máquinas la información comprendida en el llamado código operacional es más rica que en las máquinas “de tres direcciones”, ya que distingue más categorías de operaciones. En ciertos casos, la orden se refiere a una operación dada que debe efectuarse sobre los números de las direcciones A_1 y A_2 , y cuyo resultado se conserva en la parte aritmética. En otros casos, el código especifica la operación que debe realizarse entre el contenido de la parte aritmética y el número de la dirección A_1 . En tal circunstancia, la dirección A_2 indica el lugar de la memoria adonde debe dirigirse el resultado del cálculo.

En las máquinas con una sola dirección, la información comprendida en el código operacional es más rica todavía que en las máquinas con dos direcciones. Por regla general, en aquéllas la operación indicada por el código se efectúa entre el número comprendido en la parte aritmética (en el registro acumulador, R) y el número de la dirección expresada en la orden respectiva. Ahora bien, la orden puede también incluir la indicación de extraer información de la parte aritmética y registrarla en la dirección indicada por aquella orden. En el cuadro 4, puede consultarse el código de las máquinas CIFA-1 y CIFA-2³.

Al resolver distintos problemas, y al margen de los cálculos aritméticos, surge la necesidad de verificar una determinada condición lógica. Así, puede solicitarse que se calcule el módulo de un número dado. Pero se sabe que

$$|a| = \begin{cases} a & a \geq 0, \\ -a & a < 0, \end{cases}$$

de modo que, según sea el número a positivo o negativo, el cálculo se desenvuelve distintamente. En efecto, si $a > 0$, entonces el módulo de a es precisamente a : $|a| = a$. Pero si $a < 0$, entonces $|a| = 0 - a$.

La bifurcación del cálculo se hace introduciendo en el código de la máquina una operación extra respecto de las operaciones aritméticas, denominada *t r a n s f e r e n c i a c o n d i c i o n a d a*, que puede hallarse en “más” o en “menos”. Como se sabe, por toda regla general, después de una orden la máquina pasa a la orden siguiente. Pero si interviene la operación de transferencia condicionada, ya no se pasa automáticamente a la orden siguiente, sino que se considera antes el contenido de la parte aritmética. Si se produce una orden de transferencia condicionada a más, en forma de

$$m \quad 06 \quad n,$$

donde m es el número de la orden en el programa respectivo; 06, el código de la transferencia condicionada a más, y n , el número de una orden del programa, entonces la orden m se interpreta así: si el número de la parte aritmética es positivo, se pasa a la orden n ; en caso contrario, se ejecuta la orden $m + 1$.

3. Cf. VAIDA. *Utilizări*.

Código	Denominación	Explicación
00 - 000	Stop	El ordenador se para.
01 - x	Lectura	Se lee la palabra registrada en la memoria a la dirección x y se introduce en R .
02 - x	Suma	Se suma a la palabra de R la palabra registrada en la memoria a la dirección x ; el resultado en R .
03 - x	Resta	A la palabra de R se le resta la palabra registrada en la memoria a la dirección x ; el resultado en R .
04 - x	Multiplicación	Análogamente.
05 - x	División	Análogamente.
06 - x	Transferencia a más	Se pasa a ejecutar la instrucción de dirección x si la palabra de R es mayor o igual a 0. Si la palabra de R es menor que 0, se pasa a la siguiente instrucción del programa.
07 - x	Transferencia a menos	Análogamente.
10 - x	Transferencia incondicionada	Se pasa a ejecutar la instrucción de dirección x .
11 - x	Escritura	La palabra de R se inscribe en la memoria a la dirección x .
12 - 000	Teleimpresor*	Se pone en marcha el teleimpresor.
13 - 000	Salto	Se pasa a la instrucción siguiente del programa.
14 - 000	Desplazamiento a la izquierda	La palabra de R se desplaza una posición binaria a la izquierda; el resultado en R .
15 - 000	Desplazamiento a la derecha	Análogamente.
16 - 000	Escribir a máquina	Convertir la palabra de R al sistema octal o decimal y escribirla a máquina.
17 - 000	Desplazar el carro	Desplazar el carro de la máquina de escribir.

* Sólo en CIFA-2.

De un modo semejante pueden darse también órdenes de transferencia condicionada a menos. Como ilustración al funcionamiento general de las órdenes de transferencia condicionada, vamos a examinar el programa de cálculo para el módulo de un número x contenido en la dirección p , cuyo resultado se inscriba en la dirección 2. Se precisa que en la dirección 000 se registre el número 0. Entonces, si $x > 0$, el programa corresponde al del cuadro 5.

Número de la orden	Código	Dirección	Significación
1	01	p	Se lleva el número x de la dirección p en la memoria a la parte aritmética.
2	06	m	“Transferencia a más” y paso a la orden número m , todavía sin precisar.
m	11	2	El número de la parte aritmética se escribe en la dirección 2.
$m + 1$	16	000	Se detiene la máquina.

Cuadro 5

En cambio, si $x < 0$, se aplica el programa del cuadro 6.

Número de la orden	Código	Dirección	Significación
1	01	p	Se escribe en la parte aritmética (registro) el número 0 de la dirección 000. Se resta del número del registro, el número procedente de la dirección p .
2	06	m	
3	01	000	
4	03	p	
5	11	2	
6	16	000	

Cuadro 6

Obsérvese que las dos órdenes pueden refundirse en una sola, si tomamos $m = 5$:

1	01	p
2	06	5
3	01	000
4	03	p
5	11	2
6	16	000

Cuadro 7

5.4. Esquemas lógicos de cálculo

Debemos subrayar el hecho de que el programa para la realización de un determinado problema puede escribirse a distintos niveles.

En primer lugar, hay un nivel de resolución lógica. En esta etapa sólo interesa el algoritmo respectivo, y se ignoran todas las cuestiones relativas a la transcripción de este algoritmo al lenguaje de los ordenadores digitales.

En otro nivel, se atiende a la escritura del programa para un ordenador digital y se indica la circulación de la información por las distintas direcciones.

Por fin, en función del ordenador respectivo, es decir de su código, se escribe el programa efectivo para que sea ejecutado.

A su vez, también los métodos de escritura de programas, incluso en su forma general algorítmica, admiten diversos procedimientos. En efecto, existe, p. ej., un método llamado de los esquemas lógicos de cálculo, en el cual el algoritmo viene dispuesto en forma gráfica. En este caso, las operaciones quedan especificadas mediante distintos símbolos gráficos a fin de que el desarrollo del cálculo pueda seguirse más fácilmente. En el método intervienen varias operaciones fundamentales, que indicaremos a continuación.

Por bloque de cálculo se entiende el cálculo de una expresión en función de una regla dada. Así, p. ej., el algoritmo para calcular la expresión

$$f(x) = x^5 - 3x + 1 = Ax$$

se representa gráficamente en la figura 12.

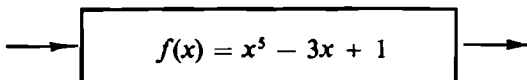


Figura 12

Las condiciones lógicas se indican mediante un rombo con una entrada y dos salidas (cf. figura 13), en el que se anota, eventualmente, la condición lógica que debe verificarse. Las dos salidas corresponden a las dos alternativas resultantes.

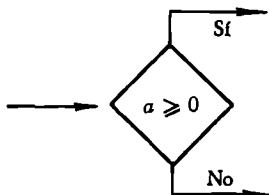


Figura 13

En el curso de los cálculos, surge a menudo la posibilidad de operaciones indiciales, tal como ocurre, p. ej., cuando se calculan los términos de una serie. Si los índices proceden en razón 1, ello se indica a base de

$$i + 1 \rightarrow i,$$

lo que significa que en el programa todos los índices i se sustituyen por los índices $i + 1$. Simbólicamente, esta operación se presenta como en la figura 14.

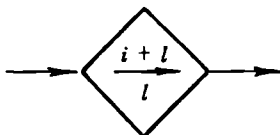


Figura 14

Los datos de entrada se expresan por medio de un círculo en el que se escribe IN, mientras que la terminación de los cálculos se indica con un círculo donde se escribe la palabra STOP (cf. figuras 15a y 15b).

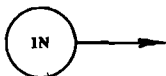


Figura 15a

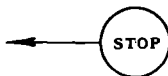


Figura 15b

Cuando quiere ponerse de manifiesto el hecho de que, en una determinada etapa del cálculo, la máquina continúa funcionando a partir de una ramificación del programa, llamada *subprograma* o *subrutina*—lo que, en rigor, es un caso análogo al de calcular una expresión cualquiera según una regla dada—, ello se especifica como en la figura 16.

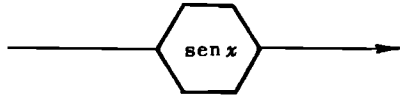


Figura 16

Es necesario, por lo demás, indicar el bloque para la forma inicial, que representa aquella parte del programa en función de la cual las instrucciones, los parámetros o las direcciones, cuyo contenido ha cambiado, son devueltas a su forma inicial. En la figura 17 se representa el bloque para la forma inicial que

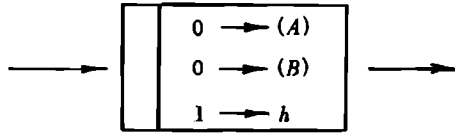


Figura 17

exige anular el contenido de las direcciones A y B, mientras que el parámetro h es llevado al valor 1. El contenido de las direcciones se anota mediante letras entre paréntesis.

Para simplificar el dibujo de un determinado esquema lógico de cálculo, se recurre al **conector fijo** (cf. figura 18a), donde entre los dos círculos



Figura 18a

notados con la misma letra griega existe una línea que no se dibuja. Las ramificaciones del programa se indican mediante el **conector variable** (cf. figura 18b), que consta de un círculo terminal marcado con una letra griega

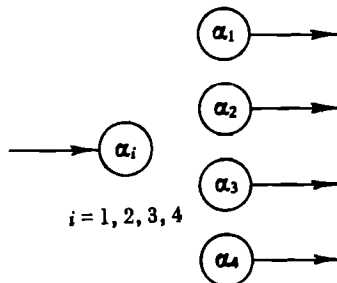


Figura 18b

subindicada por una letra latina. Las distintas continuaciones posibles se expresan, a su vez, por círculos marcados con la misma letra griega subindicada por cifras arábigas.

Antes de ejemplificar la composición de un programa por este método de esquemas lógicos de cálculo, consideremos el esquema lógico por el que se forma un reflejo condicionado. Como se sabe, en los reflejos condicionados se ponen de manifiesto dos estímulos, que notaremos a base de a y b . Normalmente, al comienzo de la experiencia, el estímulo a produce tan sólo la reacción A y, a su vez, el estímulo b , la reacción B , pero no A (cf. 4.8). Así, p. ej., si se introduce alimento (estímulo a) en la boca de un animal, éste segrega saliva (reacción A). Pero si suena un cascabel (estímulo b), el animal atiesa las orejas (reacción B) sin segregación de saliva. Por tanto, al iniciar el experimento, el estímulo b no produce la reacción A .

En las experiencias que determinan la formación de reflejos condicionados, se llega a una situación tal que el estímulo b produce por sí solo la reacción A , incluso en ausencia de a . El reflejo condicionado presupone una situación según la cual los estímulos a y b aparezcan simultáneamente, es decir supone los eventos $c = a \wedge b$.

Admitamos que los experimentos para la formación del reflejo condicionado hayan tenido lugar en los momentos t_n . Notando mediante S_n el estado del animal en el momento t_n , puede escribirse:

$$S_n = S_{n-1} + d_n + e_n,$$

donde e_n es un número aleatorio tal que $\lim_{N \rightarrow \infty} \frac{\sum_1^N e_n}{N} = 0$, y d_n , un número que depende de la operación simultánea de los estímulos a y b según la ley⁴

$$d = \begin{cases} 0 & b = 0 \\ +f & ab = 1 & (f > 0) \\ -g & a'b = 1 & (g > 0), \end{cases}$$

donde se ha adoptado la siguiente notación: una variable (a o b) tiene, en un momento t_n , valor 1 si existe en aquel momento y valor 0 si no existe; a' expresa la negación de a , de modo que $a' = 1$ si $a = 0$, y $a' = 0$ si $a = 1$.

En otras palabras, la aparición simultánea de a y b fortalece el reflejo condicionado, mientras que la aparición de b , en ausencia de a , lo debilita. A ello se le añade un ruido casual e . El estado del animal evoluciona a lo largo del tiempo. Si en un momento dado se han producido suficientes asocia-

4. [En la expresión anterior se indica, pues, que el límite, cuando N tiende a infinito, de la serie $\frac{\sum_1^N e_n}{N}$ ($n = 1 \dots N$) es cero].

ciones de los estímulos a y b , entonces $S > T$, donde T es un valor de umbral dado, y la sola presencia de b es suficiente para desencadenar la reacción A .

El organigrama que representa la formación del reflejo condicionado tal como acabamos de exponer viene dado en la figura 19.

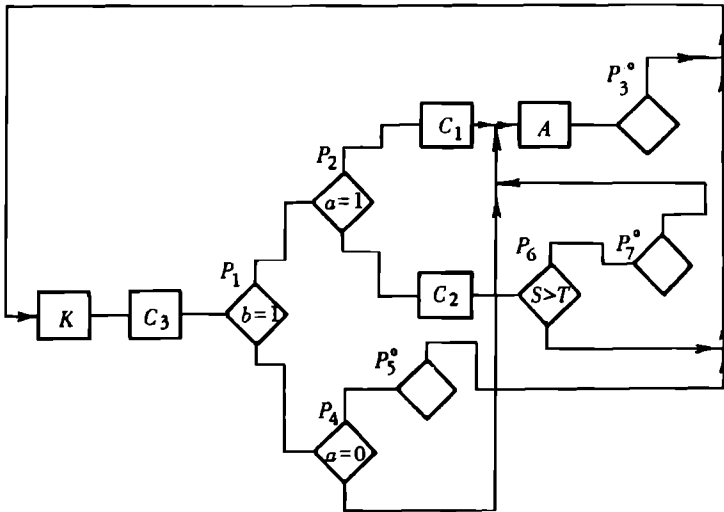


Figura 19

En cada momento t_n , el generador K produce un par de valores booleanos a y b . Los bloques de cálculo C_1 y C_2 permiten el paso de S a $S + f$ y a $S - g$, respectivamente; con C_3 se pasa a $S + e$, donde e viene generado independientemente por esta fuente K representada en el organigrama. A es el operador que produce la reacción A . Las condiciones lógicas están notadas en la figura: P_1 corresponde a $b = 1$; P_2 , a $a = 1$; P_3 , P_5 y P_7 son símbolos de paso incondicionado; P_4 corresponde a $a = 0$, y P_6 , a $S > T$.

Este mismo organigrama puede expresarse mediante un grafo, como el de la figura 20, donde β , α , γ indican la realización de las condiciones P_1 , P_2 y P_6 , y δ' , la no realización de la condición δ ($\delta = \alpha, \beta, \gamma$).

Otro ejemplo de esquema lógico se refiere a la resolución de una ecuación por aproximaciones sucesivas. Dada la ecuación $F(x) = 0$, se pide determinar la raíz que existe en el intervalo $a \mapsto b$. Supongamos que en este intervalo la ecuación tiene una sola raíz con un grado impar de multiplicidad, de modo que, en las extremidades del intervalo, la función $F(x)$ cambia de signo. El método consiste en considerar la mitad del intervalo $a \mapsto b$ —digamos $c = (a + b) / 2$ (cf. la figura 21)— y en determinar el signo del valor de $F(c)$. Entre estos dos nuevos

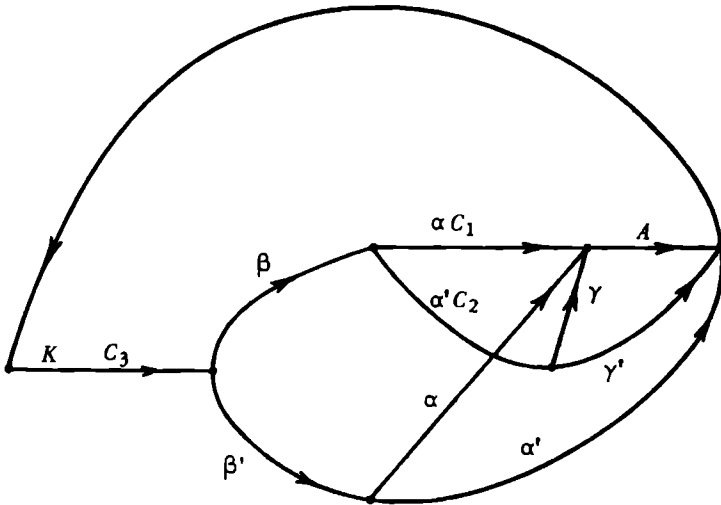


Figura 20

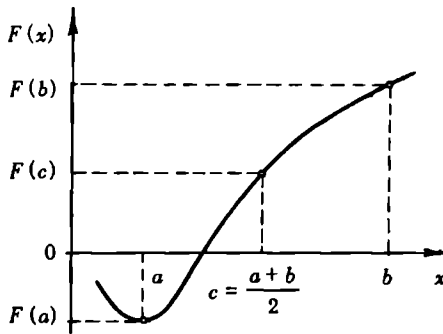


Figura 21

intervalos creados, $a \rightarrow c$ y $c \rightarrow b$, se elige aquel en cuya extremidad la función tiene signo distinto (condición [6], en la figura 22) y se repite luego el proceso de un modo cíclico. En estas repeticiones sucesivas, cada vez se verifica (9) en tanto que la diferencia entre las extremidades del intervalo sobrepase o no la precisión con que queramos determinar la raíz. En cuanto se llega al valor receptivo, el proceso se detiene. El esquema lógico de cálculo que corresponde a este método aparece en la figura 22.

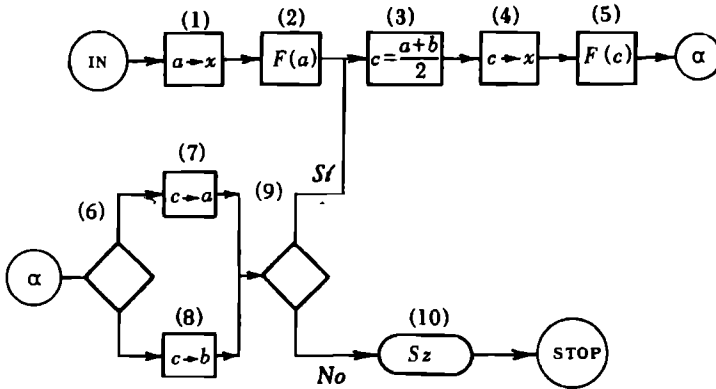


Figura 22

5.5. Algoritmos de resolución

En el método operacional se ponen de manifiesto diversas categorías de elementos: datos iniciales, operadores, condiciones lógicas, resultado final. Si se atribuye el símbolo y al resultado final y el símbolo x a los datos iniciales, entonces cabe escribir en forma abreviada un programa:

$$xA = y,$$

donde A es el operador que, aplicado a los datos iniciales, conduce al resultado buscado. En caso en que el operador A consista en la aplicación sucesiva de varios operadores A_i , se escribe⁵:

$$A = A_1 A_2 \dots A_n = \prod_{i=1}^n A_i.$$

Ahora bien, tal como se ha dicho, en los cálculos también intervienen a veces ciertas condiciones lógicas, que notaremos mediante P_j (cf. 4.2 y ss.).

Una fórmula operacional tiene, por lo común, el aspecto

$$\prod A_i P_1 \uparrow B \downarrow C!$$

Este simbolismo se interpreta en el sentido de que los operadores se aplican de izquierda a derecha. Así, sobre los datos iniciales se aplica primero la sucesión de operadores A_i , y luego se verifica la condición lógica P_1 . En caso de satisfacerse ésta, se pasa al operador B . Pero si no se satisface, entonces, tal como in-

5. [La expresión \prod indica justamente orden de multiplicación para A_i , entre los valores comprendidos de 1 a n para i].

dica el vector que va tras P_1 , se pasa al operador C , después de lo cual se detiene la sucesión de cálculos, punto indicado por el signo !. Si P_1 se verifica, se aplica antes el operador B y a continuación C .

Esta sucesión de operaciones puede representarse a base de la fórmula

$$A_i P_1 \xrightarrow{\quad} C \longrightarrow !$$

(Note: In the original image, an arrow labeled 'B' points from P1 to C, and another arrow points from C to '!')

Por ejemplo, el organigrama que determina la formación del reflejo condicionado (figura 19) admite la siguiente representación algorítmica (cf. 4.8):

$$\downarrow\downarrow KC_3 P_1 \uparrow P_2 \uparrow C_1 \downarrow\downarrow AP_3^0 \uparrow\downarrow P_4 \uparrow P_5^0 \uparrow\uparrow C_2 P_6 \uparrow P_7^0 \uparrow.$$

5.6. Automatización de las programaciones

Para que un problema dado sea eficazmente resuelto por un ordenador, es indispensable disponer de un programa completo para el tratamiento de los datos iniciales. En los primeros ordenadores los programas se escribían efectivamente, teniendo en cuenta el código de la máquina.

Pero el paso del esquema lógico —en forma operacional o bien de organigrama— al programa propiamente dicho es una operación laboriosa y rutinaria. Precisamente por ello los ordenadores modernos están previstos con capacidad de autoprogramación. Para ello, se escribe primero el programa lógico sobre el desarrollo de los cálculos, sin prestar atención a las particularidades concretas de la máquina respectiva, pero utilizando un alfabeto “comprensible” para aquella. Las órdenes se introducen en ella en forma lo más cercana posible al lenguaje matemático habitual. La máquina está provista, como decimos, de la facultad de interpretar estas órdenes y de redactar sola el programa adecuado pasando de las instrucciones dadas a las que se ejecutan efectivamente.

Entre los lenguajes empleados en la automatización de las programaciones, destaca en primer lugar el ALGOL-60⁶. Este lenguaje utiliza los siguientes grupos de símbolos fundamentales: cifras, letras, valores lógicos y limitadores. Las cifras son arábigas (10, en total) y las letras, las del alfabeto inglés, minúsculas o mayúsculas (52, en total). Hay dos valores lógicos: **true**, ‘verdadero’, y **false**, ‘falso’. Conviene puntualizar desde el principio que en la lengua ALGOL-60, todas las palabras utilizadas están tomadas del inglés. A su vez, los limitadores son de diversas categorías: operadores, paréntesis, signos de puntuación, especificadores. Las operaciones aritméticas empleadas son: + suma, — resta, × multiplicación, / división, ↑ elevación a potencia. Las relaciones de órdenes consideradas son las acostumbradas:

$$<, \geq, =, \neq, >, \leq.$$

6. Cf. BOTTENBRUCH. “Structure”.

Los operadores lógicos considerados son: \vee disjunción, \wedge conjunción, \neg negación, \supset implicación, \equiv equivalencia, definidas por las siguientes tablas de verdad, donde 1 y 0 representan, respectivamente, los valores lógicos “verdadero” y “falso”:

A	0 1	A	B	$A \wedge B$	$A \vee B$	$A \supset B$	$A \equiv B$
$\neg A$	1 0	0	0	0	0	1	1
		0	1	0	1	1	0
		1	0	0	1	0	0
		1	1	1	1	1	1

Cuadro 8

Además de los signos, la lengua ALGOL-60 emplea también 22 palabras tomadas del inglés, que se escriben en **negrita**. Se trata de las siguientes:

go to (ir a); **if** (si); **then** (entonces); **else** (de otro modo); **for** (para); **do** (ejecutar); **step** (paso); **until** (hasta); **while** (mientras); **comment** (comentario); **begin** (comienzo); **end** (final); **own** (propio); **Boolean** (booleano); **integer** (entero); **real** (real); **array** (ordenación); **switch** (conmutador); **procedure** (procedimiento); **string** (serie); **label** (marca); **value** (valor).

Para ilustrar el empleo del ALGOL-60, examinaremos algunos ejemplos.

EJEMPLO 1. Dados los números a, b, c , se pide calcular $y = \frac{a \cdot b + c}{a \cdot b - 4c}$

Formulación del proceso en el lenguaje ALGOL	Explicación
Entrada (a, b, c)	Dados tres números, introducidos en la entrada de dispositivo, se denominan a, b, c .
$h = a \times b$	Se calcula el producto $a \cdot b$ y el resultado se nota mediante h .
$g: (h + c)/(h - 4c)$	Se calcula la expresión $\frac{a \cdot b + c}{a \cdot b - 4c}$ y el resultado se nota mediante g .
Salida (g)	Se perfora a la salida el resultado denominado g en la etapa precedente.

Cuadro 9

Este programa también puede escribirse a base de:

Entrada (a, b, c); $h = a \times b$; $g = (h + c)/(h - 4 \times c)$; *salida* (g), donde las palabras *entrada* y *salida* son identificadores u operadores de procedimiento.

El cálculo de la expresión aducida puede hacerse también con ayuda de un programa distinto, en el que se tiene en cuenta la posibilidad de que la variable a tenga valor 0:

```
L: entrada (a, b, c)
if a = 0 then go to M;
h: = a × b;
salida ((h + c)/(h - 4 × c))
go to L;
M: salida (-0.25).
```

En este tipo de programación se verifica primero la condición lógica $a = 0$. En caso de que se cumpla la condición, se pasa directamente a M . En caso contrario, se imprime a la salida el resultado g , después de lo cual se vuelve a L .

Conviene señalar que en el lenguaje ALGOL no se pone coma entre 0 y los números decimales, sino punto. Se escribe, pues, -0.25 y no $-0,25$, etc.

EJEMPLO 2. Puede plantearse el problema de perforar una tablilla con los valores de la expresión

$$\frac{3a + 4}{5a + 6}$$

para $a = 1, 2, \dots, 100$.

Existe la posibilidad de especificar el programa con o sin la introducción de un ciclo. Así, pues, el programa puede darse de las dos maneras siguientes:

```
1. for a: = 1 step 1 until 100 do
   salida ((3 × a + 4)/(5 × a + 6))
```

o bien,

```
2. a: = 1
   L: salida ((3 × a + 4)/(5 × a + 6))
   a: = a + 1;
   if a ≤ 100 then go to L.
```

El segundo programa ha puesto de manifiesto un operador de ciclización.

Para las potencias existe una notación especial. Por ejemplo, el programa de cálculo para las expresiones

$$\frac{2 \times i + 3}{4 \times i + 5} \text{ y } \left(\frac{2 \times i + 3}{4 \times i + 5} \right)^2$$

para $i = 1, 2, \dots, 100$, se escribe así:

```
for  $i$ : = 1 step 1 until 100 do
  begin  $h$ : =  $(2 \times i + 3) / (4 \times i + 5)$ ;
  salida ( $h$ ,  $h \uparrow 2$ )
  end
```

Hay que subrayar el hecho de que en lugar de escribir h^2 , se utiliza la notación $h \uparrow 2$. En general, si E_1 y E_2 son dos expresiones aritméticas, entonces la expresión $E_1^{E_2}$ en lenguaje ALGOL se indica por medio de $(E_1 \uparrow E_2)$.

El conmutador se adopta para una escritura más simplificada del programa. Supongamos que en un programa para el ALGOL hay que incluir un operador que pueda producir la aplicación de cinco algoritmos diferentes, sean $L, P, Q, L2, L1$, en correlación con i , que tiene uno de los valores 1, 2, 3, 4, 5, y que el programa dado sea:

```
if  $i = 1$  then go to  $L$  else if  $i = 2$  then go to  $P$ 
else if  $i = 3$  then go to  $Q$  else if  $i = 4$  then go to  $L2$ 
else if  $i = 5$  then go to  $L1$ .
```

Este programa puede presentarse con mayor sencillez mediante el operador de conmutación:

```
switch  $s = L, P, Q, L2, L1$ .
```

El operador que dirige a s se escribe

```
go to  $s [i]$ .
```

A lo largo del programa intervienen también comentarios. El símbolo **comment** va seguido de una explicación, en lenguaje habitual, sobre el significado de algunas letras. Puede escribirse, p. ej.,

comment *si en este punto el programa se detiene, es que hay un error en los datos iniciales.*

if $a = b$ **then begin** $x : g; y : = h$ **end.** *Se trata de un caso muy raro.*

También existe la posibilidad de componer una descripción formal para el lenguaje ALGOL, pero no insistiremos sobre ello aquí.

Precisemos que, en general, en ALGOL se utilizan las siguientes notaciones: expresión aritmética, E ; expresión booleana, B ; una expresión cualquiera, G ; operador, S ; constante, V ; operador no condicionado, U ; elemento de un ciclo, L ; identificador, I ; descripción, D .

Veamos un ejemplo de escritura en lenguaje ALGOL para un programa más complejo.

EJEMPLO 3. Se plantea calcular la expresión

$$E = \sum_{s=0}^{10} \frac{1}{s! (n+s)!} \left(\frac{x}{2}\right)^{2s+n},$$

lo que representa los primeros 11 términos de la función Bessel modificada, para $n \geq 0$. El correspondiente programa se escribe así ("fac" es el operador factorial):

```

begin real x0, delta x, x max, g, term, sum, x; integer n, s, i, n fac;
entrada (x0, delta x, x max, n);
for x: = x0 step delta x until x max do
  begin n fac: = 1;
  for i: = 2 step 1 until n do n fac: = n fac x i;
  comment en el cálculo de n! se conoce de antemano el valor para s = 0;
  if x = 0  $\wedge$  n = 0 then begin sum: = 1; go to P end;
  sum: = 0; g: = (x/2)  $\uparrow$  2; term: = (x/2)  $\uparrow$  n/n fac;
  comment la variable sum significa suma parcial;
  for s: = 0 step 1 until 10 do
    begin sum: = sum + term;
      term: = term  $\times$  g/((s + 1)  $\times$  (s + 1 + n))
    end;
  end:
  P: salida (x, n, sum)
end
end

```

EJEMPLO 4. En ALGOL, el término **procedure** representa un subprograma. Si, p. ej., se solicita calcular el área de los triángulos con los vértices en tres de los cuatro puntos dados a, b, c, d , puede escribirse el siguiente programa (don-

de el operador “*sqrt*” indica la extracción de la raíz cuadrada, en inglés *square root*):

```

begin real a, b, c, d;
  procedure área del triángulo (x, y, z);
    begin real s, a;
      s := 0.5 × (x + y + z);
      if s ≥ x ∧ s ≥ y ∧ s ≥ z then
        begin a := sqrt (s × (s - x) × (s - y) × (s - z));
          salida (a)
        end
      else salida (-1)
    end área del triángulo;
  L. entrada (a, b, c, d);
  área del triángulo (a, b, c);
  área del triángulo (a, b, d);
  área del triángulo (a, c, d);
  área del triángulo (b, c, d);
  go to L
end

```

El lenguaje ALGOL-60 se emplea hoy a gran escala en los países europeos para la escritura de programas en ordenadores digitales.

6. El lenguaje cósmico

6.1. Introducción

En nuestra época, cuando el hombre ha logrado ya penetrar en el espacio extraterrestre y ha conseguido dominar energías gigantescas gracias a los notables progresos de la ciencia y de la técnica, se hace factible entablar comunicación a distancias realmente cósmicas. En la actualidad se emiten desde la Tierra señales electromagnéticas de alta intensidad, que podrían ser interpretadas, por seres extraterrestres dotados de inteligencia, como prueba de la existencia, sobre nuestro planeta, de unos seres evolucionados que tratan de establecer contacto con otros seres del universo llegados a un alto grado de civilización. Se intenta igualmente conseguir la recepción de señales eventualmente emitidas con el mismo fin por seres extraterrestres.

El establecimiento de comunicaciones bilaterales a escala cósmica plantea una serie de dificultades.

Para poder transmitir un mensaje con significado, el hombre recurre generalmente a una lengua determinada, la cual exige un aprendizaje previo en todas sus partes. En la Tierra, sin embargo, este proceso de aprendizaje se desarrolla en condiciones distintas con respecto al cosmos, ya que cabe convenir que un mismo objeto —conocido tanto por el que aprende como por el que enseña— reciba una misma denominación. Así ocurre, p. ej., con *libro*, donde se establece una asociación entre la forma sonora o escrita de la palabra y el objeto respectivo. Del mismo modo se procede para explicar el significado de las distintas acciones.

En unas condiciones comunicativas a nivel cósmico no existe, evidentemente, la posibilidad de presentar los objetos o las acciones como tales a fin de asociarlas a determinados símbolos sonoros o escritos. Por ello, el problema del lenguaje cósmico requiere un planteamiento muy distinto al que se produce en el aprendizaje de una lengua extranjera.

El Dr. Hans Freudenthal, Profesor de Matemáticas en la Universidad de Utrecht, ha tratado de establecer un método racional que permita a los eventuales extraterrestres la comprensión del pensamiento humano. Así ha llegado a la lengua *Lincos*¹.

La idea se basa en establecer relaciones lo más evidentes posible entre ciertos símbolos y conceptos muy generales, susceptibles de ser interpretados en la práctica de un modo único. A partir de ello, se desarrolla a continuación el curso de lenguaje cósmico.

Como los conceptos más generales que admiten una transmisión cómoda en forma de señales simples son los aritméticos, H. Freudenthal intentó construir el lenguaje entero a partir de la aritmética. Para escribir en *Lincos* conviene, en primer lugar, sentar un sistema de escritura cómodo.

Con este propósito, cabe utilizar un alfabeto de tipo Morse formado por señales de distinta duración. A base de estos símbolos se construyen los demás símbolos, secundarios, y dotados de un significado bien determinado. En lo que sigue, damos sólo una representación gráfica de los símbolos secundarios.

He aquí el comienzo del curso de *Lincos*.

Al principio, se transmiten grupos de palabras sin sentido, espaciadas, para facilitar su reconocimiento en contextos. A continuación, se transmiten grupos de señales que expliquen el significado de ciertos conectivos:

1. # > ... # etc.
2. # ... < # etc.
3. # = # etc.
4. # + .. = # etc.
5. # - .. = # etc.

El signo # marca el comienzo y el final de un texto.

La fórmula en metalengua "etc." muestra que en el texto sólo se indica un ejemplo de cada serie.

Se introduce luego la escritura de los números, empleando, p. ej., el sistema binario. Así, escribiremos:

$$\begin{aligned} \# . &= 1 \\ \# .. &= 10 \\ \# ... &= 11 \\ \# &= 100 \\ \# &= 101 \\ \# &= 110 \\ \# &= 1101 \# \text{ etc.} \end{aligned}$$

1. Cf. FREUDENTHAL, *Cosmic Language* y, sobre todo, *Lincos*.

Se introduce después un algoritmo de adición. Se escribirá, por ejemplo:

$$\begin{aligned} \# 111 &= 110 + 1 = 101 + 10 = 100 + 11 = 11 + 100 = \\ &= 10 + 101 = 1 + 110 \# \end{aligned}$$

Los ejemplos de este tipo se repiten muchas veces con objeto de facilitar el significado del símbolo “+”.

Conviene notar que en este método el significado de los símbolos resulta sólo a partir de los ejemplos, desde el momento en que no existe la posibilidad de hacer aclaraciones en alguna metalengua. Justamente por este motivo los ejemplos deben ser suficientemente numerosos para que la interpretación de los símbolos sea unívoca.

Se observa sin dificultad que si en nuestra transcripción parece evidente la interpretación de los signos en Lincos que hemos presentado hasta ahora, ello se debe al hecho de que en nuestra mente existen, de antemano, asociaciones precisas entre el signo “=” y su significado, lo mismo que entre los signos “1” y “0” y la transcripción de los números en sistema binario. Pero, en rigor, las señales recibidas en Lincos ofrecen una forma absolutamente diferente. En efecto, si se modifican las notaciones tal como sigue:

alfabeto usual	.	1	0	=	#
nuevo alfabeto	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i> <i>f</i>

entonces las relaciones precedentes, por las que se ha introducido la escritura binaria de los números, se convierte en

f a b e c a
a b b e c d a
a b b b e c c a
a b b b b e c d d a f

Estas series ya no son tan sugestivas como cuando se empleaban los símbolos familiares 0, 1, =, etc.

Este problema tiene que ver con otro más complejo todavía, que consiste en la interpretación de las señales o, en una formulación más adecuada a nuestra situación, al desciframiento de textos con significado desconocido (cf. 3.26).

Desde luego, cuanto mayor es la extensión de un texto, tanto mayores son las posibilidades de descifrarlo correctamente.

Volviendo al curso de Lincos, veamos cómo se introduce la noción de variable. Con este propósito, se transmiten los mensajes:

$$\begin{aligned} \# 100 &> 10 \\ 100 + 11 &> 10 + 11 \\ 100 + 1 \ 101 &> 10 + 1 \ 101 \\ 100 + 1 &> 10 + 1 \\ 100 + 110 &> 10 + 110 \\ 100 + 11 \ 111 &> 10 + 11 \ 111 \\ 100 + a &> 10 + a \ \# \text{ etc.} \end{aligned}$$

La última relación introduce el nuevo concepto de variable, descrito, en el caso presente, por a .

Se puede indicar con bastante facilidad la propiedad conmutativa. Para ello, se transmiten mensajes como:

$$\begin{aligned} \# 100 + 1 &= 1 + 100 \\ 100 + a &= a + 100 \\ a + b &= b + a \ \# \end{aligned}$$

La transmisión de los mensajes presupone también la existencia de un sistema de puntuación. En la lengua hablada, la puntuación queda expresada por la pausa y en la escrita por medio de signos gráficos especiales. En Lincos se puede convenir igualmente que determinados símbolos se pongan en relación con signos de puntuación. Pero a nosotros nos interesa, en primer lugar, establecer el valor de estos símbolos.

Como es costumbre en la lógica matemática, los signos de puntuación separan entre sí grupos de símbolos. Imaginemos que hemos de utilizar paréntesis de distintos órdenes. Normalmente, en el álgebra elemental se emplean a lo sumo paréntesis de tres o cuatro órdenes, provistos por signos gráficos especiales.

En el Lincos se han establecido 32 tipos de signos equivalentes a los paréntesis. A continuación, presentamos una tabla donde se indican los signos gráficos y el valor jerárquico del signo respectivo en tanto que paréntesis.

$$\begin{aligned} &. 1 \cdot 2 : 3 \cdot 4 : 5 : 6 : 7 \\ &| 8 | \cdot 9 | \cdot 10 | : 11 | \cdot 12 | : 13 | : 14 | : 15 \\ &| 16 | \cdot | 17 | \cdot | 18 | : 19 | \cdot 20 | : 21 | : 22 | : 23 \\ &|| 24 | \cdot | 25 | || \cdot 26 | || : 27 | || \cdot 28 | || : 29 | || : 30 | || : 31 \end{aligned}$$

Las conectivas utilizadas en Lincos adoptan diferentes formas en algunos casos.

Así, se define el símbolo de implicación, \rightarrow , por medio de las relaciones:

$$\begin{aligned} \# a > 100. \rightarrow .a > 10 \\ a > 101. \rightarrow .a > 1 \\ a > 11. \rightarrow .a > 11 \# \end{aligned}$$

etcétera.

Se define luego la implicación bilateral, \leftrightarrow , a base de:

$$\begin{aligned} \# a > 100. \leftrightarrow .a > 100: \\ a > 100. \leftrightarrow 100 < a \\ a > b \leftrightarrow b < a \\ a + b = c \leftrightarrow 'a = .c - b \# \end{aligned}$$

A continuación se puede definir la relación de desigualdad mediante ejemplos de tipo

$$\# 1 \neq 10. 1 \neq 11. 1 \neq 100. 1 \neq 110 \ 111 \#$$

junto con algunas propiedades de dicha relación:

$$\begin{aligned} \# a > b. \rightarrow 'a \neq b' \\ a < b. \rightarrow .a \neq b' \\ a \neq b. \rightarrow 'a + c. \neq .b + c \# \text{ etc.} \end{aligned}$$

Se introduce después la disyunción lógica mediante relaciones de tipo

$$\begin{aligned} \# a \neq 101. \rightarrow 'a < 101. \vee .a > 101: \\ a \neq b. \rightarrow 'a < b. \vee .a > b: \\ a = b. \vee .a < b. \vee .a > b \#, \end{aligned}$$

siendo la última relación una tautología.

Para la relación \leq , se transmitirá:

$$\begin{aligned} \# 1 \leq 1 . 1 \leq 10 . 1 \leq 11 . 1 \leq 10 \ 101. \\ 10 \leq 10 . 10 \leq 11 \cdot 10 \leq 111 \quad \text{etc.,} \# \end{aligned}$$

y, para la conjunción lógica, \wedge , se indicará:

$$\begin{aligned} \# a \leq 111 . \wedge . b \leq 10 \rightarrow 'a + b. \leq 1 \ 001: \\ a > 110 . \wedge . b \geq 10 \rightarrow 'a + b. > 1 \ 000 \# \end{aligned}$$

a lo cual se añadirán algunas de sus propiedades:

$$\# a = b . \wedge . c = d \rightarrow . a = b$$

$$a = b . \wedge . c = d \rightarrow . c = d . \wedge . a = b$$

$$a = b . \wedge . c = d \wedge . e = f : \leftrightarrow : a = b . \wedge c = d . \wedge e = f \#$$

Para introducir el símbolo ? se transmitirá:

$$\# ? x . x + 10 = 111$$

$$x + 10 = 111 . \rightarrow . x = 101$$

$$? x : a < b . \wedge . x + a . = b$$

$$a < b . \wedge . x + a . = b : \rightarrow . x = . b - a \# \text{ etc.}$$

Conviene notar que las últimas relaciones introducen una idea nueva consistente en una magnitud desconocida, o incógnita, que satisface una determinada relación. De este modo surge la posibilidad de formular preguntas. En otras palabras, al contrario de los métodos simbólicos usuales, aquí se posibilita la formulación simbólica de preguntas. Recordemos que, en la matemática usual, lo que se formaliza es un conjunto de proposiciones verdaderas y no se pretende descubrir tales proposiciones mediante el acto de responder a una pregunta formalmente expresada.

Volviendo a la aritmética, cabe introducir luego la noción de operador de sustracción y de número negativo:

$$\# 11 - 1 . = 10:$$

$$11 - 10 . = 1:$$

$$11 - 11 . = 0 \#$$

Fácilmente se observa que así se ha introducido también el concepto de cero.

En todo caso, pueden añadirse las leyes de la adición y de la sustracción de los números enteros, operaciones que se realizan sin dificultad.

Se pasa después a la ordenación de los enteros negativos y a la regla para trabajar con variables y con el operador -:

$$\# - . -a = a \#$$

A continuación, se dan indicaciones sobre la multiplicación con enteros, con 0 y con números negativos.

Prosiguiendo, se pasa a definir la operación de división:

$$\# 1100/10 = 110.$$

$$110/10 = 11.$$

$$11/10 = 1,1.$$

$$1,1/0 = 0,11 \#$$

En el sistema Lincos, la idea “y así sucesivamente” se indica abreviadamente por medio de “Etc.”. Hay que precisar que en los textos en Lincos dados hasta el momento ha intervenido sólo etc., expresado en metalengua, como indicación para el que transmitía de que existen otros ejemplos del género respectivo. La palabra “Etc.” se introduce en el Lincos mediante fracciones decimales periódicas, a base de ejemplos como:

$$\begin{aligned} \# 1/11 &= 0,01010101 \text{ Etc.} \\ 1/101 &= 0,0011001100110011 \text{ Etc.} \\ 1/110 &= 0,001010101 \text{ Etc.} \end{aligned}$$

Digamos también que en el lenguaje Lincos las abreviaciones se hacen, en general, partiendo de la denominación latina de los distintos conceptos. Así, el símbolo Num (fL n u m e r u s) contiene el significado de “número natural” (la abreviatura “fL” significa *from Latin*, es decir ‘del latín’) y la palabra Int (fL i n t e g e r) se refiere a “número entero”.

Se escribirán así:

$$\begin{aligned} \# a \in \text{Num} &\leftrightarrow a = 1 \cdot \vee a = 10 \cdot \vee a = 11 \cdot \vee a = 100 \\ &\vee a = 101 \cdot \vee \text{ Etc:} \\ a \in \text{Int} &\leftrightarrow a = 0 \cdot \vee a = 1 \cdot \vee a = -1 \cdot \vee a = 10 \cdot \\ &\vee a = -10 \cdot \vee a = 11 \cdot \vee \text{ Etc.} \# \end{aligned}$$

Una nueva conectiva, \wedge , con valor de “para todo”, se escribirá así:

$$\begin{aligned} \# a = 1 \cdot \rightarrow a^{10} > 0 \cdot \wedge a = 10 \cdot \rightarrow a^{10} > 0 \cdot \wedge a = 11 \cdot \\ &\cdot a^{10} > 0 \cdot \wedge \text{ Etc.:} \\ \wedge a \cdot a \in \text{Num} &\rightarrow a^{10} > 0 \# \end{aligned}$$

De un modo análogo —mediante una disyunción infinita— puede introducirse el símbolo #, que significa “existe un elemento que”. Se escribirá:

$$\begin{aligned} \# a = 1 \cdot \wedge x = 10a \cdot \vee a = 10 \cdot \wedge x = 10a \cdot \\ \vee a = 11 \cdot \wedge x = 10a \cdot \vee a = 100 \cdot \wedge x = 10a \cdot \vee \text{ Etc:} \\ \leftrightarrow a \in \text{Num} \cdot \wedge x = 10a \cdot \\ x = 10 \times 1 \cdot \vee x = 10 \times 10 \cdot x = 10 \times 11 \cdot \vee x = \\ = 10 \times 100 \cdot \vee \text{ Etc:} \\ \leftrightarrow : \vee a \cdot a \in \text{Num} \cdot \wedge x = 10a \# \end{aligned}$$

Y así continúa la construcción de toda la aritmética.

Para facilitar la escritura se introducen palabras correspondientes a los conceptos de divisor (Div), número positivo (Pos), número primo (Pri), núme-

ro racional (Rat), número real (Rea), número complejo (Com), junto con el símbolo de la inclusión: $\text{Int} \supset \text{Num} \supset \text{Pri}$.

Los números racionales se introducen formalmente mediante una definición que coincide en la práctica con la de Dedekind.

Con arreglo al método general, después de haber operado con conjuntos Int , Num , Pri y con relaciones entre conjuntos (\supset), se introduce también el concepto de conjunto (Agg). Adoptando el punto de vista de Quine, se escribirá $a \in a$ para un conjunto que tenga un solo elemento, es decir a .

Puede darse luego un método que transforme un conjunto en otro conjunto con un solo elemento. Se trata del signo $\ulcorner \urcorner$, que transforma un conjunto en un conjunto de un solo elemento, es decir el conjunto respectivo:

$$\begin{aligned} \# \forall C . C \in \text{Agg} . \wedge . \text{Pri} \in C : \\ \wedge x \cdot x \in C . \leftrightarrow . x = \text{Pri} : \rightarrow . C \in \text{Agg} : \\ \wedge x \cdot x \in C . \leftrightarrow . x = \text{Pri} : \leftrightarrow . C = \ulcorner \text{Pri} \urcorner : \\ A \in \text{Agg} . \rightarrow \wedge x \cdot x \in \ulcorner A \urcorner . \rightarrow . x = A : \wedge \cdot \ulcorner A \urcorner \in \text{Agg} : \\ x \in \ulcorner x \urcorner : \quad \bullet \\ y \in \ulcorner x \urcorner \cdot \leftrightarrow \cdot x = y \# \end{aligned}$$

Las relaciones entre conjuntos se introducen de un modo usual:

$$\begin{aligned} \# x \in . A \cap B \cdot \leftrightarrow x \in A . \wedge . x \in B \cdot \\ x \in . A \cup B \cdot \leftrightarrow \cdot x \in A . \vee . x \in B \cdot \\ x \in . A \setminus B \cdot \leftrightarrow \cdot x \in A . \wedge . x \notin B \cdot \\ A \in \text{Agg} . \wedge . B \in \text{Agg} \cdot \rightarrow : A \cap B . \in \text{Agg} \cdot \\ \wedge \cdot A \cup B . \in \text{Agg} \cdot \wedge A \cdot B . \in \text{Agg} \# \end{aligned}$$

Para el conjunto vacío ($\ulcorner \urcorner$) y la noción de número cardinal (Car) se utilizan los mensajes:

$$\begin{aligned} \# x \neq y \cdot \leftrightarrow \cdot \ulcorner x \urcorner \cap \ulcorner y \urcorner \cdot = \ulcorner \urcorner \# \\ \# \wedge x : x \in A . \leftrightarrow \cdot x = a . \vee x = b . \vee . x = c \cdot \\ \wedge \cdot a \neq b . \wedge . a \neq c . \wedge . b \neq c : \\ \rightarrow \cdot \text{Car } A . = 11 : \\ \wedge x : x \in A . \leftrightarrow x = a \cdot \rightarrow \cdot \text{Car } A . = 1 : \\ a \neq b \cdot \rightarrow \wedge x : x \in A \cdot \leftrightarrow . x = a . \vee . x = b \cdot \rightarrow \cdot \text{Car } A . = 10 : \\ \text{Car } \ulcorner x \urcorner \cdot = 1 : \\ \text{Car } \ulcorner \urcorner . = 0 \# \end{aligned}$$

A continuación pueden introducirse las nociones de función, conjunto ordenado, grupo, grupo abeliano.

De la lógica se emplean conceptos tales como “verdadero”, Ver (fL v e r u m); “falso”, Fal (fL f a l s u m); “proposición”, Prp (fL p r o p o s i t i o); “pregunta”, Qus (fL q u a e s t i o); ‘juicio’, Iud (fL i u d i c i u m) con el significado de valor de verdad, en oposición al cual se utiliza también el operador de negación, por medio de \neg .

Se escribirá, por ejemplo:

$$\begin{aligned} & \# 1 + 1 = 10 . \in \text{Ver:} \\ & 1 + 1 = 11 . \in \text{Fal} \# \\ & \# p \in \text{Ver} . \rightarrow . p \in \text{Prp} \cdot \\ & p \in \text{Fal} . \rightarrow . p \in \text{Prp} \# \\ & \# \neg . p \wedge q \leftrightarrow \neg p . \vee . \neg q : \\ & \neg . p \vee q \leftrightarrow \neg p . \wedge . \neg q : \\ & p \rightarrow q \leftrightarrow \neg p . \vee q : \\ & p \leftrightarrow \neg \neg p \# \end{aligned}$$

Pese a que el aparato léxico y sintáctico empleado hasta ahora en Lincos parece bastante amplio, no es todavía suficiente para responder a todas las necesidades (cf. 3.27).

6.2. Nociones temporales

Resulta particularmente importante establecer un método con que expresar nociones relativas al tiempo.

Para alcanzar este propósito se transmiten simultáneamente dos señales diferentes, de modo que algunas tengan un carácter especial con respecto al tiempo: éstas se denominan *señales de tiempo* y vienen representadas por líneas horizontales.

En la escritura, se utilizarán letras góticas para las señales de tiempo o para las señales vinculadas a relaciones temporales. Así, se escribirá:

$$\# \text{Dur} \text{—————} = \text{Sec } a \#$$

En nuestro lenguaje, es decir en metalengua, se leería: “la duración de la señal de tiempo indicada por la línea horizontal es de *a* segundos”. Para ello se usa la notación Dur (fL d u r a t i o, ‘duración’).

Este programa se repetirá varias veces con duraciones distintas a fin de mostrar el paralelismo entre la duración variable de transmisión de las señales temporales y los números *a*.

Otro método, en todo caso semejante, consiste en transmitir mensajes de forma:

$$\# \text{ Nos } \text{-----} = \mathfrak{a} \#$$

donde Nos significa “número de oscilaciones”.

Cabe igualmente utilizar una relación de tipo:

$$\# \text{ Fre } \text{-----} = \text{Sec}^{-1} \mathfrak{a} \#$$

donde Fre (fL f r e q u e n t i a) significa la “frecuencia de las oscilaciones”.

Entonces, una típica sucesión de mensajes destinados a introducir las nociones temporales básicas será como sigue:

$$\# \text{ Dur } x = \text{Sec } a . \rightarrow . a > 0 .$$

$$\text{Fre } x = \text{Sec}^{-1} b . \rightarrow . b > 0 .$$

$$\text{Nos } x = c . \rightarrow . c > 0 :$$

$$\text{Dur } x = \text{Sec } a . \wedge . \text{Fre } x = \text{Sec}^{-1} b . \wedge . \text{Nos } x = c . \rightarrow . ab = c :$$

$$\text{Sec } a \times \text{Sec}^{-1} b = ab .$$

$$c / \text{Sec } a = \text{Sec}^{-1} c = c/a .$$

$$c / \text{Sec}^{-1} b = \text{Sec } c/b \#$$

Tras un programa de este tipo, se pasa a definir las nociones de “comienzo” y de “final”, expresadas por la relación

$$\# \text{ Fin } \text{-----} = \text{Ini} \text{-----} \#$$

Para realizar esto, conviene que la segunda señal de tiempo empiece en el momento en que termine la primera y que ambas señales tengan una longitud de onda distinta. Se emplean los símbolos Fin (fL f i n i s, ‘final’) y Ini (fL i n i t i u m, ‘comienzo’).

Los problemas relativos al tiempo permiten la introducción de muchas otras nociones. Así, p. ej., se transmitirá:

$$\# \text{ Ec } \text{-----} :$$

$$\vee \lceil x . y \rceil . \text{Fre } x = \mathfrak{a} . \wedge . \text{Fre } y = b . \wedge . \text{Ini } y = \text{Fin } x \#$$

que introduce la palabra Ecc (fL e c c e, ‘he aquí’), con la que se anuncia una señal cuya explicación debe hallarse fuera de ella.

Para la noción de “antes”, Ant (fL a n t e), se transmite

$$\# \text{ Ecc } \text{-----}$$

$$\vee \lceil x . y \rceil : \text{Fre } x = \mathfrak{a} . \wedge . \text{Fre } y = b . \wedge . x \text{ Ant } y \#$$

Para la noción de “después”, Pst (fL p o s t), se utiliza el mensaje

$$\# x \text{ Ant } y . \leftrightarrow . y \text{ Pst } x \#$$

La palabra Pre (fL p r e c e d i t, ‘precede’) se refiere a una relación de sucesión inmediata, que se introducirá a base de la relación:

$$\begin{aligned} \# x \text{ Pre } y . \leftrightarrow : \text{Fre } x . &= \mathfrak{m} \cdot \wedge \cdot \text{Fre } y . = \mathfrak{m} \cdot \wedge \cdot x \text{ Ant } y . \\ \wedge . \overline{\quad} \cdot \vee z : \text{Fre } z . &= \mathfrak{m} \cdot \vee \cdot x \text{ Ant } z . \wedge \cdot z \text{ Ant } y \# \end{aligned}$$

donde nos limitamos al caso de las señales temporales de frecuencia ω , en la que ω es una constante del metatexto.

En esta frecuencia ω funcionará un emisor modulado por un reloj que, p. ej., marque los segundos. De esta manera se introduce una noción nueva relativa a la medida de las duraciones. Sobre el funcionamiento del reloj, los receptores obtendrán la información necesaria mediante un determinado programa.

La palabra Mom significa “momento” y alude a un momento dado de los que marca el reloj. Así, p. ej., el momento 0 viene expresado por la fórmula:

$$\begin{aligned} \# \text{Ini } x . = \text{Mom } n : \wedge \cdot x \text{ Pre } y : \rightarrow : \text{Ini } y . &= \text{Mom } (n + 1) \\ \text{Ini}^- &= \text{Mom } 0 \# \end{aligned}$$

A continuación se puede introducir la palabra Tem (fL t e m p u s, ‘tiempo’). Con ayuda del reloj, eventos tales como Ini x y Fin x pueden considerarse como elementos del conjunto de todos los momentos Mom a (con a real, positivo o negativo):

$$\begin{aligned} \# \text{Ini } x \in \text{Tem} . \text{Fin } x \in \text{Tem} \\ a \in \text{Rea} . \rightarrow . \text{Mom } a \in \text{Tem} \# \end{aligned}$$

A partir de ello, cabe introducir asimismo las nociones de pasado y de “hasta”, notada en Lincos por Usq (fL u s q u e), a base del mensaje

$$t_1 \mathfrak{p} \overset{1}{\uparrow} t_1 \text{ Usq } t_2 \text{ Fit } \mathfrak{p} \#$$

donde \mathfrak{p} es una letra del metatexto, p. ej. una perturbación de radio, es decir un grupo de segundos que no pueda ser interpretado como una palabra. El mensaje transmitido significa que de t_1 hasta t_2 ha ocurrido el evento notado. Aparece también la nueva palabra Fit (fL f i t, ‘ocurre’).

De un modo análogo, puede introducirse también la noción de futuro.

6.3. Comportamiento

Se accede luego a la tercera parte del programa, que trata sobre el comportamiento.

Hasta el presente, disponemos de reglas generales relativas a la descripción de los objetos aritméticos y de reglas cronométricas. Pero carecemos aún de nociones referidas al comportamiento, relativamente más difíciles de introducir y, sin embargo, absolutamente necesarias.

Una posibilidad consiste en introducir de antemano ciertas personas que actúen. La única acción inmediatamente expresable en relación con una persona es el acto del habla. En el vocabulario Lincos esta actividad viene indicada por medio de la palabra *Inq* (fL i n q u i t, 'dice').

Se introducen también símbolos que representen las personas que intervienen en la acción. Las personas se notan mediante *Ha*, *Hb*, *Hc*, etc. y pertenecen al conjunto *Hom* (fL h o m o, 'hombre').

Se añaden luego dos valores, *Ben* (fL b e n e, 'bien') y *Mal* (fL m a l e, 'mal').

Un evento elemental se escribe a base de:

$$Ha \text{ Inq } Hb \mu$$

lo cual, en metalengua, se leería: "*Ha* dice a *Hb* μ ". Aquí, μ es un sustituto metatextual para lo que se diga: puede ser una pregunta, una respuesta, una orden, un grupo de palabras sin sentido, etc.

Para ello se acompañan las siguientes precisiones:

$$\# Ha \notin \text{Num. } Ha \notin \text{Com. } Ha \in Ha.$$

$$Hb \notin \text{Num. } Hb \notin \text{Com. } Hb \in Hb$$

$$\text{Etc. } \#$$

He aquí ejemplos textuales de este tipo:

$$\# Ha \text{ Inq } Hb. ? x . 100 x = 1 \text{ 010:}$$

$$Hb \text{ Inq } Ha. 1010/100:$$

$$Ha \text{ Inq } Hb \text{ Mal:}$$

$$Hb \text{ Inq } Ha. 1/10:$$

$$Ha \text{ Inq } Hb \text{ Mal:}$$

$$Hb \text{ Inq } Ha. 101/10:$$

$$Ha \text{ Inq } Hb \text{ Ben } \#$$

No es difícil introducir los pronombres interrogativos mediante preguntas como:

$$\# \hat{1}_1 Ha \text{ Inq } Hb? x. 100x = 1 \text{ 010}^{-12}:$$

$$Hb \text{ Inq } Hc? y: t_1 t_2 Ha \text{ Inq } y \cdot ?x. 100 x = 1 \text{ 010:}$$

$$Hc \text{ Inq } Hb \text{ Hb } \#$$

Se introduce la palabra Utr (fL u t r u m, 'sí' interrogativo) por la regla:

$$\# \uparrow_1 H d \text{ Inq } H d : t t' H b \text{ Inq } H c : ? y y = . \text{Jud Etc'}$$

$$\leftrightarrow ' t t' H b \text{ Inq } H c. \text{ Utr Etc.} \#$$

Para simplificar la escritura, se adopta la convención de notar # ? = Etc. # en lugar de # ? y y = Etc. # , y así sucesivamente (cf. 3.27).

6.4. Espacio, movimiento, masa

El último grupo de nociones presentadas en lenguaje Lincos se refiere al espacio, movimiento y masa.

Recordemos que para introducir las personas que actúan, es decir los hombres —Hom— la única necesidad que hubo que cumplir fue la de tener también a disposición un canal para el tiempo. Pero si bien es relativamente fácil disponer de un canal suplementario para las nociones temporales, ya no lo es tanto para las espaciales.

Se empieza por introducir la palabra Loc (fL l o c u s, 'lugar'). Precisemos que un retraso de las señales representa una diferencia de lugar.

La palabra Spa (fL s p a t i u m, 'espacio') representa a su vez el conjunto de todos los puntos espaciales. Todos los lugares se hallan en el espacio. La noción Dst (fL d i s t a n t i a, 'distancia') se introduce con relativa facilidad aprovechando que es proporcional al tiempo del retraso. En rigor, se trata de un método provisional, ya que utiliza el tiempo en tanto que no aparezcan otras posibilidades más adecuadas a la señalización.

Para la medición de las distancias se recurre a la unidad centímetro, Cmt. A partir de ella se dan los axiomas sobre el espacio métrico, se define la línea recta y los axiomas del espacio euclidiano, se ofrecen ciertas propiedades geométricas de este espacio y se definen los conceptos de vector, espacio vectorial y volumen de un cuerpo.

Con ayuda de estos conceptos espaciales y temporales pueden comunicarse igualmente otros conceptos físicos.

Se empieza por la teoría de las oscilaciones, definiendo nociones como amplitud (Aml), frecuencia (Fre), fase (Pha) y punto de equilibrio de una oscilación (EqI).

Nos preguntamos cómo pueden las personas recibir mensajes emitidos en otro lugar. Hay tres modalidades. En primer lugar, por transporte (convección) del mensaje, pero es un procedimiento lento. Un método más rápido consiste en utilizar las ondas que se propagan por el Aire (Aer), caracterizado por la velocidad de propagación del sonido. La tercera posibilidad se basa en las ondas (Und) que se propagan a la velocidad de la luz, c.

Disponiendo de una unidad de longitud, puede transmitirse el espectro del hidrógeno y la constante de Rydberg.

A continuación, se pasa a la noción de masa y a la de velocidad (C_{el}). Por medio del choque elástico entre dos cuerpos se introduce la noción de ley natural (Lex Nat). Después se llega al campo gravitacional, al movimiento kepleriano, a la cosmología.

Una vez expuestos otros capítulos de la física y de la técnica, p. ej., la teoría de los gases y el radar, el curso Lincos termina exponiendo la teoría de la relatividad restringida.

El Lincos constituye un intento particularmente interesante de introducir las bases del lenguaje de un modo estrictamente lógico, recurriendo sólo a las reducidas posibilidades que brindan dos canales de comunicación por los cuales sólo pueden enviarse ondas de radio moduladas en forma de señales simples.

Evidentemente, aunque por el momento las aplicaciones prácticas del lenguaje Lincos no parecen aprovechables en un futuro inmediato, merece, en cambio, la mayor atención el intento, realizado por el Prof. H. Freundenthal, de introducir, de una manera lógica y con un mínimo de recursos, tanto la riqueza del lenguaje humano, con su vocabulario y sus reglas gramaticales, como los principales resultados obtenidos en el dominio científico.

Así, como se ha dicho también al principio del capítulo, el problema esencial que se plantea en el lenguaje Lincos, y en todos los lenguajes similares, consiste en la interpretación unívoca de los mensajes transmitidos.

Las investigaciones que se efectúan en esta dirección se hallan todavía al principio, pero es de esperar que muy pronto se obtengan también otros resultados fuera del sistema Lincos.

Pese a las críticas que se le pueden hacer, el Lincos permanece como una obra importante en el terreno de la formalización total del lenguaje humano.

7. Traducción y composición automáticas

7.1. Nociones introductorias

Entre los logros más espectaculares que la cibernética ha conseguido en el terreno de la lingüística destaca, sin duda, el de la traducción automática.

Bien es verdad que la posibilidad de realizar traducciones con ayuda de máquinas fue vislumbrada hace ya mucho. En 1933, el investigador soviético P. Smirnov-Troianski obtuvo en Moscú una patente para una máquina que permitía efectuar una traducción simultánea y a distancia en varias lenguas. Pero la verdadera época de las traducciones automáticas no empezó hasta el perfeccionamiento de los ordenadores electrónicos. En el año 1957, en Nueva York, la firma I. B. M. hizo la primera demostración pública de traducción automática de una lengua a otra, en aquel caso del ruso al inglés. Poco tiempo después, en la Unión Soviética se llevó a cabo una traducción automática en sentido inverso.

En Rumanía, las primeras traducciones automáticas tuvieron lugar en abril de 1962, en Timişoara, bajo la dirección del académico Gr. C. Moisil. La experiencia se realizó entre el inglés y el rumano a partir de un algoritmo elaborado por E. Domonkoş¹.

7.2. El tratamiento de los datos

Para comprender la capacidad de los calculadores electrónicos en efectuar traducciones automáticas conviene que recordemos de nuevo que estas máquinas no funcionan con números o con letras, sino con señales eléctricas. Y en tanto que estas señales se ponen en correspondencia con números, símbolos lógicos o letras, el aparato puede considerarse como una máquina aritmética, lógica o de

1. Cf. DOMONKOŞ. *Algoritmul de traducere*.

traducción, respectivamente. A fin de simplificar la exposición, las señales eléctricas de los ordenadores se representan a menudo mediante cifras, convención que nosotros también adoptaremos para las letras del alfabeto. Entonces, escribiremos:

$$a = 1, \quad b = 2, \quad c = 3, \dots$$

o bien en un sistema binario

$$a = 00001, \quad b = 00010, \quad c = 00011, \text{ etc.}$$

En el segundo sistema, y para ganar en uniformidad, cada letra va indicada por el mismo número de cifras, lo que permite separarlas sin equívoco. De este modo, en lugar de escribir *ac*, se escribirá

0000100011.

Gracias a este sistema, todo texto queda traducido como una sucesión de símbolos binarios, expresables luego por una sucesión de impulsos eléctricos adaptados al tratamiento del ordenador digital (cf. 5.1). La segunda etapa para la comprensión del proceso de traducción automática atañe a la posibilidad de automatizar el acto mismo de traducir². En este acto se parte de un determinado texto, expresado en una lengua denominada *l e n g u a - o r i g i n a l*, y en el que aparecen series de letras agrupadas en forma de palabras flexionadas según las reglas gramaticales que se hallan en vigor en el ámbito de la lengua respectiva.

Para efectuar la traducción, es necesario establecer, teniendo en cuenta estas reglas gramaticales, cuáles son las palabras no-flexionadas que intervienen en el texto, así como el tipo de flexión sufrida por cada una de las demás palabras. De esta suerte, p. ej., si en el texto aparece la palabra *copiilor*, ‘de/a los niños’, hay que transcribirla del siguiente modo: “copil (‘niño’), plural, genitivo o dativo”; si aparece la palabra *alergaseși*, ‘habías corrido’, se escribirá: “a alerga (‘correr’), pluscuamperfecto, indicativo, singular, segunda persona”, y así sucesivamente.

Con ayuda de este procedimiento, toda frase queda expresada por sucesiones de palabras que pueden buscarse en un diccionario bilingüe —de la lengua original a la *l e n g u a - i m a g e n*, es decir a la lengua a que se traduce— y acompañadas de las indicaciones gramaticales correspondientes.

Para traducir el texto, una vez buscada en el diccionario la respectiva sucesión de palabras, se obtiene otro donde en el lugar de las palabras de la lengua original aparecen las de la lengua-imagen. Utilizando entonces la gramática de esta última lengua y las indicaciones gramaticales pertinentes, se flexionan las

2. Cf. BOOTH, “Practical realization”; BOOTH-BRANWOOD-CLEAVE, *Mechanical Resolution*; NICOLAU, *Utilizare*; PANOV, *Avtomaticheskij perevod*; YNGVE, “Programming language”; ZARECHNAK, *Faster than Thought*; [y GOUGHENHEIM *et al.*, *Problèmes*].

palabras y se compone una traducción en bruto en la lengua-imagen. Evidentemente, la traducción así obtenida resulta imperfecta desde el momento en que se ha omitido la posibilidad de un orden de palabras distinto, así como la eventual existencia de homonimias o de modismos, que pueden aparecer en los textos literarios. En cuanto a los textos científicos, sin embargo, no se plantean por lo general estos problemas, puesto que la redacción suele evitar sistemáticamente toda ambigüedad interpretativa.

A la luz de los conocimientos de que se dispone sobre los ordenadores digitales, es fácil comprobar que todas estas operaciones son susceptibles de un tratamiento asequible (cf. 5.2). Efectivamente, en la memoria del ordenador pueden introducirse señales eléctricas que, con arreglo al código cífico mencionado, compendien el diccionario bilingüe que se requiere en la traducción. De un modo similar, también se introducen en la memoria las reglas gramaticales tanto de la lengua-original como de la lengua-imagen. Desde luego, no hay en ello la menor dificultad. Puede convenirse, p. ej., en que el nominativo quede expresado a base de 01; el genitivo, de 10, y así sucesivamente. Igualmente pueden aplicarse indicaciones cíficas para el número, el modo, el tiempo, la persona, etc.

Vale la pena subrayar el hecho de que la cibernética, a través de la traducción automática, ha dado un nuevo ímpetu a los estudios de gramática y ha demostrado, entre otras muchas cosas, que en ninguna lengua existe una gramática completa, capaz de comprender todas las situaciones posibles, sin excepción.

En Rumania, las primeras investigaciones en este sentido se deben al académico Gr. C. Moisis, el cual ha realizado una serie de importantes contribuciones a la microgramática del verbo, del nombre y del adjetivo³.

A este propósito, hay que notar el hecho de que, para muchos especialistas en cibernética, la gramática es una ciencia de observación. El material bruto del que se parte está formado por la totalidad de los textos existentes en el uso corriente, en un determinado momento y en una lengua cualquiera, sobre los cuales el gramático determina, por una parte, el fondo léxico y, por otra, las reglas morfológicas, sintácticas, etcétera. Salvando las lenguas artificiales, tipo esperanto, para ninguna de las lenguas naturales existentes hoy podemos disponer de una gramática acabada.

Contra lo que ocurre en las teorías axiomáticas, donde, al margen de una lengua-objeto, hay posibilidad de hacer comentarios en una metalengua, en las lenguas naturales no es fácil distinguir dos categorías de lengua, una lengua-objeto y una metalengua, entre las cuales quepa establecer una preeminencia de una sobre la otra.

Pero, aunque no hay ninguna gramática completa para las lenguas naturales, sí en cambio las hay suficientemente cumplidas como para permitir la tra-

3. Cf. MOISIL, "Problèmes" y "Probleme".

ducción automática, al menos, de textos científicos. Éste es el caso del rumano y del inglés en las primeras experiencias aludidas de traducción automática.

7.3. El algoritmo de traducción automática del inglés al rumano

En lo que sigue, expondremos los principales rasgos del algoritmo utilizado en la traducción del inglés al rumano por E. Domonkos. En la figura 23 se representa el organigrama para esta categoría de traducción. En principio, el organigrama es válido para cualquier traducción de una lengua flexiva a otra lengua flexiva o, más exactamente, de una lengua cuyas palabras puedan presentar modificaciones desinenciales a otra lengua con declinación, conjugación y concordancia.

Ante todo, el texto a traducir se perfora sobre una banda de acuerdo con un código adecuado al alfabeto de la lengua inglesa. En una segunda etapa, se introduce cada palabra por separado en el ordenador y se busca en el diccionario. Existe, entonces, la posibilidad de que se encuentre o de que no se encuentre. En el primer caso, se registra la palabra en la memoria del ordenador, en la llamada matriz de las reservas de investigación⁴. Si se considera todo el proceso de traducción desde un punto de vista abstracto, se comprueba que, en conjunto, esta operación consta de varias etapas. En primer lugar, hay un proceso de transformación del texto de la lengua-objeto a una sucesión de símbolos numéricos que comprenden diversas categorías de información:

- a) la localización de las palabras en el glosario inglés, que fija una parte del significado respectivo e informa sobre qué palabra del diccionario se trata;
- b) la determinación de los índices gramaticales de la palabra inglesa, que permite su asignación a una determinada clase de palabras;
- c) la determinación del caso que tiene la palabra respectiva en la proposición que se traduce.

Evidentemente, las traducciones a partir del inglés se benefician de la ventaja de que en esta lengua la flexión de las palabras apenas existe.

Desde luego, cuando las palabras del texto se encuentran bajo la misma forma en el diccionario, es que estamos en presencia de un nombre en nominativo, un verbo en infinitivo, etcétera. Pero las palabras no se encuentran siempre así. Sin embargo, basta efectuar una serie de operaciones preliminares para que las palabras de una proposición se hallen registradas en la memoria de la máquina, concretamente en un grupo de direcciones que constituyen la matriz de las reservas de investigación. En ella, cada palabra ocupa dos reservas: al principio de la primera reserva se registra la dirección donde la palabra ha sido

4. Es decir, en la matriz de las expresiones analizadas.

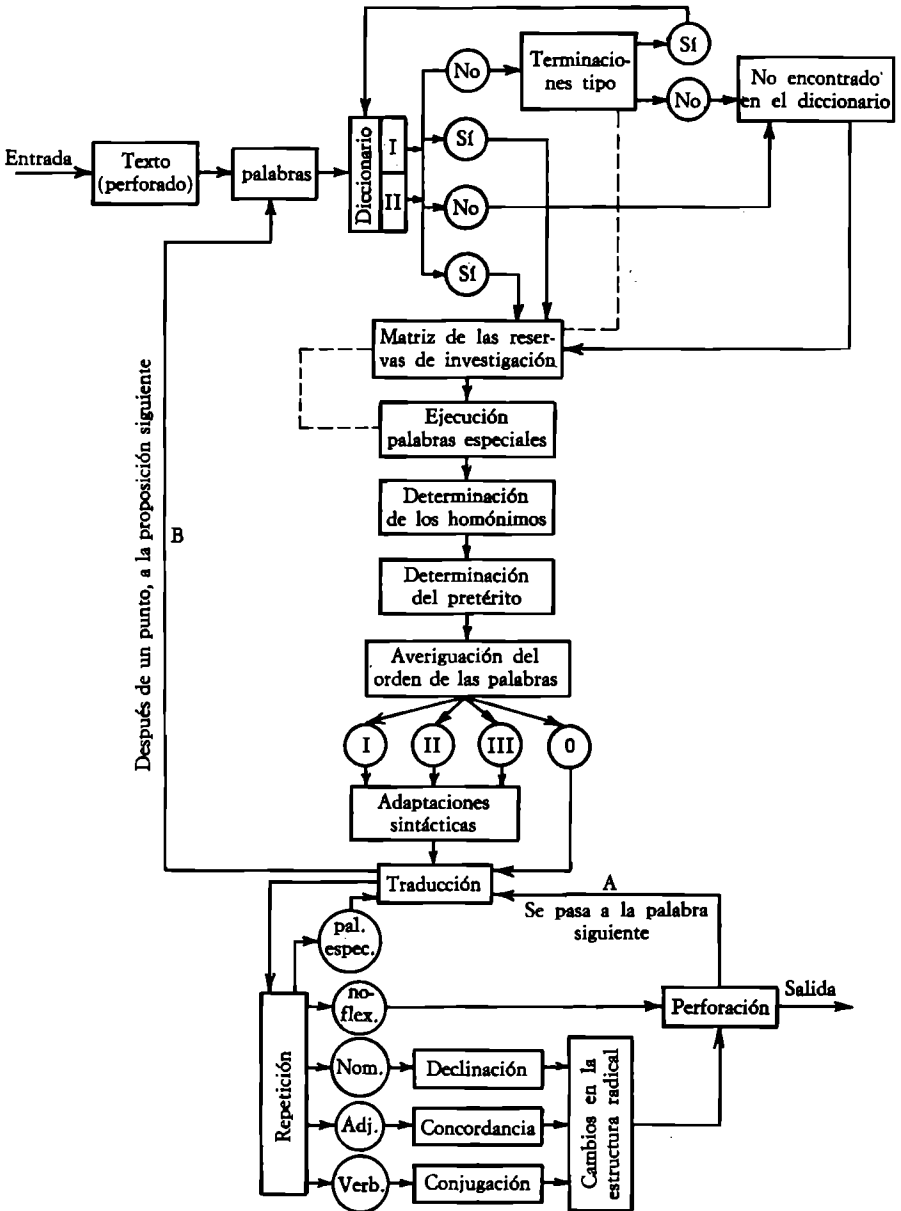


Figura 23

localizada en el glosario; al final de la misma dirección se consignan los índices gramaticales de la palabra en cuestión, a fin de que en la segunda reserva se centralicen las informaciones gramaticales relativas a la forma de la palabra.

Con estas indicaciones generales, podemos volver al organigrama de la traducción.

Si una palabra no se encuentra en el diccionario, entonces se sigue un programa de eliminación de las terminaciones-tipo, el cual verifica las principales desinencias gramaticales del inglés: *ed, s, 's, s', ing, ly, er, est*, cuyo significado detallamos en el cuadro 10.

Terminación	Significado gramatical
<i>ed</i>	pretérito
<i>s</i>	plural en los nombres; 3. ^a persona del singular en los verbos
' <i>s</i>	genitivo sajón singular
<i>s'</i>	genitivo sajón plural
<i>ing</i>	gerundio
<i>ly</i>	adverbio
<i>er</i>	comparativo
<i>est</i>	superlativo

Cuadro 10

Si la palabra no se encuentra en el diccionario ni termina con ninguna de las desinencias que acabamos de examinar, se considera que no figura en el diccionario y se retiene como tal en su forma primitiva para que reaparezca así en el texto traducido. Pero si la palabra presenta una de las terminaciones anteriores, entonces se separa la terminación y se busca de nuevo en el diccionario la porción restante. Al mismo tiempo, en la segunda reserva destinada a esta palabra dentro de la matriz de reservas de investigación se inscribe una información relativa al significado gramatical de la desinencia separada. En caso de que una terminación ofrezca una cierta ambigüedad por corresponder a dos o más alternativas posibles, se consignan todas las informaciones en la matriz de reservas de investigación con el objeto de que la máquina seleccione ulteriormente el significado que corresponda sobre la base de otros índices. Si la palabra no se encuentra en el diccionario, pese a tener una terminación-tipo, se considera inexistente en el diccionario, pero si se encuentra efectivamente, entonces se procede como en la primera búsqueda.

Al final de esta etapa aparecen cadenas de símbolos abstractos que representan la proposición que va a traducirse. La segunda porción de las palabras no comprende en esta etapa más que las raíces gramaticales obtenidas por eli-

minación de unas terminaciones-tipo, pero en el caso del inglés, algunas informaciones no se hallan contenidas en el cuerpo mismo de las palabras, sino en palabras distintas, normalmente situadas antes de la palabra respectiva. Se trata de las llamadas "palabras especiales", que no se traducen al rumano [ni, en algunos casos, a ninguna lengua románica]. Este programa se refiere a las palabras del cuadro 11, donde se especifica su contenido gramatical.

Hay que destacar el hecho de que, a veces, la misma información puede obtenerse de varias maneras. Así, p. ej., el comparativo se manifiesta por la terminación *-er* o bien por la palabra especial *more*. Una vez ejecutados todos estos programas especiales, se entiende que ha quedado recogida toda la información gramatical pertinente.

La siguiente etapa del programa se refiere a la identificación de los homónimo a fin de que quepa establecer si se trata de una de las categorías gramaticales de nombre, verbo o adjetivo.

Antes de pasar al texto rumano, debe cambiarse el orden de las palabras conforme a las leyes de la sintaxis rumana. Esto se efectúa a partir de dos programas.

Palabra especial	Contenido gramatical
<i>of</i>	genitivo
<i>the</i>	artículo
<i>shall</i>	futuro
<i>will</i>	futuro
<i>should</i>	condicional
<i>would</i>	condicional
<i>more</i>	comparativo
<i>most</i>	superlativo
<i>have</i>	perfecto compuesto
<i>had</i>	pluscuamperfecto
<i>do</i>	—
<i>does</i>	3.ª persona del singular
<i>did</i>	perfecto

Cuadro 11

En primer lugar, se averigua el orden de las palabras de forma que la máquina reconozca la presencia de construcciones donde sea necesario efectuar cambios. Un segundo programa realiza el cambio mismo, en el que las palabras modifican su lugar en la matriz de las reservas de investigación con arreglo a los principios de la sintaxis rumana. Con ello, el proceso de traducción queda virtualmente terminado.

Veamos ahora algunos detalles relativos al programa por el que cada palabra inglesa se reemplaza por su equivalente rumano según el contenido del diccionario bilingüe.

La primera etapa —convencionalmente denominada de “repartición”— comprende la adscripción de la palabra a una determinada clase. Posteriormente, se ponen en marcha unos subprogramas en correlación con las clases de palabras establecidas anteriormente, según consta en el cuadro 12.

Categoría de la palabra	Subprograma
Nombre	Declinación
Verbo	Conjugación
Adjetivo	Concordancia
Palabras no-flexivas y conjunciones copulativas	Imprimir sin modificación
Palabra-especial	No se traduce y se pasa a la palabra siguiente

Cuadro 12

El subprograma de declinación identifica los siguientes elementos: caso, número y artículo. En función de ello, el programa promueve la adición de la desinencia de rigor a la palabra rumana teniendo en cuenta para ello la asignación del nombre a los tipos comprendidos en el modelo de esta lengua.

El subprograma de conjugación se refiere al modo, tiempo y persona, para que la palabra de la lengua-imagen flexione en virtud de estos factores.

El subprograma de concordancia establece los elementos que aluden al grado de comparación. A partir del nombre con artículo determinado, se especifica el número y se analizan las indicaciones léxicas del mismo, junto con el género. Se verifica si el adjetivo presenta terminación en *-e* o si hay homonimia morfológica en el paradigma del adjetivo⁵. En una nueva etapa, es necesario cambiar la estructura de la raíz, para lo cual se prevén tres instrucciones:

a) se elide una determinada letra, dada por los índices del diccionario (por ejemplo, *dreaptă* — *drepte*, ‘recta / rectas’);

b) se añade una determinada letra, igualmente dada por los índices del diccionario (p. ej., *frumos* — *frumoasă*, ‘hermoso / hermosa’);

5. [En rumano, los adjetivos en *-e* suelen tener sólo dos desinencias morfológicas de distribución irregular, con tres paradigmas diferentes, por lo menos].

c) se modifica la última letra de la raíz (p. ej., *frumos* — *frumoși*, ‘hermoso / hermosos’, *drept* — *drepti*, ‘recto / rectos’, *ied* — *iezi*, ‘chivo / chivos’)⁶.

Para ejecutar la instrucción c), en el alfabeto cífico del rumano, el orden es distinto del habitual, puesto que aparece la sucesión ...d, z..., mientras que en los demás casos se mantiene la sucesión normal (...s, ș..., d, ț..., z). El programa previene, pues, en c), el cambio de una unidad en la última letra de la palabra.

A veces, las palabras encontradas se envían a estos subprogramas, es decir: los nombres femeninos terminados en -ă y en plural siguen el programa a); los adjetivos no terminados en -e y en femenino singular pasan al programa b); los mismos nombres masculinos en plural se envían al subprograma c).

La máquina verifica si los índices del diccionario prevén este tipo de cambios en la estructura radical. Evidentemente, en caso de que no se dé esta indicación, no se efectúa ninguno de los subprogramas respectivos. Desde luego, aquellas tres instrucciones no agotan todas las posibilidades. Precisamente por ello en la bibliografía especializada se indican también los métodos exhaustivos⁷.

La última etapa del programa consiste en ordenar la impresión del texto a la máquina eléctrica de escribir. En cada paso se imprime la traducción de una sola palabra del texto original en inglés. A este propósito, conviene subrayar que una palabra de este tipo puede expresarse en rumano a base de un número de palabras comprendido entre 0 y 3. En ciertos casos, la palabra carece de traducción —se trata de las palabras especiales—. En otros, para una palabra inglesa hay un equivalente en rumano. Pero existe la posibilidad también de expresar una palabra inglesa mediante dos o más palabras rumanas; v. gr.: *better* = *mai bine* (‘mejor’) y *the best* = *cel mai bine* (‘el mejor’). Tales construcciones reciben un tratamiento especial.

Para las eventuales palabras auxiliares se prevén dos reservas y para las normales, tres. A veces (en el perfecto compuesto, adjetivo comparativo, nombre sin artículo y en caso genitivo, verbos en futuro, subjuntivo, optativo), basta una sola palabra auxiliar. Para los adjetivos superlativos, los verbos en futuro anterior, el subjuntivo pasado o el optativo pasado, se necesitan dos palabras auxiliares. El correspondiente subprograma contiene las siguientes instrucciones:

a) si en la primera reserva hay algo registrado, se imprime la indicación y luego se acciona sobre la tecla de “espacio” en la máquina de escribir;

6. [En todos los casos, se trata de reglas morfológicas típicas del rumano. Salvo algunas excepciones debidas a préstamo o cultismo, por lo general (v. gr. *coastă*, ‘costa’ / *costă*, ‘cuesta’), el fonema /o/ diptonga en /oa/ en posición tónica y cuando en la sílaba subsiguiente se encuentra /ă/, /a/ o bien /e/; el fonema /a/ hace lo propio en /eă/ en posición tónica y con /ă/, /a/ en la sílaba siguiente. Y en consonancia a c), se refiere a la oposición igualmente fonológica entre palatalización / no-palatalización del consonantismo rumano en posición final (cf. 8.15.)].

7. Cf. SLUTSKER, “Poluchenie”.

b) si en la primera reserva no hay nada registrado, se pasa a la segunda reserva sin accionar la tecla de "espacio";

c) si en la segunda reserva hay algo registrado, se imprime del modo correspondiente y se acciona la tecla de "espacio";

d) si en la segunda reserva no hay nada registrado, se pasa a la palabra propiamente dicha sin espacio;

e) se imprime la palabra a partir de sus tres reservas sin espacios intermedios. Una vez impresa toda la palabra, se acciona la tecla "espacio".

La operación de imprimir se realiza primero en las letras rumanas en el mismo orden en que aparecen dentro de la palabra respectiva. Para ello, un subprograma especial controla el número máximo de letras que caben en una línea del texto que se escribe a máquina y ordena el desplazamiento automático del carro. El mismo subprograma controla el paro del ordenador al final del texto, así como la impresión del punto tras una proposición, etc.

El problema de la traducción automática guarda estricta relación con la capacidad de la memoria del ordenador con que se opera. Evidentemente, para realizar la traducción en buenas condiciones, es necesario disponer de una memoria completa, que permita el registro de un vocabulario lo más abundante posible y una cantidad cuanto más grande mejor de reglas gramaticales. En la experiencia descrita se empleó el ordenador MECIPT-1.

La introducción de una sola palabra en el diccionario requiere un número de direcciones calculable en función de la longitud máxima de las palabras. En aquel experimento, se admitió que la palabra inglesa tiene una longitud máxima de 12 letras. Y como cada letra exige cinco posiciones binarias, resulta que cada palabra ocupaba $12 \times 5 = 60$ bits (cf. 8.8). Por lo tanto, en la memoria del ordenador cada palabra inglesa disponía de dos direcciones de 30 cifras binarias cada una, además de la cifra-signo (*sign bit*). Para la palabra rumana se admitió una longitud máxima de 18 letras, con lo que ocupaba tres direcciones. A las palabras de ambas lenguas se les añadió todavía una dirección de índices gramaticales.

Teniendo en cuenta la capacidad memorística de la máquina, cabe deducir el número máximo de palabras que puede comprender el diccionario bilingüe. En nuestro caso, se utilizó un vocabulario de 80 palabras tan sólo. Debido a esta constricción, la selección de las palabras a partir del diccionario se hizo considerando diversos criterios: se eligió un determinado modelo de lengua, correspondiente a textos científicos; el diccionario comprendía todos los tipos de nombres y verbos de aquel modelo lingüístico; al mismo tiempo, tenía en cuenta algunas irregularidades frecuentes del inglés; se previó, en fin, la disposición recíproca de las palabras en un contexto.

Las palabras se buscan en el diccionario a base de sustracciones sucesivas. Como cada palabra inglesa comprende dos direcciones, la primera sustracción se hace entre la segunda célula o dirección de la palabra buscada y la segunda célula de las palabras del diccionario. En el momento en que el resultado de la sustracción equivale a 0, se sustraen las primeras direcciones. La palabra se considera localizada en el diccionario sólo cuando coinciden ambas direcciones.

El orden de las palabras en el diccionario es indiferente desde el punto de vista teórico, pero cuando se intenta traducir una determinada categoría de textos, vale la pena estudiar una disposición óptima. Esto, claro, en el caso de un diccionario de gran capacidad.

En el diccionario, toda palabra va acompañada de una célula de índices. En el cuadro 13 se indica el rango de las distintas señales binarias de una dirección junto a su significación respectiva.

Rango de la cifra binaria	Significación
30-28	Adscripción a una clase de palabras
27	Predeterminante para <i>of</i> para traducir por <i>din</i> ('de/desde')
26	Predeterminante para <i>of</i> para traducir por <i>de</i> ('de')
25	Postdeterminante para <i>of</i> para traducir por <i>din</i>
24	Postdeterminante para <i>of</i> para traducir por <i>de</i>
23-14	Subprograma de las palabras especiales

Cuadro 13

La adscripción a una clase de palabras se hace a base de tres bits, lo que permite un máximo de $2^3 = 8$ clases de palabras (cuadro 14).

Símbolos binarios	Clase de palabras
000	Nombre
001	Adjetivo
010	Verbo
011	Palabra especial
100	Palabra no flexiva
101	Conjunción copulativa
111	Homónimo

Cuadro 14

Rango de la cifra binaria	Significado
30-26	Parte invariable en una raíz con estructura variable
25	Índices de la célula variable (0 = segunda célula, 1 = tercera célula)
18	Se aumenta la última letra en una unidad
17	Tras la letra indicada por el grupo 30-26, se intercala <i>a</i>
16	Pierde la letra que sigue al grupo 30-26
10-1	Asignación a subcategorías dentro de la clase de palabras

Cuadro 15. Presenta los índices de la palabra rumana

En el caso de las palabras flexivas, las informaciones gramaticales se obtienen bien por la eliminación de las terminaciones-tipo, bien ejecutando el programa de las palabras especiales. Las informaciones pertinentes se registran en la segunda dirección de la palabra respectiva a partir de la cadena de reservas de investigación. La estructura de esta célula o dirección se indica en el cuadro 16.

Rango de la cifra binaria	Significado
30-28	Formas nominales
27-26	Formas adjetivas
25-20	Tiempo, modo
19-12	Persona
16	Pretérito en función adjetiva
15	Genitivo sajón

Cuadro 16

En el cuadro 17 aparece la especificación de las formas nominales.

Nótese que, en el rango 28, 0 indica nominativo y 1, genitivo. En el rango 29, 0 significa singular y 1, plural. Y en el rango 30, 0 representa ausencia de artículo y 1, presencia de artículo.

Para la especificación de las formas adjetivas, cf. el cuadro 18.

Para las formas verbales, se procede en dos etapas. El tiempo y el modo quedan indicados por los rangos 25-20 (cf. cuadro 19), y la persona, por los rangos binarios 17, 18 y 19 (cf. cuadro 20).

Rango de las cifras binarias			Significado
28	29	30	
0	0	0	Nominativo singular sin artículo
0	0	1	Nominativo singular con artículo
0	1	0	Nominativo plural sin artículo
0	1	1	Nominativo plural con artículo
1	0	0	Genitivo singular sin artículo
1	0	1	Genitivo singular con artículo
1	1	0	Genitivo plural sin artículo
1	1	1	Genitivo plural con artículo

Cuadro 17

Rango de las cifras binarias			Significado
26	27		
0	0		Positivo
0	1		Comparativo
1	0		Superlativo

Cuadro 18

La información relativa a la persona viene dada por factores tales como el pronombre personal y la orden de eliminar la desinencia *s* o la palabra *does* para la tercera persona del singular.

Los problemas sintácticos que plantea la traducción automática del inglés al rumano determinaron la elaboración de un subprograma para el “establecimiento del orden de las palabras”, en el cual se reconocen cuatro casos:

- a) grupo de dos palabras, cuyo orden se invierte;
- b) grupo de tres palabras, cuyo orden se invierte;
- c) grupo de cuatro palabras, cuyo orden se invierte;
- d) otros casos.

Para el primer caso (grupo de dos palabras) se compuso un subprograma con objeto de prever el grupo [nombre + nombre], en el que el primero es un

genitivo sajón y el segundo representa la cosa poseída; v. gr. *father's book* = *cartea tatălui* ('el libro del padre'). Igualmente se ha estudiado el grupo [adjetivo + nombre]; v. gr. *good book* = *carte bună* ('libro bueno').

Para la segunda categoría (grupo de tres palabras), se consideraron grupos de forma [adjetivo₁ + adjetivo₂ + nombre]; v. gr. *good Rumanian book* = *carte românească bună* ('libro rumano bueno').

Para el tercer caso, se tuvieron en cuenta construcciones a base de [adjetivo₁ + coma + adjetivo₂ + nombre] o bien [adjetivo₁ + conjunción copulativa + adjetivo₂ + nombre]; v. gr. *good and interesting book* = *carte bună și interesantă* ('libro bueno e interesante').

Rango de la cifra binaria	Recogida de la información	Forma verbal
25	Eliminación de la desinencia <i>-ed</i>	Perfecto compuesto
24	Eliminación de la desinencia <i>-ing</i>	Gerundio
23	Verbo precedido de <i>to</i>	Infinitivo precedido de <i>a</i> ⁸
22	<i>shall will should would</i>	Tras una palabra auxiliar de futuro o de optativo, poner verbo en infinitivo
21	<i>shall will should would</i>	Subjuntivo, optativo
20	Participio pasado precedido de <i>had</i>	Pluscuamperfecto

Cuadro 19

Rango de las cifras binarias			Persona
17	18	19	
0	0	0	3. ^a plural
0	0	1	2. ^a plural
0	1	0	3. ^a singular
0	1	1	1. ^a plural
1	0	0	1. ^a singular

Cuadro 20

8. [En rumano, el infinitivo va precedido de *a*: *a lucra*, 'trabajar'; *a hotărî*, 'decidir', etcétera. Por lo demás, sólo en una fase muy provisional de traducción cabe establecer la presente equivalencia con el inglés *to* + verbo].

El orden de las palabras se reconoce del siguiente modo. Se forma un número de 12 cifras binarias, obtenido de la adscripción a la clase de palabras de cuatro palabras conocidas (así, pues, a cada palabra le corresponden tres cifras). De este número se resta la estructura de la construcción *c*), ya conocida. Si el resultado equivale a 0, la construcción *c*) queda ya identificada. Si no, se eliminan las últimas tres cifras del número respectivo, del cual se resta a continuación la estructura de la construcción *b*), y así sucesivamente hasta obtener el resultado 0.

Mediante el programa de reordenación sintáctica cambia el lugar de las palabras en la cadena de las reservas de investigación. En efecto, si se notan a base de 1, 2, 3, 4, 5, 6, 7 y 8 las direcciones consecutivas que corresponden al caso *c*) y se tiene en cuenta el hecho de que una palabra contiene dos células, entonces, a partir de la modificación del orden y en conformidad con las reglas del rumano, se llega a la formación 7, 8, 5, 6, 3, 4, 1, 2.

El programa, en rigor, era más amplio, en el sentido de que permitía, por lo general, la reordenación de cantidades incluso mayores de palabras.

Para la homonimia se elaboró asimismo un programa que cubría algunos de los casos más frecuentes en el modelo lingüístico adoptado. A grandes rasgos, cabe distinguir tres categorías de homónimos: las preposicionales, las gramaticales y las determinables por métodos gramaticales. En algunos casos de homonimia en las preposiciones *of* y *to*, el algoritmo analiza las palabras adyacentes recurriendo, pues, al procedimiento sintagmático.

A su vez, para determinar la función concreta del pretérito inglés se verifica un análisis especial. Si funciona como verbo, se traduce por el perfecto compuesto, pero si funciona como adjetivo, se traduce por el participio pasado (al que se le asigna género y número en concordancia con el nombre determinado) y se trata luego como adjetivo en el momento de considerar el orden de las palabras. El subprograma comprende, así, las siguientes especificaciones:

a) si va precedido de un pronombre personal, es un verbo y se traduce por el perfecto compuesto;

b) precedido de *had*, funciona como verbo y se traduce por el pluscuamperfecto;

c) precedido de un artículo o seguido de un nombre funciona como adjetivo.

El diccionario contiene también ciertos homónimos identificables por métodos gramaticales, tales como los de nombre-verbo (v. gr. *work*), de adjetivo-verbo (v. gr. *complete*) o de nombre-adjetivo (v. gr. *electron*).

A su vez, el programa mismo determina los casos en que estas homonimias son identificables por medios gramaticales. Así, p. ej., si una palabra lleva una indicación de nombre, resulta que en los rangos 30-28 de la célula de las informaciones gramaticales aparecerá un número distinto de 0. De ello se deduce que estamos en presencia de un nombre.

El programa de las palabras especiales es más extenso. Se distinguen, en primer lugar, dos categorías de palabras especiales: las llamadas “palabras provisionalmente especiales” y las “palabras especiales propiamente dichas”. En el grupo de las primeras entran los pronombres personales y las formas irregulares de las palabras inglesas. A éstas se les asignan índices de palabra especial ya en el diccionario, de modo que puedan pasarse al subprograma individual. Tras ejecutar éste, se cambian los índices a “no flexivo en el pronombre”, o en el “verbo”, “nombre”, “adjetivo”, o bien en las formas irregulares (*better, best, begun, began, wrote, written*, etc.). En la segunda categoría figuran palabras como *of, the, shall, will, should, would, do, does, did, more, most* (cf. cuadro 11), las cuales conservan los índices inicialmente, pero desaparecen en el texto traducido.

Las palabras provisionalmente especiales son tratadas por los programas generales, mientras que las palabras especiales propiamente dichas lo son por programas individuales. En lo que sigue examinaremos algunos de estos programas específicos.

Para los pronombres personales se busca el verbo que va a continuación y se obtiene con ello la información relativa a la persona.

Para la palabra *of* se confeccionó una estadística sobre 500 expresiones que la contenían en textos matemáticos. Se llegó a la conclusión de que la palabra *of* puede traducirse al rumano de tres maneras diferentes: por *de* (‘de’), *din* (‘de / desde’) o genitivo (cf. cuadro 13). Se elaboraron asimismo criterios lógicos a fin de distinguir estas tres alternativas. Entonces, p. ej., la palabra *of* se traduce por *de* antes o después de ciertos nombres, pero sólo en el caso de que éstos vayan sin artículo y no vayan precedidos de un adjetivo cuantitativo o de un numeral ordinal, o no vayan seguidos de un numeral cardinal, o bien cuando *of* no preceda a un nombre propio, una fórmula matemática o la notación de una letra. He aquí algunos ejemplos de predeterminantes:

nombres: *collection of, solution of, number of*;
 adjetivos: *inside of, independent of*;
 verbos: *to dispose of, to make use of*.

Y como ejemplos de postdeterminantes: *order, character, class, type*.

La palabra *of* se traduce por *din* (‘de / desde’) o por *dintre* (‘de [entre]’) tras pronombres cuantitativos (*some, each, most, and, all*), numerales ordinales y cardinales y aun tras pronombres ordinales (*last*).

La palabra *of* se traduce por el genitivo cuando precede a un nombre con artículo o cuando el nombre va precedido de un adjetivo demostrativo (*this, that, these, those*) y también cuando precede a una fórmula matemática, la notación de una letra o un nombre propio. Igualmente por genitivo se traduce cuando el nombre va precedido de un adjetivo cuantitativo, de un numeral car-

dinal u ordinal, de un artículo indeterminado e incluso cuando el nombre va seguido de un pronombre personal o demostrativo, así como en expresiones de tipo *because of, in spite of*.

Como puede observarse, los criterios enumerados se excluyen mutuamente, con lo que la determinación se vuelve unívoca en el modelo de lengua utilizado (cf. 4.5.7).

No está exento de interés conocer también las principales etapas de verificación experimental del algoritmo de traducción que acabamos de describir. Como hemos dicho, las experiencias las realizó, entre febrero y abril de 1962, E. Domonkoş con una máquina MECIPT y bajo la supervisión directa del académico Gr. C. Moisil.

En la primera fase, se verificó la traducción en bruto, sin flexiones, de algunas proposiciones simples, en las que bastaba la traducción palabra por palabra. El corpus utilizado era el siguiente:

- | | |
|---|--|
| — <i>I am content.</i> | — <i>Eu sînt mulţumit</i> ('Yo estoy contento'). |
| — <i>He is beautiful.</i> | — <i>El este frumos</i> ('Él es guapo'). |
| — <i>I explain how I verify and help.</i> | — <i>Eu explic cum eu verific şi ajut</i> ('Yo explico cómo yo verifico y ayudo'). |

En una segunda fase, se desarrollaron los subprogramas para la flexión de las palabras. Es decir, se llevó a cabo la traducción de dos proposiciones con flexión, pero sin adjetivos:

- | | |
|---|---|
| — <i>Verifying the operations she halted the computer.</i> | — <i>Verificînd operaţiile ea a oprit calculatorul</i> ('Verificando las operaciones, ella paró el ordenador'). |
| — <i>You explain the development of the science and we help with the description of the examples.</i> | — <i>Dumneavoastră explicaţi dezvoltarea ştiinţei şi noi ajutăm la descrierea exemplelor</i> ('Usted explica el desarrollo de la ciencia y nosotros ayudamos en la descripción de los ejemplos'). |

En una tercera fase, se verificaron algunas partes del algoritmo que comprendían problemas sintácticos y la ampliación del modelo de lengua, incluyen-

do el cambio en el orden de las palabras y otras adaptaciones sintácticas. Se tradujeron entonces proposiciones como:

- *The speed of the translation does not depend of the examples and depends of the number of operations.*
- *Viteza traducerii nu depinde de exemple și depinde de numărul de operații* ('La velocidad de la traducción no depende de los ejemplos y depende del número de operaciones').
- *They obtained good and rapid translation with computers.*
- *Ei/ele au obținut traducere rapidă și bună cu calculatoare* ('Ellos/ellas obtuvieron traducción rápida y buena con ordenadores').

Se eligió el primer ejemplo para verificar el subprograma de la palabra *of* y el subprograma de la negación⁹. El segundo, en cambio, comporta una modificación sintáctica relativa al orden de cuatro palabras sucesivas.

En el momento presente todavía no puede considerarse absolutamente resuelto el problema de la traducción automática. En este dominio, como en cualquier otro, la máquina electrónica de cálculo no es más que un auxiliar del hombre al que permite efectuar determinadas tareas de rutina.

Sin duda, en nuestros días, los ordenadores electrónicos pueden traducir textos científicos e incluso de lengua corriente siempre y cuando no aparezcan dificultades estilísticas importantes. Pero no es de esperar que, en un futuro próximo, lleguen a realizar por sí solos la traducción de obras literarias a causa de los inconvenientes todavía insolubles que ofrece este tipo de textos.

Hemos de señalar asimismo que a partir de las investigaciones emprendidas en el ámbito de la traducción automática se han realizado experimentos de síntesis sobre algunos textos, empleando algoritmos de formación de estructuras gramaticales y cadenas de números al azar. Evidentemente, ello no constituye más que una experiencia audaz y sin aplicaciones prácticas por el momento. En esencia, el problema de la traducción automática puede examinarse desde dos puntos de vista completamente distintos: el teórico y el práctico.

Desde el punto de vista práctico, o del interés económico, es posible que pase mucho tiempo aún antes de que los traductores humanos sean rentablemente sustituidos en ciertas clases de textos, incluidos los de habla corriente.

9. [Los textos originales en inglés aparecen, sin embargo, un tanto predisuestos. Así, p. ej., en lugar de **to depend of*, hay que decir *to depend upon/on*, 'depender de'. A su vez, la construcción *He is beautiful*, 'él es bonito', de la primera fase, resultaría más natural con el adjetivo *handsome*, 'guapo'].

Pero las experiencias en la traducción automática y, en parte también las de síntesis automática de los textos, demuestran de un modo sobresaliente la capacidad del hombre para averiguar las leyes que rigen en la esfera de las actividades intelectuales. Conociendo los algoritmos que utiliza el intelecto humano en determinadas actividades, cabría trasladarlos a los ordenadores electrónicos que, como hemos visto, trabajan a una velocidad de cálculo incomparablemente mayor que la del cerebro humano.

No cabe la menor duda que, en un futuro no muy lejano, las traducciones automáticas resultarán interesantes incluso desde el punto de vista de su utilización práctica. Hoy por hoy, sin embargo, representan una etapa importante en la dominación de las leyes del pensamiento humano.

7.4. Puntos de vista sobre la traducción automática

A principios de la década pasada, con las primeras traducciones de una lengua a otra, nació la esperanza de crear con cierta rapidez unos sistemas de programación que permitieran traducir prácticamente todos los textos posibles. Pero también se pudo comprobar bastante pronto que el problema presenta muchos más inconvenientes de lo que se suponía a primera vista. Algunas de las dificultades son de índole gramatical: ninguna lengua dispone de gramáticas y de vocabularios exhaustivos.

Por otro lado, como las lenguas están sujetas a una dinámica permanente, es difícil suponer que pudieran confeccionarse gramáticas de este tipo con una validez duradera, digamos de seis a ocho siglos. Pero, incluso para periodos más limitados, las traducciones literarias tropiezan con graves dificultades ante construcciones muy especiales, creadas *ad hoc* por diversos literatos y que constituyen precisamente algunas de las características estilísticas de los autores en cuestión. A ello hay que añadir los problemas de vocabulario, polisemia, etcétera.

Sin caer en el extremo opuesto, el de negar todo valor a la traducción automática, en el momento presente se ha llegado, en ciertas instituciones, a una solución de compromiso que consiste en componer sólo una traducción en bruto para textos técnicos y científicos, sobre los cuales la máquina traduce palabra por palabra sin ulterior elaboración textual. De este modo se obtiene un texto con indicios suficientes para que el interesado averigüe de qué se trata y complete con su propia actividad intelectual lo que no hizo la máquina. Como, por otra parte, la máquina funciona mucho más rápidamente, la operación resulta, en último término, más barata. Dicho de otro modo, durante el mismo intervalo de tiempo, una misma máquina puede satisfacer a muchos más beneficiarios. Para alcanzar una mayor rapidez operativa, se impone, además, mejorar la búsqueda de las palabras. Con este propósito, las palabras se incorporan a la memoria de la máquina en virtud de su frecuencia en orden decreciente; método que ha promo-

vido la confección de diccionarios de frecuencia para la electrónica¹⁰. Con ello, se ha consagrado el nuevo término de ingeniería lingüística, o lingüística ingeniérica, que se ocupa de los aspectos mecánicos de la lingüística, tales como el de mejorar el modo de ejecutar los programas para la traducción, p. ej., del ruso al francés, mejorar el modo de organizar los índices léxicos (llamados “tesoros”), etcétera¹¹. En lo que atañe al último problema, el de cubrir un dominio o corpus de datos a base de criterios categorizadores, o descriptores, y organizar los tesoros, merecen mencionarse las interesantes contribuciones teóricas allegadas por el profesor Aurel Avramescu, que ha establecido una serie de relaciones importantes en este dominio¹².

Como existe la parte de preparación de un texto a partir de la lengua formal con la que se describe el contenido semántico en la memoria del calculador, se plantea, entonces, el problema de si cabe o no generar automáticamente un texto cualquiera. Visto en general, todo se reduce a determinar los algoritmos que permitan construir concatenaciones de letras en un alfabeto dado. Esta suerte de problema se presenta en lingüística, en las artes plásticas y en música, pero con rasgos específicos para cada caso.

Abordemos, de momento, el dominio de la música. La historia de la música consigna la existencia de ciertos métodos de composición aleatoria mediante un “diccionario” atribuido a Mozart. Se trata de una composición automática de valsos a base de ir echando dados como proceso generador de números aleatorios. Según el número obtenido con el lanzamiento de los dados, se elegía un determinado compás del diccionario, el cual se añadía a otros anteriores formando secuencia. La operación continuaba arrojando de nuevo los dados y seleccionando sucesivamente los correspondientes compases. Evidentemente, un tal proceso puede modelarse a partir de un autómata finito.

Hace algún tiempo, la escuela contrapuntista elaboró una serie entera de algoritmos con que podían generarse temas, y sus correspondientes armonías, partiendo de un tema original escogido al azar. Más recientemente, el director Mihai Brediceanu ha defendido una tesis doctoral donde demuestra que a partir de una línea melódica dada pueden obtenerse otras a base de ciertas transformaciones geométricas transferibles al calculador. Merece mencionarse también aquí la corriente aleatoria en música.

Por otro lado, numerosos musicólogos —como, p. ej., Aurel Stroe— sostienen la estricta necesidad del uso de calculadores por parte del compositor actual. Este juicio se justifica de diversas maneras. En primer lugar, el compositor con-

10. Para la traducción del inglés al ruso, cf. Alekseiev-Piotrovski, *Chastotnyi anglo-russkii*. Cf. también 9.4.

11. Cf. *Inzhenernaia lingvistika*.

12. Cf. Avramescu, “Geometric model”.

temporáneo debe operar con masas sonoras mucho mayores que en el pasado, ya que el número de instrumentos ha aumentado con respecto a la orquesta clásica debido a la inclusión de instrumentos exóticos o simplemente nuevos, electrónicos, etc. En segundo lugar, se ha abandonado la gama clásica, con lo que se ha incorporado todo tipo de sonido y se ha incrementado, en consecuencia, tanto el número de sonidos admitidos en la composición como su cualidad misma. En resumen, abandonando las reglas clásicas de la armonía, cabe utilizar cualquier combinación sonora, por lo que el compositor se enfrenta a unas posibilidades muy superiores entre las que debe escoger. En otras palabras, nos hallamos ante unos fenómenos donde la entropía es mucho más elevada (cf. 8.9). La elección, entonces, puede realizarse más fácilmente por cuanto el compositor tiene la posibilidad de transferir al calculador determinadas partes de su trabajo, como las de calcular ritmos, secuencias sonoras, etcétera.

Naturalmente, el problema se puede formular de un modo cómodo, pero falta aún por resolver ciertas dificultades técnicas, como lo prueba, p. ej., el que no todas las composiciones que emplean el calculador reciban el beneplácito del gran público o de los especialistas. En todo caso, hay que destacar el hecho de que en música las cosas discurren con cierta sencillez desde el momento en que no existen problemas semánticos, como en las artes plásticas o en la literatura. La música no es, por lo común, figurativa y ello simplifica mucho las reglas para la concatenación de sonidos.

En el dominio de las artes plásticas, los hechos se hallan en un estado bastante avanzado de desarrollo. Los calculadores disponen de programas para trazar figuras en dos y tres dimensiones y existen, desde unos diez años para acá, rutinas para unir puntos o curvas mediante líneas dotadas de determinadas características. Los programas de este tipo se han elaborado para cubrir principalmente las necesidades del diseño industrial, para simplificar el trabajo de proyección y dibujo e incluso para la preparación automática de piezas con ayuda de máquinas adecuadas de manejo automático. La ejecución de un diseño es una operación bidimensional, en tanto que, p. ej., la elaboración de un programa para fabricar automáticamente una hélice de navío constituye una operación referida a la realización de un objeto tridimensional.

Es, pues, relativamente simple plantear el problema de conformar estos programas al propósito de obtener unos dibujos dotados de unas características dadas en cuanto se elabora la imagen respectiva mediante el calculador. De ello derivan nuevas posibilidades. Así, una de ellas consiste en componer, de acuerdo con unas reglas establecidas, ciertas imágenes de la realidad. Otra posibilidad se refiere a la generación de unas imágenes partiendo de programas generales y de un generador de números aleatorios.

Ilustremos ambos casos. En el primero, puede tomarse una imagen de la realidad, pongamos un retrato, que se transforma en señales eléctricas con ayuda de

unas cámaras tomavistas, como en televisión. Estas señales eléctricas pueden elaborarse de diversos modos. Uno de ellos consiste en cuantificar la iluminación o los colores reduciendo mucho la imagen. Por ejemplo, remitiéndonos al blanco y negro, pueden retenerse sólo aquellas partes de la imagen cuyo gradiente sobrepase un determinado valor. Se forma, de esta manera, un contorno. O bien cabe plantear el problema de modo que, a partir de un cierto punto del plano, se reduzca la imagen a una serie de líneas de distinta anchura y longitud que pasen por aquel punto elegido.

Otro procedimiento se basa en fijar ciertos algoritmos que modifiquen la imagen de un modo cinético. Así, por la superficie de una pantalla aparece un cubo que gira. El cubo no existe en la realidad; sólo está definido matemáticamente en la memoria del calculador, donde existe además el programa de giro, es decir, el programa que genera las secuencias de tensión transformadas en los puntos luminosos que aparecen en la pantalla. Hay asimismo programas que generan imágenes que permiten el paso de una imagen plana, *A*, a otra imagen plana, *B*; p. ej., de una botella de bebida refrescante al mapa de un continente, etcétera, etcétera.

En cuanto a las generaciones aleatorias, puede llevarse a cabo un programa con el que se formen imágenes al estilo de un grafista conocido. Por lo general, se utilizan imágenes no figurativas. En la "Gesellschaft für Mathematik und Datenverarbeitung" de Bonn existen programas para generar diseños aleatorios a la manera de Paul Klee.

Otra vía consiste en generar líneas decorativas mediante recursos puramente electrónicos y sin ninguna relación necesaria con un estilo conocido. Se trata de ciertos programas que se concretan en líneas trazadas sobre la pantalla de un tubo catódico, eventualmente en color, en forma de dibujos realizados con el llamado "plotter", dispositivo empleado en los sistemas de cálculo para el trazado automático de las líneas. En el caso de los diseños al estilo de Paul Klee, el programa preveía el trazado con el plotter de unas estructuras formadas por líneas rectas y círculos de tal manera que las líneas resultaran predominantemente paralelas en dos direcciones dadas, perpendiculares entre sí. Las reglas para el trazado contenían además determinados parámetros que en cada realización tomaban, aleatoriamente, otros valores, con lo que cada diseño quedaba individualizado en alguna medida.

En la actualidad, la gráfica con ordenador se ha extendido suficientemente y cuenta con interesantes realizaciones desde el punto de vista decorativo.

La posibilidad de ejecutar unos programas que contemplen la realización de estructuras tridimensionales ha sido el punto de partida para ciertas investigaciones teóricas y para experimentos concretos referidos a la escultura concebida con el calculador. Se entrevén nuevas fases en esta dirección, empezando por una elemental, en la que el ordenador no hace más que calcular determinadas superficies

conforme a unos programas rígidos que el artista introduce con todo detalle, hasta el extremo opuesto —en rigor, irrealizable hoy por hoy— en el que el calculador crearía totalmente sus propias obras.

El ordenador también puede utilizarse para la creación de objetos arquitectónicos, según opina Gordon Pask.

Todo ello nos demuestra que, de hecho, el calculador empieza a emplearse intensamente en dominios creativos nuevos tales como el de las formas musicales —y, por tanto, auditivas— o plásticas —en última instancia, visuales— a partir de unos algoritmos que el artista delimita con mayor o menor detalle. Desde el punto de vista informacional, cibernético, todo se reduce a la creación de unas estructuras en el marco de unas gramáticas generativas dadas.

Sin entrar a discutir ahora el valor estético de las obras producidas, es preciso subrayar un hecho esencial. Por el momento, los logros más interesantes se encuentran en el dominio de la acústica a causa de una razón simple: en el campo acústico el contenido semántico del mensaje es incomparablemente más pequeño que en la plástica y aún menor que en la literatura. Hasta hoy, los intentos de formalizar el contenido semántico se hallan tan sólo en sus comienzos. El hecho de que un autorretrato de Rembrandt impresione mucho más que la gráfica obtenida por calculador se explica sencillamente si tenemos en cuenta que para el hombre un retrato evoca unos significados mucho más amplios que un dibujo sobre la pantalla de un tubo catódico, donde ciertas líneas se entrecruzan armoniosamente.

En el centro de los dominios creativos subsiste, sin embargo, el problema de la literatura, cuya riqueza semántica todavía se resiste a la automatización. Hay, con todo, experimentos dedicados a la creación automática de estructuras poéticas¹³.

La poesía, sobre todo la de factura moderna, se presta a la automatización, incluso de un modo relativamente fácil cuando se toman determinadas precauciones. En lo que sigue ofrecemos algunos de los resultados obtenidos por uno de los autores (Ed. Nicolau) en el dominio de la creación automática de estructuras poéticas.

Un método bastante simple consiste en seleccionar un corpus poético suficientemente homogéneo, del cual se escogen al azar ciertas porciones que, al ensamblarse, constituyen nuevas estructuras. Como se ve, se trata de una creación combinatorial. Hasta un cierto punto, este procedimiento es comparable al que se produce en el mundo vivo, donde cada organismo asimila del medio ambiente determinadas sustancias que luego elabora. Las plantas absorben elementos químicos y los animales tan sólo macromoléculas.

En un primer estadio experimental, se procedió a tomar macrosegmentos de

13. Cf. Nicolau, *Omni informajonal*.

Omar Khayyam, y luego de T. S. Eliot. He aquí los resultados del ensamblamiento:

I-1 *Căci înlăuntru și'n afară, dincolo, în jur și în jos,
se rotește cenușa —și ea prosperă și acum
micul ei copil calcă greșit în întunecime?
Floarea ce-a'nflorit o dată e moartă pentru'ntotdeauna¹⁴.*

II-1 *Atunci, la acest pământesc tot mai în jos
ale cărui porți sînt vînturi de ploaie
ca și zăpada pe rațiune din patul meu
coboară noi înșine nu sîntem nici aici nici acolo¹⁵.*

Precisemos que las rubayyatas combinatorias fueron compuestas a base de una técnica de montaje aleatorio, empleando la creación de Omar Khayyam de Nishapur en la versión de Edward Fitzgerald y traducida al rumano por Edmond Nicolau (1971).

En la variante I se utilizaron versos enteros; en la II, mitades de verso. En ambos casos, la elección fue aleatoria. Transcribimos al inglés las mismas rubayyatas¹⁶:

I-2 *For in and out, above, about, below,
Turns ashes —or it prospers; and anon,
Her little children stumbling in the dark?
The flower that once has blown for ever dies.*

II-2 *Then to this earthly about, below
Whose doorways are the winds like rain
Like snow upon reason from my bed
Descend ourselves is neither here nor there!*

Evidentemente, como en toda técnica aleatoria, combinatorial, en que se combina, en un número finito, un número finito de elementos, también el número de posibilidades resultante es finito. Con la primera técnica existe la posibilidad de formar cerca de 10^7 estrofas, mientras que con la segunda el número crece hasta alrededor de 10^{14} .

Otra técnica distinta consiste en utilizar una gramática generativa con que

14. ['Porque por dentro y por fuera, más allá, alrededor y debajo/ gira la ceniza—y ella prospera y ahora/ ¿su pequeño niño pisa errado en la oscuridad?/ La flor que floreció una vez está muerta para siempre'].

15. ['Entonces, en este terrenal cada vez más abajo/ cuyas puertas son vientos de lluvia/ como la nieve sobre la razón de mi lecho/ descendi nosotros mismos no estamos ni aquí ni allá'].

16. [Hay que advertir que la traducción no es absolutamente literal].

combinar palabras a partir de un vocabulario dado. Desde luego, en este caso aparecen ya determinados problemas semánticos. La elección del vocabulario es, sin embargo, esencial. He aquí dos muestras de creación realizadas con esta técnica:

*B-2 mîine, peste infinit,
nesfîrșit, clopotul bate
nesfîrșit*¹⁷.

*A-14 Singurătatea ta să fie suavă ca marmura.
Dă-mi cuvinte albastre, verzi și palide,
rinduri scrise bine.
Pentru leacurile plecate, te-am ridicat,
sorocită să fii lîngă suflul,
copila mea*¹⁸.

Se ha empleado, en este caso, un vocabulario de palabras basado en la obra poética de L. Blaga, para el texto *B*, y de T. Arghezi, para el *A*.

Para terminar, veamos un texto compuesto a partir de poemas de T. S. Eliot:

*So closely! I myself can hardly understand.
Let us go and make our visit
Puote veder del amor che a te mi scalda
(Another hid his eyes behind his wing)
There will be time to murder and create.*¹⁹

En la actualidad, la creación de textos en sentido amplio puede realizarse sin dificultad en la música y eventualmente en las artes plásticas. En el dominio poético hay posibilidades y logros, pero la perspectiva es aún bastante limitada. Por lo que se refiere a la generación de textos épicos o dramáticos, no se registran realizaciones significativas. Es necesario, para ello, que el calculador pueda ser programado de antemano a fin de manipular también el contenido semántico de los mensajes y de sus palabras respectivas²⁰.

17. ['mañana, más allá del infinito./ sin fin, la campana tañe/ sin fin'].

18. ['Tu soledad sea suave como el mármol./ Dame palabras azules, verdes y pálidas./ líneas bien escritas./ Para los remedios idos, te he elevado./ destinada a estar junto al alma./ niña mía'].

19. ['¡Tan cerca! Yo mismo apenas puedo comprender./ Vayamos y hagamos nuestra visita/ Puede ver del amor que por ti me inflama/ (Otro ocultaba los ojos tras el ala)/ Habrá tiempo para asesinar y crear'].

20. Cf. NICOLAU-BĂLĂCEANU, "Semnificat".

8. Probabilidad, información y entropía en la lengua

8.1. Leyes cuantitativas en la lengua

El fenómeno de la sinonimia, ampliamente extendido en la lengua, permite que cada uno imprima determinadas particularidades al propio lenguaje, que se manifiestan como un cierto modo de expresar las ideas, una manera específica de construir las frases o una suerte de preferencia de uso en favor de algunas palabras y en detrimento de otras. Las posibilidades de la lengua rumana [o de cualquier otra, sin excepción] para proveer de nuevas combinaciones de palabras, o nuevas modalidades en cuanto es necesaria la expresión de ideas y sentimientos nuevos, es inagotable. Las particularidades de estilo surgen en especial en la lengua de los grandes escritores. Por ello, quienquiera que conozca algo la literatura rumana contemporánea identificará un poema de Arghezi o de Blaga, un cuento de Sadoveanu, un reportaje de Geo Bogza o un artículo de G. Călinescu. La gran cantidad de estilos posibles y la impresión misma de que disponemos de una amplia libertad en lo que se refiere al modo de expresarnos y de hacernos comprender no contradicen, sin embargo, la regularidad que existe en la lengua ni el carácter sistemático de ésta. Los sonidos no se combinan al azar cuando se componen palabras con ayuda de ellos, ni las palabras se ligan arbitrariamente al formar una frase. Nuestras posibilidades para utilizar la lengua en un sentido o en otro tienen determinados límites. Así, p. ej., no podemos expresarnos ni hacernos entender sin emplear conjunciones o preposiciones, sin respetar un cierto orden en las palabras de la frase y algunas reglas de concordancia. Pero el carácter sistemático de la lengua no se reduce sólo a estos rasgos, visibles incluso para quien haya aprendido un poco de gramática, un poco de morfología y sintaxis. Los lingüistas han descubierto una multitud de hechos menos visibles, que revelan relaciones mutuas complejas entre los diversos aspectos de la lengua. En efecto, se ha observado que, normalmente, las palabras empleadas más a menudo en rumano pertenecen al fondo primitivo de la lengua,

en especial a las heredades del latín y a los préstamos del eslavo. A medida que una palabra se hace más utilizada, aparecen modificaciones en ella, tanto en sus significados como en su pronunciación. Se ha observado igualmente que, al menos en ciertas lenguas, existe una clara relación entre el grado de uso de una palabra y la naturaleza de los sonidos que la forman. En francés, p. ej., en las palabras muy frecuentes se manifiesta una ligera preponderancia de labiales¹. Todos estos rasgos operan sobre cualquier hablante, independientemente del deseo y de la personalidad de éste.

Como vemos, el carácter sistemático de la lengua se refiere a veces a situaciones cuantitativas dadas y a ciertas proporciones entre los distintos elementos de la lengua. Esta comprobación nos lleva a la idea de que la matemática puede desempeñar un importante papel explicativo en este tipo de problemas. Como se sabe, los aspectos cuantitativos de los fenómenos son precisamente objeto de estudio en matemática (cf. 2.1, 2.19 y 2.20). Un estudio cuantitativo que evitara el empleo de los números sería incompleto, falto de precisión. Y, por otra parte, tampoco hemos de caer en el extremo opuesto forzando todo aspecto cuantitativo que presente la lengua y obligándolo a entrar en el estrecho marco de una fórmula matemática.

8.2. Precisiones numéricas en algunas cuestiones de vocabulario

La regularidad de la lengua, desde el punto de vista de sus aspectos cuantitativos, se manifiesta a veces de un modo más bien oculto. Para ponerlo en evidencia, hay que estudiar numerosos textos y emplear ciertas estadísticas bastante meticolosas, cálculos, gráficos e incluso determinadas fórmulas matemáticas. En este párrafo y en el siguiente, daremos algunas ilustraciones a este respecto.

Las investigaciones estadísticas efectuadas sobre diversas lenguas han facilitado interesantes conclusiones sobre la distribución de las palabras en un texto cualquiera. Indiquemos con f el número que se refiere a las veces que aparece una palabra en un texto determinado, o sea f será la frecuencia de esta palabra en el texto. Se ha comprobado que si colocamos las palabras de un texto en orden decreciente a su frecuencia, las primeras 15 palabras de la lista cubren aproximadamente un cuarto del texto, las primeras 66 palabras cubren aproximadamente el 72% del texto. El término “aproximadamente” indica que se trata de verdades no rigurosamente exactas, sino válidas con una cierta aproximación. Sin embargo, cuanto más largo es el texto examinado, tanto mayor es la aproximación, es decir tanto menor es el error, de modo que podemos reducirlo tanto como queramos con tal de que el texto sea lo bastante largo. Este tipo de verda-

1. Las labiales son sonidos vocálicos o consonánticos en cuya articulación participan los labios, como en /o/, /u/, /b/, /f/, /m/, /p/, /v/.

des se denominan *verdades estadísticas*. Todas las leyes que trataremos en lo que sigue tienen un carácter estadístico.

De lo dicho hasta ahora se deduce que, en un texto dado, un número muy pequeño de palabras suministra la mayor parte del texto. Se puede ir más lejos y plantear el problema de si del conjunto de las palabras que componen un texto dado no se puede extraer un número reducido de palabras que cubra la mayor parte de un texto cualquiera. La resolución de un problema como éste es mucho más difícil y requiere bastantes estudios preliminares. Aquí nos basta tener en cuenta que el léxico de una lengua moderna utilizada por un pueblo con un alto grado de civilización se eleva a varias decenas de miles de palabras y supera, a veces, las 100 000. Así, el *Dictionarul Limbii Române Moderne*, 'Diccionario de la Lengua Rumana Moderna', editado por la Academia de la República Socialista de Rumanía, contiene más de 50 000 palabras. Para algunas lenguas, se han confeccionado los llamados diccionarios de frecuencias, es decir diccionarios donde las palabras se hallan inscritas con sus frecuencias en la lengua (cf., para el rumano, 2.20)². Por frecuencia de una palabra en la lengua se entiende una media de sus frecuencias en textos que pertenezcan a estilos diversos: artístico, literario, científico, jurídico, periodístico, etcétera. Cuanto mayor es el número y la extensión de los textos considerados y cuanto más variados estilísticamente sean estos textos, tanto mayor es la aproximación a la verdad sobre la frecuencia de las palabras en la lengua. El diccionario de frecuencias sólo comprende una pequeña parte (a lo sumo, algunos miles de palabras) del léxico de la lengua. Las palabras se incorporan en orden decreciente de frecuencia, de manera que las palabras figuran en número mayor cuanto más grande sea su frecuencia en la lengua. La composición de diccionarios de frecuencias ha permitido hacer ciertas comprobaciones particularmente interesantes. Así, p. ej., las primeras 100 palabras del diccionario de frecuencias cubren el 60 % de cualquier texto; las primeras 1 000 palabras, el 85 %, y las primeras 4 000, el 97,5 %. Por supuesto que ya anteriormente se sabía que una gran parte de las palabras de una lengua tienen un uso extremadamente raro; no obstante, las verdades estadísticas que acabamos de observar aducen una precisión numérica al conocimiento de este fenómeno.

El método estadístico descubre ciertas particularidades interesantes de los distintos estilos. Para dar sólo un ejemplo, citaremos las investigaciones emprendidas hace un par de decenios en la Unión Soviética por V. A. Nikonov³. Este autor ha demostrado que en los textos políticos y científicos escritos en

2. [Para el español, disponemos de JULLAND-CHANG, *Frequency Dictionary*, sobre el cual cf. BUSTOS, "Nuevo recuento". Cf., también, GARCÍA, *Vocabulario* y, para cuestiones generales, MÜLLER, *Introduction*, y KOCK, *Introducción*].

3. Cf. NIKONOV, "Statistiki padezhei".

ruso los nombres prefieren el caso genitivo (recordemos que, en ruso, la declinación nominal tiene seis casos). He aquí los datos:

	Nominativo	Genitivo		Nominativo	Genitivo
Lenin	26,0	36,7	Karamzin	33,0	17,7
Timiriázev	26,7	45,7	Pushkin	33,8	18,6
Pavlov	21,9	36,2	Chejov	33,7	17,1
Iofe	20,6	39,2	Gorki	33,0	16,7

Cuadro 21

Los datos de la tabla deben interpretarse del siguiente modo: de 100 nombres tomados al azar en textos de Chejov, 34 aparecen en nominativo y sólo 17 en genitivo.

He aquí, pues, que las diferencias que existen entre los procedimientos artísticos de Karamzin, Pushkin, Chejov y Gorki, el hecho de que cada uno de ellos se perfile como una personalidad poderosa, como un escritor excepcionalmente original y fácilmente reconocible a partir de unos pocos textos, no impide la existencia de fuertes rasgos comunes que se manifiestan como expresión de una ley o de una característica de los textos literarios. Observaciones análogas se desprenden asimismo de los textos políticos y científicos.

Estas verdades examinadas tienen un carácter estadístico. Podría ocurrir que, tomando una sola página de Gorki, desapareciera la correspondencia apreciada en la tabla entre el nominativo y el genitivo de los nombres. Pero a medida que crezca la extensión de los textos considerados, las cifras de la tabla recibirán una confirmación cada vez más exacta.

El lector puede comprobar en qué medida se verifican en español los datos obtenidos por V. A. Nikonov para el ruso. Cabría comparar, en este sentido, los resultados extraídos de textos pertenecientes a las ciencias naturales con los extraídos de textos literarios.

Fácilmente se comprende, entonces, la importancia de los estudios estadísticos en la composición de manuales de lenguas extranjeras. Estos estudios permiten que el vocabulario introducido en un manual de este tipo responda a una selección científica a fin de requerir el menor esfuerzo posible a la memoria del lector.

Ahora bien, las investigaciones estadísticas sobre la lengua no se detienen aquí. Es importante estudiar la frecuencia de cada letra y de cada grupo de letras, de cada sonido y de cada grupo de sonidos, de cada parte del discurso

(nombre, adjetivo, verbo, etc.), de los modos y los tiempos verbales y, en fin, la frecuencia de las distintas configuraciones sintácticas. Este tipo de estudio puede hacerse sobre autores y estilos diversos. Sería interesante, p. ej., emprender estudios estadísticos comparando el ámbito de diferentes disciplinas: matemática, física, etc.

8.3. Fenómenos lingüísticos que obedecen a determinadas fórmulas

Sea un número natural L y consideremos la totalidad de los textos que tengan una extensión L . Por extensión de un texto entendemos el número de las palabras que lo forman, teniendo en cuenta que cada palabra debe contarse tantas veces como aparezca en él⁴. Por vocabulario de un texto se comprende el número V de las palabras distintas que aparecen en él.

Supongamos que las palabras de un texto cualquiera, de extensión L , han sido colocadas en orden decreciente respecto de su frecuencia en el texto y que las palabras con una misma frecuencia se agrupan juntas sin seguir un orden interno determinado. El número de orden de una palabra en una lista de esta clase se llama *rango* de la palabra en el texto. Hay una ley fundamental, tal vez la más importante de las leyes estadísticas del vocabulario, que afirma que el producto del rango de una palabra por su frecuencia es constante⁵. Esta ley debe interpretarse en el sentido de que el producto resultante de multiplicar el rango y la frecuencia de una palabra cualquiera de un texto tiene un valor constante. Tal valor cambia, no obstante, en cuanto se cambia el texto y, a su vez, la exactitud con que se verifica la ley es tanto mayor cuanto más grande sea la extensión L del texto. Notando por medio de $K(L)$ el valor de la constante correspondiente a un texto de extensión L , obtenemos:

$$1 \cdot f_1 = K(L), \quad 2 \cdot f_2 = K(L), \quad 3 \cdot f_3 = K(L), \dots, \nu \cdot f_\nu = K(L),$$

donde f_i es la frecuencia de la palabra de rango i . Deducimos:

$$if_i = f_1 \quad (i = 1, 2, \dots, \nu).$$

De modo que podemos escribir:

$$\begin{aligned} L &= f_1 + f_2 + \dots + f_i + \dots + f_\nu = \\ &= f_1 \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{i} + \dots + \frac{1}{\nu} \right). \end{aligned}$$

4. Dos formas flexivas de una misma palabra (v. gr., *yo, mi, me*) se consideran como palabras diferentes.

5. Esta ley fue señalada en 1916 por J. B. Estoup, autor de un manual de estenografía, y luego confirmada, en 1949, por G. K. Zipf. De ahí que se llame "ley de Estoup-Zipf".

En el paréntesis del segundo miembro de la igualdad obtenida aparecen los primeros ν términos de una serie célebre, la serie armónica⁶. Cabe, por tanto, decir que el vocabulario de un texto se distribuye según una serie armónica.

La ley que acabamos de describir se denomina ley de rango-frecuencia. En un sistema de coordenadas cartesianas en que el rango represente la abscisa y la frecuencia la ordenada, esta ley se manifiesta como una hipérbola equilátera (fig. 24).

Normalmente, para disminuir las dimensiones del gráfico, se opera con las llamadas coordenadas logarítmicas, que vienen a representar, en el eje de las abscisas, a $\lg_{10}r$ en lugar de r y, en el de las ordenadas, a $\lg_{10}f$ en lugar de f . En este tipo de sistema de coordenadas, la ley de rango-frecuencia queda representada por una línea recta de pendiente -1 (fig. 25).

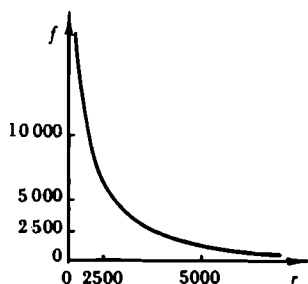


Figura 24

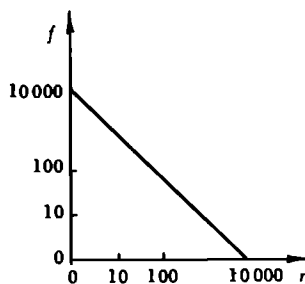


Figura 25

He aquí una ilustración de la ley de rango-frecuencia sobre un texto de extensión igual a 60 000:

Rango	Frecuencia	Rango \times frecuencia
10	2 653	26 530
100	265	26 500
1 000	26	26 000
10 000	2	20 000
29 000	1	29 000

Cuadro 22

Emparentada con la ley de rango-frecuencia se halla la ley de número-frecuencia, que afirma que a cada número natural L le corresponde un número C_L tal que, en todo texto de extensión L , el producto del cuadrado de

6. [Es decir, $\sum_{n=1}^{\infty} \frac{1}{n}$].

un número f por el número n_f de las palabras de frecuencia igual a f en el texto respectivo es igual a C_L , cualquiera que sea el número natural f , inferior o igual a la frecuencia de la palabra de rango 1:

$$n_f \cdot f^2 = C_L.$$

Véase la representación gráfica de la ley de número-frecuencia, tanto en coordenadas cartesianas normales (fig. 26) como en coordenadas logarítmicas (fig. 27).

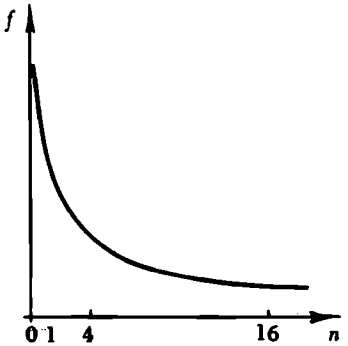


Figura 26

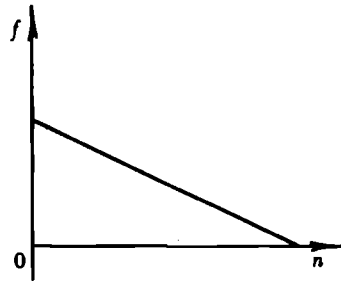


Figura 27

Puede demostrarse que esta ley es una consecuencia de la ley de rango-frecuencia y que, igual que ésta, sólo es verdadera aproximadamente a causa de su carácter estadístico. He aquí una ilustración de la ley de número-frecuencia para los primeros diez valores de f :

f	n_f	$n_f \cdot f^2$
1	16 432	16 432
2	4 776	19 104
3	2 194	19 746
4	1 400	22 400
5	900	22 500
6	770	27 720
7	480	23 520
8	370	23 680
9	300	24 300
10	220	22 000

Cuadro 23

Sería interesante verificar la ley de rango-frecuencia y la de número-frecuencia sobre textos tomados al azar entre manuales escolares a fin de confrontar los gráficos obtenidos con los gráficos teóricos aducidos antes. Es recomendable trabajar sobre papel milimetrado y, si los textos son muy extensos, conviene utilizar coordenadas logarítmicas para que el gráfico resulte más espacioso.

Como se ve en las tablas anteriores, las leyes de rango-frecuencia y de número-frecuencia a veces se cumplen un tanto precariamente, al menos para los valores bajos de rango y, respectivamente, de frecuencia.

Una descripción más fiel sobre el carácter dependiente de la frecuencia con respecto al rango viene dada por la relación:

$$f_r \cdot r^\alpha = K(L),$$

donde f_r es la frecuencia de la palabra de rango r , α es un número cercano a 1 (normalmente, mayor que 1), mientras que el valor $K(L)$ depende de la extensión L del texto y, en menor grado, del texto mismo. En este caso, obtenemos:

$$\begin{aligned} L &= f_1 + f_2 + \dots + f_i + \dots + f_\nu = \\ &= K(L) \left(\frac{1}{1^\alpha} + \frac{1}{2^\alpha} + \dots + \frac{1}{i^\alpha} + \dots + \frac{1}{\nu^\alpha} \right), \end{aligned}$$

por lo que podemos decir que el vocabulario de un texto se distribuye de acuerdo con una serie armónica generalizada. Se sabe que una serie de este tipo es convergente para $\alpha > 1$ y divergente para $\alpha \leq 1$. Si $\alpha > 1$, entonces, cuanto menor es la diferencia $\alpha - 1$, tanto más despacio converge la serie, mientras que para valores muy pequeños de $\alpha - 1$, la serie se hace, prácticamente, divergente. Justamente esta situación se produce aquí, por cuanto los valores de α son, para la mayoría de los textos, muy cercanos a 1.

La serie $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ fue estudiada por Riemann. La suma de ella es una función de α , la cual se indica, por lo común, mediante la letra griega ζ (dseta) y se denomina función ζ de Riemann.

B. Mandelbrot mejoró la ley de rango-frecuencia proponiendo para la frecuencia la siguiente función de rango, que también llamaremos función de Estoup, Zipf y Mandelbrot:

$$f(r) = PL(r + \rho)^{-\beta}, \quad (1)$$

donde L es la extensión del texto, es decir el número de todas las palabras (distintas o no) contenidas en él, β y ρ son dos parámetros cuyos valores dependen del texto ($\beta > 0$), mientras P viene dado por la fórmula:

$$\frac{1}{P} = (1 + \rho)^{-\beta} + (2 + \rho)^{-\beta} + \dots + (\nu + \rho)^{-\beta}, \quad (2)$$

donde ν es el número de palabras distintas del texto. Una descripción "grosso

modo" de un texto se adquiere estudiando los parámetros β , ρ y ν , que determinan la función de Estoup, Zipf y Mandelbrot aplicable al texto en cuestión.

En sus estudios experimentales, Zipf concedió una especial atención a textos que se caracterizan por una particular variedad. Puede demostrarse, sin embargo, que precisamente por la variedad del vocabulario de estos textos el valor de β está muy cerca de 1, lo que acerca, para r grande, la fórmula (1) a la fórmula inicial de Estoup. A esta fórmula Zipf le concedió equivocadamente una validez general. Por lo demás, de los parámetros que aparecen en la fórmula (1), β es el más importante. Su especial significación surge sobre todo para $\beta > 1$, lo que ocurre en la mayoría de los textos. En este caso, cuanto menor es β , es decir cuanto más cerca está de 1, o cuanto mayor es $1/\beta$, tanto más rico y variado resulta el vocabulario del texto respectivo, donde aparecerán numerosas palabras inusitadas o poco frecuentes con una frecuencia apreciable. Por este motivo, el número $1/\beta$ se denomina *temperatura de información* del texto. Los textos escritos en un estilo que emplea recursos variados y un vocabulario rico se caracterizan por una temperatura de información muy grande (es decir, muy cercana a 1). La temperatura de información de los textos infantiles es disminuida, pero se acerca a 1 a medida que los niños crecen y desarrollan su inteligencia.

Hay muy pocos textos donde el valor de β resulte inferior a 1. Puede, no obstante, ocurrir cuando existe el propósito de transmitir el máximo de información en un texto lo más corto posible.

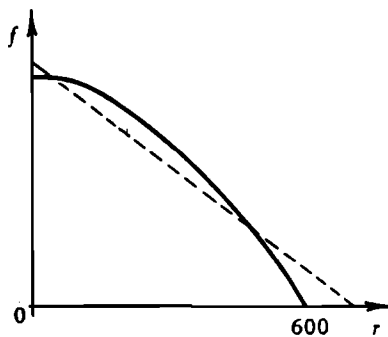


Figura 28

En la figura 28 se observa cómo se manifiesta, en un sistema de coordenadas logarítmicas, la curva de rango-frecuencia, tal como resulta de la función de Estoup, Zipf y Mandelbrot (la línea intermitente se refiere a la ley de Estoup-Zipf).

Se advierte que, a medida que r crece, el gráfico de la función $f(r)$ se hace más y más asimilable a una línea recta. Ello resulta de un pequeño cálculo. En efecto, tenemos que:

$$\lg f(r) = \lg (PL) - \beta \lg (r + \rho)$$

y cuanto mayor es r , la relación

$$\lg (r + \rho) = \lg r$$

es más verdadera ofreciendo un error cada vez más pequeño. Así, pues, a medida que r aumenta, la función de Estoup, Zipf y Mandelbrot se acerca, en las coordenadas logarítmicas, a una función lineal, representada por la función $\lg f(r) = \lg (PL) - \beta \lg r$, cuyo gráfico refleja una recta de pendiente igual a $-\beta$. La presencia del parámetro ρ introduce, para r pequeño, una determinada desviación en la función de Estoup, Zipf y Mandelbrot con respecto a una función lineal.

Suponiendo que $\beta > 1$, lo que, como hemos demostrado, sucede en la mayoría de los textos, la función:

$$f(x) = \frac{1}{(x + \rho)^\beta}$$

es decreciente para x positivo, de modo que⁷:

$$\frac{1}{(n + \rho)^\beta} \leq \int_n^v \frac{dx}{(x + \rho)^\beta} \leq \frac{1}{(n - 1 + \rho)^\beta}.$$

Teniendo en cuenta que estas desigualdades son válidas para todo valor entero de n , comprendido entre 1 y v , deducimos por suma:

$$\sum_{n=1}^v \frac{1}{(n + \rho)^\beta} \leq \int_0^v \frac{dx}{(x + \rho)^\beta} \leq \sum_{n=1}^v \frac{1}{(n - 1 + \rho)^\beta},$$

y, por tanto:

$$0 \leq \int_0^v \frac{dx}{(x + \rho)^\beta} - \sum_{n=1}^v \frac{1}{(n + \rho)^\beta} \leq \frac{1}{\rho^\beta}$$

o, teniendo en cuenta la relación (2),

$$0 \leq \int_0^v \frac{dx}{(x + \rho)^\beta} - \frac{1}{P} \leq \frac{1}{\rho^\beta},$$

de forma que el error que se comete al considerar P igual a

$$\frac{1}{\int_0^v \frac{dx}{(x + \rho)^\beta}}$$

es inferior a $\frac{1}{\rho^\beta}$.

7. [Como se sabe, aquí x , limitada entre $n-1$ y n , es la variable de la función integral].

Véase en la figura 29, también en coordenadas logarítmicas, algunas curvas de rango-frecuencia deducidas por G. K. Zipf a partir de datos numéricos correspondientes a textos que él mismo estudió.

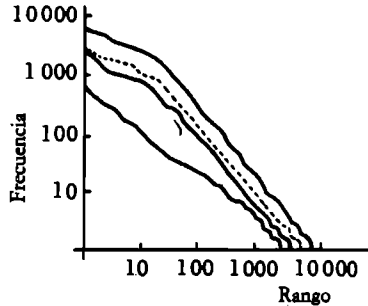


Figura 29

Como se ve, la función de Estoup, Zipf y Mandelbrot corresponde más al fenómeno real que la función de Estoup y Zipf.

Investigaciones más recientes han demostrado que la ley de rango-frecuencia conserva la validez especialmente para valores ni demasiado pequeños ni demasiado grandes del rango.

Es sabido que ciertas palabras pueden tener más de un significado. En el *Dictionarul limbii române moderne*, como en la mayoría de los diccionarios, las acepciones de una palabra van numeradas. El análisis del diccionario muestra que el número n_s de las palabras que tienen un número s determinado de significados es inversamente proporcional al cuadrado del número de significados:

$$n_s \cdot s^2 = C.$$

También esta ley tiene un carácter estadístico, donde el número C que figura en el segundo miembro de la relación aducida depende del número de palabras que ofrece el diccionario y, por tanto, varía de uno a otro diccionario.

Si no hay posibilidad de efectuar esta comprobación sobre un diccionario entero, puede hacerse sobre otro más pequeño: tomando, p. ej., al azar un número determinado de palabras de un diccionario grande. He aquí el resultado obtenido para 545 palabras extraídas en estas condiciones:

s	n_s	$s^2 \cdot n_s$
1	364	364
2	100	400
3	37	333
4	21	336
5	8	200
6	9	324
7	3	147
8	1	64
9	0	0
10	1	100
16	1	256

Cuadro 24

El lector mismo puede verificar esta ley tomando al azar unas 600 palabras del *Diccionario de la Lengua Española*. La selección puede hacerse, p. ej., considerando la última palabra de cada página.

8.4. Preocupaciones y perspectivas en el estudio estadístico de la lengua rumana

El propósito de utilizar algunos métodos matemáticos en el estudio del rumano es más bien reciente, pese a que se advierten ciertos anticipos ya en la obra *Limba în circulațiune, 'La lengua en circulación'*, de B. P. Hasdeu, con la que emitió, a fines de 1881, una "teoría de la circulación de las palabras" (cf. 9.5).

En la actualidad, aparte del *Frequency Dictionary of Rumanian Words* de A. Juilland, P. M. H. Edwards e I. Juilland, la mayoría de los estudios estadísticos se refieren al rumano escrito⁸.

En 1954, aparece la obra del académico Al. Graur, donde el método estadístico se emplea como auxiliar para determinar la estructura léxica del rumano⁹. Como señala el autor, el método estadístico ha revelado hechos nuevos relativos a la estructura etimológica del léxico de la lengua rumana, entre los cuales cabe citar el de que las palabras eslavas ocupan un lugar importante (más de una quinta parte) en el fondo principal mismo de la lengua.

A finales del año 1958, por iniciativa del académico Al. Rosetti, se fundó, en el ámbito del Instituto de Lingüística de la Academia de la R. S. de Rumanía, un círculo de lingüística matemática. En la revista *Studii și Cercetări Lingvistice, 'Estudios e Investigaciones Lingüísticas'*, editada por este Instituto, han

8. Cf. 2.20, nota 14 y la Bibliografía. Para el rumano hablado, cf. MACREA, "Frecvență".

9. Cf. GRAUR, *Încercare*.

aparecido, en distintas ocasiones, algunos trabajos en este sentido. En un artículo de V. Şuteu, p. ej., se estudia la frecuencia de las palabras en diez escritores rumanos de la primera mitad de siglo, donde se comprueban ciertas desproporciones entre la frecuencia de las palabras y su distribución, entendiéndose aquí por “distribución” el número de autores en que la palabra aparece al menos una vez¹⁰.

En este mismo orden de ideas, cabe citar la útil contribución de Gh. Bolocan en un trabajo donde indica algunas particularidades estadísticas del estilo publicitario que lo distinguen del literario¹¹.

Un ciclo interesante de artículos relativos a la estructura estadística de la lengua es el publicado por A. Roceric-Alexandrescu¹². En otros casos, se han aplicado determinadas fórmulas de P. Guiraud al estudio de textos literarios¹³.

Los estudios estadísticos sobre el vocabulario han merecido la atención también del académico Tudor Vianu, el cual ha hecho algunas consideraciones sobre la obra más conocida de Guiraud¹⁴. Este libro ha provocado apasionadas controversias y reservas, algunas procedentes incluso de matemáticos como B. Mandelbrot. Guiraud, realmente, va muy lejos en el uso del aparato estadístico, con el que llega a formulaciones atrevidas sobre el estilo de algunos grandes escritores franceses del pasado. En todo caso, sería un error rechazar totalmente las ideas de P. Guiraud. Llevar hasta las últimas consecuencias ciertas consideraciones produce, a veces, la sospecha de que se atenta contra la complejidad y el contenido emotivo de la lengua literaria. El método estadístico debe, sin embargo, completarse con los demás métodos de investigación; y, a su vez, los resultados deberían interpretarse y subordinarse a un punto de vista que comprenda todos los métodos. En tanto que se intente transformar el método estadístico en un método exhaustivo de investigación sobre la lengua de Eminescu, p. ej., se comprobará, evidentemente, que quedan aspectos de la mayor importancia fuera del estudio, y lo que constituye un límite natural del método mismo terminará pareciendo una deficiencia imperdonable.

8.5. La necesidad de una teoría científica de la información

Gracias a su función comunicativa, la lengua se estudia hoy desde el ángulo no sólo de la lingüística, sino también de la cibernética. En efecto, según la definición del gran matemático A. N. Kolmogorov, la cibernética es la ciencia de los modos de recepción, conservación, transformación y empleo de la información

10. Cf. ŞUTEU, “Observații asupra frecvenței”.

11. Cf. BOLOCAN, “Unele caracteristici”.

12. Cf. ROCERIC-ALEXANDRESCU, “Aplicarea metodei statistice” y “Recherches statistiques”.

13. Cf. GOLOPENȚIA-TOMA, “Statistică”.

14. Cf. GUIRAUD, *Caractères*; VIANU, “Statistică lexicală”, y NICOLAU-BĂLĂCEANU, *Cibernetica*.

en máquinas, organismos vivos y sus reuniones. Este hecho pone en una nueva perspectiva algunos de los problemas más viejos de la lingüística y, como veremos, determina la aparición de otros nuevos, que no se hubieran podido formular antes de aparecer la cibernética.

La palabra *información* tiene un significado muy amplio, impregnado de elementos intuitivos. La información principal que transporta la lengua es de naturaleza semántica. Por el momento, las posibilidades de hacer un tratamiento exacto sobre este aspecto de la información son todavía bastante reducidas. Hasta hace 40 años, nadie sospechaba de qué manera podría perfilarse una teoría matemática de la información.

La información es algo muy complejo, que aparece en los supuestos más variados. El habla humana, los fenómenos de la actividad nerviosa, la transmisión de los caracteres hereditarios, los procesos económicos, la comunicación telegráfica, telefónica, de radio o televisión son, todas ellas, distintas modalidades en las que se expresa una información. Estas informaciones pueden ser enormemente variadas desde el punto de vista de su contenido o del modo en que se transmiten y reciben. Los hombres de ciencia se han preguntado si no podría, sin embargo, hacerse abstracción de este aspecto cualitativo a fin de estudiar exclusivamente el aspecto cuantitativo. Con el patrón que llamamos “metro” medimos la longitud indistintamente de un trozo de paño, de un alambre o de una línea férrea. Con el patrón llamado “kilógramo” apreciamos el peso, operación también indiferente a la naturaleza del objeto en cuestión. ¿No es posible encontrar, entonces, un patrón para medir la cantidad de información, es decir una unidad de medida, una información estándar, objetivamente determinada, con cuya ayuda midiéramos las informaciones y las comparáramos independientemente de su contenido y del modo en que fueran emitidas, transmitidas y captadas?

Precisemos desde ahora que por *aspecto cuantitativo de la información* entendemos la estructura estadística de la misma: admitiendo que la forma típica en que se nos manifiestan las informaciones es la de una cadena de símbolos, lo único que nos interesa es la frecuencia de aparición de estos símbolos y de sus distintas combinaciones. Así, p. ej., estudiamos la frecuencia de las letras, los fonemas, los morfemas y las palabras y la frecuencia de las diversas combinaciones de letras, morfemas, etcétera. Está claro que a partir de ello el concepto de información queda considerablemente empobrecido, pero en el estado actual de las investigaciones éste es el único modo de abordarlo con exactitud. Algunos intentos más recientes de Bar-Hillel, Carnap y otros de abarcar también la información semántica dentro de la esfera de la teoría de la información han resultado, al menos de momento, bastante estériles¹⁵.

15. [Cf., p. ej., BAR-HILLEL, “Dictionaries”, y CARNAP, *Meaning*].

Hasta cierto punto, las informaciones pueden compararse por vía intuitiva.

Sean dos urnas A y B tales que la primera contenga tres bolas de color diferente y la segunda cuatro, también de distinto color.

Supongamos que alguien saca de cada urna una bola y, sin que la veamos, nos dice: “He sacado de la urna A una bola verde y de la urna B una bola blanca”. En este caso hemos obtenido dos informaciones desiguales, ya que, en efecto, la información que consiste en precisar el color de una bola de A (la urna con menos bolas) es, evidentemente, menor que la información que nos dice de qué color es la bola sacada de B . Esta diferencia cuantitativa entre las dos informaciones se advierte sin necesidad de recurrir a una teoría científica de la información. ¿Acaso se infiere de ello que una tal teoría es inútil? Si contestáramos afirmativamente a esta pregunta, también deberíamos rechazar la necesidad de una teoría científica de la temperatura y, con ella, un capítulo entero de la física. Y es que, sin duda, cualquiera puede darse cuenta con sus propios sentidos de que en invierno hace frío y en verano, calor.

Una teoría científica de la información se propone, sin embargo, “prolongar” nuestras posibilidades intuitivas de apreciar y comparar las informaciones, del mismo modo que el microscopio “prolonga”, multiplica, un número enorme de veces, las posibilidades del sentido humano de la vista. Una teoría científica de la información confirmará a veces las apreciaciones intuitivas, pero en otras distinguirá matices intuitivamente inaccesibles. Las informaciones se convierten en números con que puede calcularse. También existen, naturalmente, casos en que la apreciación intuitiva de la información difiere de la información científica de una forma aparentemente negligible. Pero esta diferencia aparentemente mínima puede producir grandes desplazamientos en los resultados obtenidos a partir de los razonamientos y los cálculos efectuados con las cantidades de información.

8.6. Economía y derroche

Los que al menos una vez han tenido ocasión de enviar un telegrama habrán observado, probablemente, la calma y la indiferencia con que la empleada de telégrafos cuenta las palabras. Los telegramas contienen, por lo común, noticias muy importantes para el destinatario. Puede contener una buena noticia, pero también una gran desgracia: *Operación bien curado* y *Ven papá enfermo* son dos telegramas diametralmente opuestos desde el punto de vista de los sentimientos que provoca en el destinatario. Pero, para la empleada de Telégrafos, ambos presentan la misma situación. Puede que ni siquiera haya observado de qué se trata en ellos, ya que, en todo caso, sólo le interesa el hecho de que ambos contienen tres palabras y que, por tanto, deben ser igualmente tarifados. Todo telegrama contiene, pues, una cierta información, pero a la funcionaria no le intere-

sa su contenido, sino tan sólo el número de palabras, es decir el aspecto cuantitativo de la información.

La persona que remite el telegrama renuncia a las palabras que no son esenciales para que el destinatario comprenda la información. Así, p. ej., el texto *Operación bien curado* es una forma abreviada del texto *La operación ha terminado bien; X está curado*. Este segundo texto, aunque más largo, no contiene una información más rica o más completa que el texto abreviado, puesto que no cabe duda que el destinatario sabe bien a qué se refiere el remitente. Una tendencia natural a la economía nos determina a preferir el texto abreviado. Esta abreviación es generalmente posible debido a que, en casi todo tipo de texto, se manifiesta un cierto derroche, una cierta prodigalidad de símbolos.

Otro ejemplo significativo en lo que atañe al derroche de símbolos en un texto lo constituyen los distintos sistemas de estenografía. En la estenografía se renuncia muchas veces a notar las vocales, puesto que las consonantes solas permiten identificar ya las palabras. Las eventuales ambigüedades se resuelven empleando el contexto, es decir los términos vecinos. ¿A qué se debe, entonces, tanto derroche en la lengua? ¿Acaso los hombres son seres tan poco prácticos que no se han dado cuenta de ello y continúan malgastando la energía complicando lo que podría decirse escueta y simplemente? Desde luego, no puede negarse que en el modo de expresarse de muchos hay una cierta falta de concisión. Así, los llamados “pleonasmos” representan justamente una de las formas de despilfarro que puede y debe evitarse. Sin embargo, es no sólo inevitable una cierta cantidad de derroche, sino incluso útil. Si, p. ej., hablamos con alguien por teléfono y, en un momento dado, el paso de un autobús nos impide entender una palabra, hay muchas probabilidades de que esta palabra sea identificada a partir de su contexto o, en caso de que no ocurra así, es también probable que ello no impida la comprensión en conjunto de lo que se nos transmite desde la otra parte del hilo. Tales situaciones no sólo se producen en una conversación telefónica, sino también en la discusión directa. La posibilidad que tenemos de malgastar palabras o de repetir según convenga la comunicación asegura, pese a las eventuales perturbaciones, la comprensión por parte del interlocutor.

Como consecuencia de ello, en la transmisión de una información intervienen dos tendencias opuestas, ambas igualmente necesarias: una de economía y otra de derroche. La cuestión consiste en ensamblarlas y, por tanto, en realizar la mayor economía de símbolos que garantice la integridad de la transmisión informativa ante diversas perturbaciones y la posibilidad de comprensión para el destinatario.

8.7. Cómo nos ayuda la “teoría de la información”

La teoría de la información manifiesta, frente al contenido de un texto, la misma indiferencia que mostraba la empleada de Telégrafos del párrafo anterior. En

efecto, la teoría de la información considera el texto no desde el punto de vista cualitativo, o sea en cuanto al contenido, sino desde el punto de vista cuantitativo. Sin embargo, esta teoría aprecia el aspecto cuantitativo de una información de un modo totalmente distinto con respecto a nuestra empleada. Para ésta, lo que importa es el número de palabras, es decir la cantidad de medios materiales necesarios para la transmisión de la información. Y, como hemos observado antes, la información de *La operación ha terminado bien; X está curado* no es en absoluto ni más rica ni más completa que la de *Operación bien curado*, a pesar de que contenga más palabras [y resulte proporcionalmente más cara en Telégrafos]. Por este motivo, la teoría de la información adopta otro criterio, que consiste en medir la cantidad de información de un texto, pero entendiendo la información como un medio de precisar una determinada variante entre dos o más variantes posibles. Cuanto mayor es el grado de indeterminación que elimina una información dada, tanto mayor es la cantidad de información suministrada. Así, p. ej., el hecho de precisar que una letra dada es una vocal o una consonante representa una cantidad de información inferior al hecho de indicar cuál es exactamente la letra en cuestión. En el primer caso, se ha precisado una variante entre dos variantes posibles; en el segundo, se ha precisado una variante entre 27 variantes posibles. Por consiguiente, en el primer caso se ha eliminado una indeterminación menor que en el segundo.

8.8. Cómo se mide la información

A primera vista, parece que la información debe medirse contando las variantes posibles. Un ejemplo nos demostrará, sin embargo, que es más juicioso definir la cantidad de información no a partir del número de variantes posibles, sino a partir del logaritmo de base 2 de este número.

Supongamos que hemos de buscar, en una guía telefónica de 512 páginas, una persona que se apellida Marino. ¿Qué cantidad de información representa el hecho de indicar la página en que se encuentra este apellido? Como se trata de una opción entre 512 variantes posibles, nos vemos tentados a suponer que la cantidad de información suministrada equivale a 512. Pero es fácil advertir que esto constituiría una exageración bastante grosera sobre la supuesta cantidad informativa. Ciertas páginas, como las del principio y del final de la guía, quedan directamente eliminadas, ya que un apellido que empiece por *M* sólo puede figurar en la parte central de la guía.

Procederemos, entonces, de otro modo: abrimos el libro por la mitad, es decir por la página 256. Atendiendo a la palabra con que empieza esta página, averiguamos en qué mitad del libro se encuentra el apellido buscado. Supongamos que se halla en la segunda mitad. Dividimos de nuevo la segunda mitad y abrimos por la página 384. Igual que antes, la palabra que está al principio de esta

página nos permite determinar en cuál de las dos cuartas partes se halla la palabra buscada. Supongamos que está en la primera de ellas. Dividiendo esta cuarta parte en dos y continuando así llegamos, después de nueve operaciones a lo sumo, a identificar la página en que se halla el apellido buscado. La cifra 9, que refleja el número de operaciones, no es más que el número que indica cuántas veces puede dividirse sucesivamente por la mitad el número 512, es decir la potencia a la que debe elevarse 2 para obtener 512; en otras palabras: $\log_2 512 = 9$. Vemos, pues, que no es el número 512, sino su logaritmo de base 2 lo que nos indica fielmente la cantidad de información suministrada al precisar la página en que se encuentra el apellido en cuestión.

Vamos a considerar otro ejemplo que examinaremos más en detalle.

Supongamos que tenemos un libro de 128 páginas. ¿Cuál es la información suministrada al precisar una de las páginas del libro? (Por ejemplo, decimos a alguien: “En la página ... encontrarás la espléndida descripción de un ocaso”). Podríamos apreciar esta información a base del número 128, pero tal procedimiento ofrece demasiadas dificultades ya que lleva a números demasiado altos —sobre todo con un libro voluminoso— y no deja ver la unidad de medida empleada. Por éste y otros motivos, que presentaremos más adelante, el matemático e ingeniero americano C. E. Shannon inventó otro procedimiento, proponiendo como unidad de medida la información producida al precisar una variante entre dos variantes igualmente probables¹⁶. Se trata, desde luego, de una información elemental, la forma más simple bajo la cual puede concebirse la información, aunque, en este “estado puro”, se encuentre raras veces. La información suministrada al precisar una entre las dos caras de una moneda corresponde a este tipo de información-unidad. La información que aparece en la respuesta —positiva o negativa— a la pregunta de si el número que sale al tirar un dado (perfectamente cúbico y homogéneo) es menor de 4 constituye igualmente una información-unidad.

Esta unidad de medida para la información ha recibido el nombre de *bit*¹⁷.

Si bien Shannon puso las bases de la teoría de la información, tampoco podemos omitir el nombre de un importante precursor suyo, R. V. Hartley, que ya en 1928 introducía la noción de “cantidad de información”.

16. Cf. SHANNON, *Mathematical Theory*.

17. El origen de esta palabra es el siguiente: la indicación de una entre las cifras 0 y 1, supuestamente equiprobables, es, evidentemente, una información-unidad. Y como las cifras 0 y 1 sirven para representar cualquier cantidad en base 2, es decir en forma de

$$\frac{a_1}{2} + \frac{a_2}{2^2} + \dots + \frac{a_n}{2^n} + \dots,$$

donde los numeradores $a_1, a_2, \dots, a_n \dots$ son iguales a 0 ó a 1, las cifras 0 y 1 han recibido el nombre de “cifras binarias”, en inglés *binary digits*. La palabra *bit* está formada por las dos primeras letras de *binary* y la última de *digit*.

El problema que se plantea ahora consiste en descomponer una información compleja en un número cualquiera de bits. Naturalmente, sucede a menudo que la información considerada no es reductible a un múltiplo entero de bits. Para ilustrar más claramente el sentido del procedimiento, soslayaremos, de momento, esta dificultad que, por lo demás, resultará fácilmente superable en una etapa ulterior.

Tomemos de nuevo el problema. Tenemos un libro de 128 páginas. ¿Cuál es la información suministrada al precisar una sola de sus páginas?

Después de adoptar la unidad de medida (el bit), nuestra pregunta se formulará así: ¿Cuántos bits tiene la información producida al precisar una página de las 128 que tiene el libro? La cuestión consiste, pues, en descomponer la información considerada en una cadena de opciones entre dos variantes equiprobables.

Supongamos que la página buscada es la 58 y que un interlocutor nuestro quiere identificarla preguntándonos. Surgirá entonces el siguiente diálogo:

- La página buscada, ¿es anterior a la página 65?
- Sí.
- ¿Es también anterior a la página 33?
- No.
- ¿Está antes de la 49?
- No.
- ¿Está antes de la 57?
- No.
- ¿Y antes de la 61?
- Sí.
- La página buscada, ¿es anterior a la página 59?
- Sí.
- ¿Es también anterior a la página 58?
- No.

De esta forma, a partir de las respuestas “sí” o “no”, nuestro interlocutor ha alcanzado la información deseada en siete preguntas: la página buscada es la 58. Examinemos ahora un poco más de cerca el mecanismo por el que se obtiene esta información.

Por la respuesta a la primera pregunta, suministramos al interlocutor una información de un bit (la página buscada puede ser anterior o no a la 65 y, por tanto, existen dos variantes equiprobables). En efecto, resulta que $128 = 64 \times 2$; así que la probabilidad de que la página elegida esté entre las primeras 64 páginas del libro es igual a la probabilidad de que se encuentre en la segunda mitad del libro. Después de recibir la respuesta afirmativa a la primera pregunta, el interlocutor sabe que la página buscada se encuentra entre las primeras 64 pá-

ginas del libro. Y como ocurre que $64 = 32 \times 2$, resulta que por la respuesta que le damos a la segunda pregunta, el interlocutor recibe una nueva información de un bit. Continuando del mismo modo y observando que $16 = 8 \times 2$, $8 = 4 \times 2$, $4 = 2 \times 2$, $2 = 1 \times 2$, comprobamos que, en total, el interlocutor recibe una información de siete bits, ya que a cada pregunta la respuesta obtenida representa una información de un bit. Por otra parte, una vez recibida la respuesta a la séptima pregunta, el interlocutor está ya plenamente capacitado para identificar la página 58. En efecto, después de la respuesta afirmativa a la sexta pregunta, sabe que la página buscada es anterior a la 59, pero no anterior a la 57. Falta por decidir si la página buscada es la 57 o la 58, duda que queda resuelta precisamente por la respuesta a la última pregunta.

Por consiguiente, la información adquirida al precisar una de las 128 páginas de un libro equivale a siete bits. La proveniencia del número 7 es fácil de identificar. En cada etapa, el número de páginas restantes ha sido dividido por dos; como resulta que $128 = 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 2^7$, se han necesitado siete etapas. He aquí un aspecto interesante de la solución obtenida. El hecho de que la información considerada sea igual a siete se refiere al hecho de que elevando 2 a la potencia 7 se obtenga justamente el número de las 128 páginas del libro. Ahora bien, por definición, la potencia a que debe elevarse 2 para obtener un número n es igual al logaritmo de base 2 de n . Así, 7 es el logaritmo de base 2 de 128. Generalizando este razonamiento, se llega entonces a la siguiente regla:

La información que se obtiene al precisar una variante entre n variantes igualmente probables es igual al logaritmo de base 2 de n .

Notando esta información por medio de H_0 , podemos escribir

$$H_0 = \log_2 n.$$

Por **probabilidad** de una variante entendemos, aquí, lo que de hecho constituye la frecuencia relativa de esta variante, es decir la relación entre el número de apariciones de la variante y el número de todas las apariciones posibles. Por ejemplo, la probabilidad de que, al lanzar un dado perfectamente cúbico y homogéneo, aparezca un número par es igual a $\frac{3}{6} = \frac{1}{2}$. La previsibilidad total es igual a 1, y la imposibilidad corresponde a la probabilidad 0.

Y como la probabilidad p de aparición de una de las n variantes equivale, evidentemente, a $\frac{1}{n}$, y como $\log_2 \frac{1}{n} = \log_2 1 - \log_2 n = 0 - \log_2 n = -\log_2 n$, podemos escribir

$$H_0 = -\log_2 p. \quad (1)$$

En la experiencia anterior, sobre el libro, teníamos $H_0 = 7$.

8.9. Concepto de entropía

Tratemos de captar el significado general de la fórmula (1). Se ha considerado una experiencia con n resultados posibles y se ha supuesto que cada uno de estos n resultados tienen la misma probabilidad de aparición. En estas condiciones, la relación (1) afirma el siguiente hecho: la cantidad de información que corresponde, en promedio, a un resultado de la experiencia es igual, en bits, a $-\log_2 p$.

En el ejemplo anterior, del libro con 128 páginas, hemos tenido precisamente una experiencia del tipo descrito, puesto que los 128 resultados posibles de allí eran equiprobables.

Por lo común, en la mayoría de las experiencias reales, no suele satisfacerse la condición de la equiprobabilidad. En rumano, la consonante *s* aparece más veces que la consonante *h*, el nominativo más veces que el vocativo, el indicativo más veces que el subjuntivo...

El examen de la frecuencia de las palabras en diez obras representativas de la literatura rumana ha demostrado que, en 50 000 palabras-texto, las más frecuentes son *de*, *și*, *el*, *avea* (1887, 1871, 1712 y 1608 veces, respectivamente)¹⁸. El rango más bajo corresponde, como era de esperar, a las partículas no flexivas del habla, verbos auxiliares, pronombres; el primer nombre (*copil*, 'niño') no aparece hasta el rango 42 (atestiguado 105 veces), y el siguiente (*zi*, 'día') hasta el 46 (99 veces).

Por lo tanto, en las experiencias con carácter lingüístico apenas ocurre que los resultados sean igualmente probables. En estas condiciones, la fórmula (1) ofrece una aproximación poco atinada de la cantidad de información. Hay necesidad de una aproximación más eficiente, que tenga en cuenta las diferencias de probabilidad para los posibles resultados. He aquí cómo se consigue esta mejora.

Sean $r_1, r_2, \dots, r_i, \dots, r_n$ los n resultados posibles en una experiencia dada. Sea p_i la probabilidad de que la experiencia dé el resultado r_i ($i = 1, 2, \dots, n$). Según la fórmula (1), la cantidad de información obtenida en la hipótesis de que todos los resultados tengan la probabilidad p_i es

$$-\log_2 p_i.$$

Introduzcamos ahora la noción de **m e d i a p o n d e r a d a**. Dados n números, a_1, a_2, \dots, a_n , y otros n números, p_1, p_2, \dots, p_n , la media ponderada

18. [*De*, 'de'; *și*, 'y, también'; *el*, 'él'; *avea*, 'tener (haber)']. Cf. ȘUTEU, "Observații asupra frecvenței". [Los datos que aquí se mencionan no se corresponden plenamente con los de JULLIAND-EDWARDS-JULLIAND, *Frequency Dictionary*, a causa de las diferencias de criterio en la delimitación de las palabras y de su campo significativo y funcional. Cf., para esta problemática, KOCK, *Introducción*].

de los números a_1, a_2, \dots, a_n con las ponderaciones p_1, p_2, \dots, p_n es, por definición, la expresión

$$\frac{p_1 a_1 + p_2 a_2 + \dots + p_n a_n}{p_1 + p_2 + \dots + p_n}$$

Si, p. ej., un estudiante obtiene la nota 7 en un examen escrito y la nota 10 en un examen oral y se atribuye al examen escrito una ponderación doble respecto del oral, entonces la nota media que recibirá el estudiante será

$$\frac{2 \cdot 7 + 1 \cdot 10}{2 + 1} = 8$$

De modo que la media ponderada de los números 7 y 10, con las ponderaciones 2 y 1, es 8.

Pues bien, consideraremos, entonces, la media ponderada de las n cantidades de información obtenidas según la fórmula (1):

$$-\log_2 p_1, -\log_2 p_2, \dots, -\log_2 p_i, \dots, -\log_2 p_n.$$

Ello significa que se considera la expresión

$$\frac{p_1(-\log_2 p_1) + p_2(-\log_2 p_2) + \dots + p_i(-\log_2 p_i) + \dots + p_n(-\log_2 p_n)}{p_1 + p_2 + \dots + p_n} \quad (2)$$

Debido a que la experiencia tendrá necesariamente uno de los resultados $r_1, r_2, \dots, r_i, \dots, r_n$, cualquiera que sea su evolución, y como estos resultados son incompatibles de dos en dos, ocurrirá que:

$$p_1 + p_2 + \dots + p_i + \dots + p_n = 1,$$

de donde, notando la expresión (2) por medio de H_1 , resulta que

$$H_1 = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_i \log_2 p_i - \dots - p_n \log_2 p_n.$$

H_1 constituye una mejor apreciación de la cantidad de información que H_0 . En el caso particular en que $p_1 = p_2 = \dots = p_i = \dots = p_n$, donde p indica el valor común de las n probabilidades, obtenemos

$$H_1 = \underbrace{-p \log_2 p - \dots - p \log_2 p}_{n \text{ veces}}$$

y, como $np = 1$, resulta que $H_1 = H_0$, como era de esperar. Pero, en general, ocurre que, como se demuestra fácilmente,

$$H_0 \geq H_1,$$

siendo la igualdad posible sólo si $p_1 = p_2 = \dots = p_n$.

La extensión H_0 se denomina también entropía de orden 0 de la experiencia considerada, mientras que la extensión H_1 se llama además entropía de primer orden¹⁹. El término "entropía" es un préstamo de la termodinámica. Esta noción tiene en común, en el marco de la información y de la termodinámica, el hecho de que mide el grado de indeterminación de un fenómeno. Por lo demás, también su expresión matemática es semejante²⁰.

La entropía de primer orden ofrece un gran interés lingüístico, pero su empleo es todavía precario debido a la inexistencia de trabajos preliminares. En la expresión de la entropía de primer orden intervienen las probabilidades de los distintos resultados de la experiencia considerada, y son justamente estas probabilidades lo que permanece desconocido para la mayor parte de las experiencias con hechos de lengua. Supongamos, p. ej., que queremos determinar la entropía de primer orden del sistema casual rumano. Aquí, los resultados posibles de la experiencia son los cinco casos del nombre rumano [(nominativo, acusativo, genitivo, dativo, vocativo)]. Como aún no disponemos de una tal investigación sobre esta parte de la gramática rumana, no se conoce la probabilidad de aparición textual de los casos²¹. Ni siquiera disponemos de un estudio estadístico sobre las partes del discurso del rumano y, por ello, tampoco podemos establecer aún su entropía²².

El cálculo de la entropía lingüística se halla en estrecha relación con el estado de las investigaciones puramente lingüísticas. Ya que nos hemos referido a la estadística sobre las partes del discurso, vale la pena recordar ahora cuán débil y vulnerable es la llamada clasificación tradicional de las palabras, en la que se confunden criterios semánticos, lógicos y morfológicos. Por la práctica del análisis gramatical, se sabe que a menudo la misma palabra puede considerarse adjetivo o adverbio, pronombre o adjetivo, numeral o pronombre, verbo en forma impersonal o nombre (infinitivo) o adjetivo (participio); se discute controvertidamente sobre si el artículo es una parte del discurso o un morfema y sobre si el numeral, con sus diversas especies, no debería ser "borrado" como parte del discurso.

¿Qué valor podría tener una estadística montada sobre un material tan equivoco? La entropía calculada para unas categorías imprecisas tendrá inevitablemente un valor científico bastante reducido.

Si el fonema fuese considerado, como en tiempos de Baudoin de Courtenay, una realidad más psicológica que objetiva, difícilmente evaluaríamos la entropía

19. A veces, el término "entropía" se hace sinónimo de "cantidad de información". En rigor, sólo sus valores absolutos coinciden, pero la cantidad de información es positiva, mientras que la entropía es negativa.

20. Para el concepto de entropía en termodinámica, cf. GAMOW. *Unu, doi, trei*.

21. Para el ruso, cabe citar la estadística de NIKONOV. "Statistika padzhei".

22. Algunas indagaciones sobre las partes del discurso en francés se encuentran en GUIRAUD. *Caractères*.

pía fonológica de una lengua²³. De ahí la gran importancia que revisten, para la aplicación de la teoría de la información en lingüística, las investigaciones realizadas sobre bases posicionales, el análisis distribucional o contextual, tanto en fonología como en gramática. La modelación matemática de la lengua crea condiciones óptimas para esta aplicación.

8.10. Influencia a distancia

Lo indicado en el párrafo anterior todavía no resuelve el problema de evaluar la entropía en la lengua. Las unidades lingüísticas (fonemas, morfemas, palabras, etcétera) no sólo poseen una determinada probabilidad de aparición, sino también una determinada capacidad para combinarse entre sí. En rumano, p. ej., el fonema /a/ va más veces seguido de /r/ que de /u/. La semivocal [ɣ] (como en *greacă*, 'griega', *leagăn*, 'cuna', *vreodată*, 'alguna vez', etc.) sólo se combina con una de las vocales *a*, *o*, y siempre se encuentra precedida de una consonante. Un morfema de tiempo se combina con una base verbal, pero nunca con una base nominal. Se plantea así el problema de evaluar, a partir del establecimiento de las probabilidades de las combinaciones entre dos unidades, la entropía H_2 de segundo orden. Este problema se ha resuelto en el marco de la teoría de la información²⁴.

Intuitivamente, esperamos obtener $H_1 \geq H_2$; este hecho puede demostrarse de un modo riguroso.

Existen estadísticas sobre combinaciones de dos letras, pero, por desgracia, todavía no disponemos de trabajos análogos sobre combinaciones de dos fonemas, morfemas o sintagmas²⁵.

Teóricamente, podemos continuar definiendo las entropías de tercer, cuarto, etc. orden teniendo en cuenta las probabilidades de aparición de grupos de tres, cuatro, etc. unidades de un determinado nivel. Notando estas entropías por medio de H_3 , H_4 , etc., se comprueba que tienen lugar las desigualdades

$$H_0 \geq H_1 \geq H_2 \geq H_3 \geq \dots \geq H_n \geq H_{n+1} \geq \dots, \quad (3)$$

lo que es natural, puesto que, a medida que se toman en consideración proba-

23. Cf. GRUMĂZESCU. "Frecvența sunetelor", sobre la frecuencia estadística de los fonemas en rumano. [Para el español, cf. NAVARRO, *Estudios*, y ALARCOS, *Fonología*].

24. Para los detalles, cf. IAGLOM-DOBRUSHIN-IAGLOM. "Teoria informației" (traducción del ruso *Verozhatnost' i informatsia*), donde se enriquece, respecto del original, la parte relativa a las aplicaciones de la teoría de la información en la lingüística y se introducen, por primera vez, los experimentos de predicción de Kolmogorov. Cf. también IAGLOM-IAGLOM, *Probabilitate*, y NOVAK-PIOTROVSKY, "Experimental".

25. Una estadística para la combinación de dos letras entre sí, en textos rumanos literarios y periodísticos, se halla en NICOLAU-SALA-ROGERIC, "Observații".

bilidades de aparición de combinaciones cada vez mayores de unidades, la indeterminación disminuye²⁶.

Sin embargo, se plantea el problema de si ofrece interés, desde el punto de vista lingüístico, la consideración de estas entropías de orden superior. La respuesta no sólo es afirmativa, sino que además ilustra de una manera muy convincente la profunda significación que tiene la teoría de la información para la lingüística. Veamos de qué se trata. Un problema muy natural en lingüística consiste en establecer la influencia a distancia de las unidades lingüísticas entre sí. Si al principio de un libro encontramos la palabra *fonema*, es muy poco probable que al final del mismo libro aparezca la palabra *fermentación* y, en cambio, es muy probable encontrar la palabra *neutralización*. Hemos elegido intencionadamente un caso extremo para demostrar que la aparición de una palabra en un texto influye sobre la aparición incluso de las palabras más alejadas de ella en el texto respectivo. Cabe también dar ejemplos de significado inverso, que sugieran que, en ciertas condiciones, una palabra dada no ejerce ninguna influencia sobre la aparición de la palabra que sigue inmediatamente. El problema planteado consiste en establecer cuál es la influencia media a distancia entre las palabras de un texto. Pero, ¿qué significado preciso tiene esto? La teoría de la información nos proporciona la respuesta. La influencia a distancia entre las palabras de un texto viene dada por aquel número n , empezando por el cual, en la cadena (3), ocurre que $H_n = H_{n+1} = H_{n+2} = \dots$, donde cada entropía se refiere a las apariciones de palabras. Ahora bien, ¿cuál es el valor de n ? He aquí un problema aún no resuelto para ninguna lengua. En ello hay dificultades tanto de índole lingüística, debidas al gran volumen del vocabulario de cualquier lengua, como de índole técnica, relativas a las complicaciones de cálculo que intervienen en la evaluación de las entropías de orden superior.

El problema de la influencia a distancia ha recibido una solución casi completa en lo que atañe a la entropía sobre letras, aunque en este caso la intuición nos ayuda menos que en el de las palabras y, por tanto, puede inducirnos a error más fácilmente. Las investigaciones de Burton y Licklider para el inglés han permitido establecer el valor de 30. Esto significa que $H_{29} > H_{30} = H_{31} = H_{32} = \dots$, o, en otras palabras, que, en general, la aparición de una letra influye sobre la aparición de las siguientes 30 letras, pero que la influencia se extingue a partir de la letra 31-ava. También significa que la aparición de una letra está influida, por término medio, por las 30 letras inmediatamente anteriores²⁷.

I. M. Iaglom y A. M. Iaglom han emitido la hipótesis de que este valor de 30 resulta también válido para las demás lenguas. El valor 30 parece, desde luego, sorprendente; la intuición nos sugiere que la influencia de una letra tiene un radio de acción mucho menor. Aquí se valoriza íntegramente la eficacia de

26. Cf. IAGLOM-DOBRUSHIN-IAGLOM, "Teoria informației", [y SINGH, *Great Ideas*].

27. Cf. BURTON-LICKLIDER, "Long-range constraints".

la teoría de la información y la fineza con que registra los fenómenos que escapan al análisis directo y empírico.

Pero las letras no son lo que más interés ofrece para el lingüista. No es difícil mostrar ejemplos en que se revele que la aparición de una palabra ejerce una influencia a distancia. Precisamente este hecho ha sido considerado por N. Chomsky, al definir el concepto de "(i, j)-dependencia". En la secuencia *Si te portas bien y comes todo lo que te he dado, entonces recibirás un regalo*, la aparición de la palabra *si* convierte en probable, a pesar de la apreciable distancia intermedia, la aparición de la palabra *entonces*²⁸. En alemán, la presencia de *gefunden* al final de una oración resulta de la acción a distancia de un verbo como *ich habe, er hat* u otro del mismo tipo. Del mismo modo, en ruso la aparición de *esli*, 'si', determina la aparición —incluso a gran distancia— de *to*, 'entonces', en inglés la aparición de *either* determina la aparición de *or*, y la aparición de *if*, la de *then*.

Estos ejemplos ilustran la influencia a distancia entre palabras. En lo que se refiere a la influencia a distancia entre morfemas gramaticales, pueden aducirse fenómenos tales como la concordancia y la rección. En la secuencia de morfemas rumanos *fat-ă bogat-ă* ('muchacha rica') el segundo morfema determina la aparición del cuarto, y en la secuencia *uș-a cas-ei* ('la puerta de la casa'), el morfema de genitivo *ei*, aunque no es el único posible en esta posición, no puede ser sustituido por cualquier otro de la flexión de *casă*; p. ej., no puede aparecer, en su lugar, *a* [**ușa casa, *la puerta la casa*']. También aquí se manifiesta la influencia del segundo morfema en la secuencia considerada. El último ejemplo es característico para la mayoría de las influencias a distancia, donde dicha influencia no se manifiesta tanto por la determinación forzosa de un morfema dado, como por las restricciones en la facultad de elegir, es decir por la disminución del número de morfemas que pueden aparecer en una posición determinada (cf. 8.11).

Se plantea ahora el problema de cómo medir esta acción a distancia entre los morfemas. ¿Cómo comparar, entonces, dos fenómenos de influencia a distancia de modo que quepa establecer cuál de ellos resulta más pronunciado? Esta clase de problemas son extremadamente importantes por cuanto esta "acción a distancia" comprende, como casos particulares, tanto el fenómeno de (i, j)-dependencia considerado por N. Chomsky como el fenómeno de determinación heterosintagmática estudiado por L. Hjelmslev y considerado por él fundamental para la distinción entre morfemas y pleremas o, más en general, entre exponentes y constituyentes.

En cuanto a la apreciación global, en promedio, de la influencia a distancia que se produce entre los morfemas de una lengua, el problema se resuelve estableciendo la entropía morfemática de orden superior en la lengua en cuestión.

28. [Este hecho es bastante menos probable en español que en rumano].

Para cada morfema y para cada combinación de morfemas se averiguará, de antemano, la probabilidad de aparición y se aplicará luego la fórmula de la entropía de orden superior. Si, p. ej., calculamos las probabilidades de aparición de combinaciones formadas a lo sumo por diez morfemas, podremos establecer la entropía morfemática de décimo orden, obteniendo así una medida de la influencia a distancia entre los morfemas de la lengua respectiva y haciendo abstracción de las acciones que se ejercitan a una distancia superior a los diez morfemas. Pero aparte de una apreciación global de este tipo sobre la acción a distancia, interesa también la apreciación local, que proporcione la posibilidad de comparar entre sí dos secuencias morfemáticas dadas. Un problema semejante se plantea para las secuencias de fonemas. Intuitivamente es muy difícil, si no imposible, decir en cuál de las secuencias en rumano *casa mare* ('la casa grande') y *casa mică* ('la casa pequeña') es más fuerte la acción a distancia de los fonemas. Basándonos en las probabilidades de aparición de los grupos de fonemas *casa ma*, *casa mi*, *casa mar*, *casa mic*, *casa mare*, *casa mică*, podemos establecer la entropía de cada una de las secuencias consideradas. Así, la entropía del paso de *casa m* a *casa ma* o a *casa mi* viene dada por la expresión:

$$-p(1)\log_2 p(1) - p(2)\log_2 p(2) - \dots - p(i)\log_2(i) - \dots - p(n)\log_2 p(n),$$

donde $p(i)$ indica la probabilidad de que el fenómeno de rango i sea precedido por el grupo *cas m*; aquí, i toma los valores de 1 a n , siendo n el número de fonemas del rumano. El establecimiento de estas probabilidades es bastante difícil. Asimismo, el empleo de las fórmulas que proporcionan las entropías de orden superior conduce, del tercer orden para arriba, a cálculos muy laboriosos. He aquí los valores de las entropías calculados para algunas lenguas europeas:

Lengua	Entropía			
	H_0	H_1	H_2	H_3
Ruso hablado	5,38	4,76	3,69	0,70
Ruso escrito	5	4,35	3,52	3,01
Inglés escrito	4,76	4,03	3,32	3,10
Alemán escrito	4,70	4,10		
Francés escrito	4,70	3,98		
Rumano escrito	4,70	4,11		
Español escrito	4,70	4,01		

Cuadro 25

Para comprender el significado de los valores allegados, precisaremos algunas aclaraciones. La entropía se ha calculado sobre las letras para la lengua escrita y sobre los fonemas para la hablada. El alfabeto del ruso se ha considerado compuesto por 32 letras (se ha asimilado \ddot{e} a e y \mathfrak{b} a b ; el espacio en blanco entre palabras se ha contado como una letra aparte). En inglés, el espacio en blanco se ha asimilado a una letra aparte; en cambio, en alemán, francés, rumano y español no se ha procedido así, de modo que el alfabeto de estas lenguas ha quedado evaluado en 26 letras. La entropía del ruso hablado fue establecida en un artículo de C. Cherry, M. Halle y R. Jakobson²⁹. Hay que hacer notar, sin embargo, que el valor que estos autores conceden a H_3 es más bien dudoso, ya que el volumen textual que analizaron resulta insuficiente para determinar las probabilidades combinatorias de tres fonemas³⁰.

Para comprender de un modo práctico la diferencia entre las entropías (calculadas sobre letras) de las distintas lenguas, podemos realizar diversas experiencias simples: numeramos las letras de un texto redactado en la lengua A y las letras de la traducción del mismo texto en la lengua B . Aunque las lenguas A y B utilicen el mismo alfabeto y el contenido del texto sea el mismo, ni el número de letras resultará idéntico ni la frecuencia de cada letra tomada aisladamente será la misma.

Tomemos, p. ej., las siguientes frases rumanas y sus correspondientes traducciones al alemán [y al español]:

A) Din exemplele de mai sus, s-a observat, desigur, că declinarea substantivelor fără articol sau cu articol nehotărît e mai simplă decît cea cu articol hotărît. Ca să declinăm un substantiv nearticulat, indiferent de gen, e deci suficient să cunoaştem cuvîntul respectiv la singular şi la plural.

B) Bei den obigen Beispielen haben Sie gewiss bemerkt, dass die Beugung der Hauptwörter ohne Artikel oder mit dem unbestimmten Artikel einfacher ist als die mit dem bestimmten Artikel. Um ein Hauptwort, gleichgültig welchen Geschlechts, ohne Artikel zu beugen, genügt es also, das betreffende Wort in der Einzahl und in der Mehrzahl zu kennen.

[C) Por los ejemplos de antes, se ha observado, sin duda, que la declinación de los nombres sin artículo o con artículo indeterminado es más simple que la declinación con artículo determinado. Para declinar un nombre sin artículo, indistintamente del género, basta, pues, conocer la palabra respectiva en singular y en plural.]

El texto rumano contiene 294 letras; su traducción al alemán, 339 y al español 321 (hemos contado tanto los signos de puntuación como los espacios en blanco entre palabras). Repitiendo este tipo de experiencias, comprobamos que, por término medio, las versiones alemana y española de un texto rumano

29. Cf. CHERRY-HALLE-JAKOBSON, "Logical description".

30. Para el establecimiento de las demás entropías, pueden hallarse referencias más detalladas en el artículo de IAGLOM-DOBRUSHIN-IAGLOM, "Teoria informaţiei" y en el libro de AJMANOVA-MELCHUK-PADUCHEVA-FRUMKINA. *O tochinij metodaj*. Para el rumano, los datos se han tomado de NICOLAU-SALA-ROGERIC, "Observaţii", donde se establece el valor de H_1 a partir de textos periodísticos.

tienen más letras, son más largas. Puede demostrarse asimismo que, procediendo a la inversa, es decir considerando la versión rumana de un texto alemán o español, aparece el mismo desfase: el texto rumano será más corto³¹.

8.11. Determinación experimental de la entropía lingüística

Como hemos visto, para poder utilizar las fórmulas que determinan la entropía de orden superior es necesario establecer de antemano las probabilidades de las distintas combinaciones de letras, fonemas, morfemas, etcétera. El número de estas combinaciones es, sin embargo, muy grande.

En un trabajo sobre la entropía del inglés, Shannon propone un método para evaluar la entropía de las letras que evita las probabilidades combinatorias de las mismas³². En cambio, es indispensable establecer las probabilidades $q_i^{(N)}$ para que la N -ésima letra del texto quede correctamente indicada en la i -ésima prueba, suponiendo conocidas las $N - 1$ letras precedentes. La ventaja de este método consiste en que las probabilidades $q_i^{(N)}$ pueden determinarse experimentalmente. En efecto, todo buen conocedor de la lengua española posee una idea intuitiva sobre estas probabilidades, hecho que queda bien ilustrado mediante experimentos de adivinanza del siguiente tipo:

Se toma un texto y se le da al interlocutor la posibilidad de conocer las primeras $N - 1$ letras del mismo. Se le pide entonces que adivine la letra N -ésima y se repite la prueba hasta obtener la solución correcta; luego, partiendo del conocimiento de las primeras N letras, se le pide que adivine la letra $N + 1$ -ésima y se anota, como antes, el número de intentos necesarios hasta conseguir la solución, y así sucesivamente.

Para el valor $N = 1$, es decir al principio del experimento, la indeterminación es máxima y, por tanto, la oportunidad de adivinar la primera letra del

31. [En la traducción italiana de esta obra (p. 330), C. TAGLIAVINI halla en su versión 327 letras (en el original, se atribuyen al rumano 282 letras y al alemán 328, tal vez (?) porque los textos se han presentado en dos párrafos: nuestro cómputo aprecia en la versión italiana 335 letras). Unificando los criterios, se obtiene, pues:

alemán: 339
italiano: 335
español: 321
rumano: 294

En estas diferencias concurren, entre otros, factores tales como la ortografía (más copiosa en alemán), la tipología gramatical, sobre todo del nombre (cf. el ejemplo de antes, *usa casei* [2 palabras, 9 letras], frente a *la puertia de la casa* [5 palabras, 20 letras]) y la norma lingüística (cf. *e suficient să cunoaștem* frente a *basta conocer*, en lugar de *es suficiente que conozcamos*; pero, en cambio, *declinarea... e mai simplă decât cea cu articol* frente a *la declinación... es más simple que la declinación con artículo*, en lugar de **...que la con artículo*).

32. Cf. SHANNON, "Prediction".

texto es muy pequeña. Ello significa que las probabilidades $q_1^{(1)}$, $q_2^{(1)}$, $q_3^{(1)}$, de adivinar la primera letra en un solo intento, en dos y, respectivamente, en tres son muy escasas. Para el valor $N = 2$, las oportunidades de acertar son sensiblemente mayores. Si, p. j., la primera letra ha sido R , entonces la segunda no puede ser más que A , E , I , O o bien U . Resulta, pues, que $q_1^{(1)} < q_1^{(2)}$, $q_2^{(1)} < q_2^{(2)}$.

A medida que avanza el experimento, el interlocutor adquiere familiaridad con una porción cada vez mayor de texto, con lo que puede valorizar, además del conocimiento de la estructura estadística del español, la información semántica del texto en cuestión. Así, si en la etapa $N = 41$, o sea en el momento en que debe adivinar la letra 41.^a del texto, las primeras 40 letras (con el espacio en blanco inclusive) se presentan como

Siendo grande la redundancia del texto, s

entonces, la probabilidad $q_1^{(41)}$ de que la letra 41.^a sea adivinada al primer intento queda aumentada por el conocimiento de la información semántica proporcionada por la secuencia de las primeras 40 letras (en la hipótesis de que el interlocutor esté en condiciones de captar esta información). En efecto, es probable que la letra 41.^a sea U , de modo que el texto discorra así:

Siendo grande la redundancia del texto, su entropía es pequeña.

A pesar de ello, no podemos afirmar que $q_1^{(41)} = 1$ y que, por tanto, la aparición de U en la posición 41 esté asegurada; podría ocurrir, p. ej., que la letra 41 fuese E , con otro desarrollo textual:

Siendo grande la redundancia del texto, se verificará que...

En todo caso, la probabilidad de acertar la letra 41.^a a los pocos intentos es muy grande y, por ello, muy cercana a 1.

Realizando un gran número de experimentos de este tipo, se obtienen aproximaciones bastante afinadas sobre los valores $q_i^{(N)}$. A partir de estos valores, se obtiene asimismo una aproximación sobre las entropías H_n de orden superior, mediante las siguientes desigualdades establecidas por Shannon:

$$\begin{aligned} & 2(q_2^{(N)} - q_3^{(N)}) \log_2 2 + 3(q_3^{(N)} - q_4^{(N)}) \log_2 3 + \dots \\ & + (n-1)(q_{n-1}^{(N)} - q_n^{(N)}) \log_2 (n-1) + n q_n^{(N)} \log_2 n \leq H_n \leq \\ & - q_1^{(N)} \log_2 q_1^{(N)} - q_2^{(N)} \log_2 q_2^{(N)} - \dots - q_n^{(N)} \log_2 q_n^{(N)}. \end{aligned}$$

El método expuesto también presenta, no obstante, algunos inconvenientes: no proporciona el valor de la entropía H_n , sino tan sólo una limitación inferior y otra superior a ella, es decir un intervalo donde se encuentra H_n . Por otra parte, si bien se puede suponer que un buen conocedor del español posee intuitivamente su estructura estadística en lo que atañe a las letras, no puede decirse lo mismo en cuanto a los fonemas y morfemas. Esta observación vale para toda

lengua y puede que por ello no se hayan efectuado con éxito tales experimentos de predicción más que para determinar la entropía sobre letras. Citemos, en este sentido, la determinación, en inglés, de la entropía H_N sobre letras llevada a cabo por Shannon para $N = 1, 2, 3, \dots, 15$ y $N = 100$, y por Burton y Licklider para $N = 1, 2, 4, 8, 16, 32, 64, 128$ y $N \approx 10\,000$ ³³. Estos últimos han utilizado un material más rico y, por tanto, sus resultados son más fiables que los de Shannon. Justamente de este modo se ha obtenido el resultado mencionado en 8.10, según el cual la cadena H_N se vuelve estacionaria empezando por el rango 30; en otras palabras, $H_{30} = H_{31} = H_{32} = \dots = H_{100} = H_{101} = \dots$

Evidentemente, a medida que se progresa en la composición, sobre bases objetivas, del inventario fonemático de las distintas lenguas, se podrá encontrar un número suficientemente grande de lingüistas para realizar, mediante un conocimiento profundo tanto de la lengua como de su estructura fonemática, experimentos de predicción de fonemas que faciliten una mejor evaluación sobre la entropía de la lengua hablada. De momento, esta suerte de experimentos ha dado sólo resultados precarios³⁴.

8.12. La entropía de las palabras y su relación con la entropía de las letras

Existe una estrecha relación entre estas dos entropías. Cada una de ellas puede establecerse a partir de la otra y aun de una manera independiente. El cálculo de la entropía de primer orden para las palabras queda muy simplificado gracias a la conocida ley de Zipf. Recordemos que, según esta ley, disponiendo las palabras en forma de lista en orden decreciente respecto de su frecuencia, la probabilidad P_n de aparición de la palabra de rango n viene dada por

$$P_n = \frac{k}{n},$$

donde k es una constante del texto considerado, que se determina del siguiente modo.

Por una parte, notando mediante m el número total de palabras, resulta que

$$p_1 + p_2 + \dots + p_m = 1,$$

o, en otros términos, que la suma de las probabilidades de aparición de las distintas palabras es igual a 1, ya que la aparición de una palabra cualquiera del vocabulario es segura.

33. Cf. SHANNON, "Prediction", y BURTON-LICKLIDER, "Long-range constraints".

34. Para algunas implicaciones psicológicas derivadas de los experimentos de predicción, cf. SLAMA-CAZA y ROCERIC, "Statística fonemelor".

Por otra parte, en conformidad con la ley de Zipf, sucede que:

$$\begin{aligned} p_1 + p_2 + \dots + p_m &= \\ &= \frac{k}{1} + \frac{k}{2} + \dots + \frac{k}{m} = k \left(1 + \frac{1}{2} + \dots + \frac{1}{m} \right), \end{aligned}$$

y, por tanto,

$$k \left(1 + \frac{1}{2} + \dots + \frac{1}{m} \right) = 1$$

de donde resulta

$$k = \frac{1}{1 + \frac{1}{2} + \dots + \frac{1}{m}}$$

Shannon determinó el valor de la constante k para el inglés y obtuvo $k = 0,1$. Digamos también que el valor de k también puede determinarse experimentalmente; en efecto, como $p_1 = \frac{k}{1} = k$, entonces k es la probabilidad de aparición de la palabra más frecuente.

Una vez conocido el valor de k , podemos determinar las probabilidades $p_1, p_2, \dots, p_n \dots$, a partir de las relaciones $\frac{k}{1}, \frac{k}{2}, \dots, \frac{k}{n}, \dots$. Sustituyendo los valores de estas probabilidades en la expresión de la entropía de primer orden, obtenemos el resultado deseado. Por este camino, Shannon estableció la entropía de primer orden de las palabras en inglés para la que dedujo un valor aproximado de 11,82 bits.

Si notamos por medio de s la extensión o longitud media de las palabras de una lengua dada, entonces, dividiendo por $s + 1$ la entropía de las palabras, extraemos la entropía de las letras o de los fonemas de la misma lengua (según que las palabras hayan recibido una consideración de secuencias de letras o de fonemas). Efectivamente, si una palabra está formada por s letras, junto con el espacio en blanco que la separa de la palabra siguiente tendrá $s + 1$ letras. A una letra le corresponde, pues, la $s + 1$ -ésima parte de la entropía que corresponde a una palabra.

Y a la inversa, conociendo la entropía de una letra, la multiplicación de esta entropía por $s + 1$ proporciona la entropía de una palabra.

8.13. La entropía de los textos poéticos

La entropía derivada de un texto poético ofrece una especial complejidad. Comprendiendo este inconveniente, A. N. Kolmogorov introdujo el concepto de

flexibilidad del texto, noción que no puede definirse todavía con precisión, pero que allega, podemos decir, un término suplementario a la expresión de la entropía. Para evaluar la flexibilidad de un texto, A. N. Kolmogorov propone el siguiente tipo de experimento. Se toma un texto de una lengua extranjera. Después de que el interlocutor trabaje con conocimiento con el mismo, se pone a su disposición una traducción de buena calidad sobre un fragmento del mismo texto y se le pide que trate de adivinar el resto de la traducción. La entropía correspondiente a un experimento de este tipo, es decir el grado de indeterminación de las sucesivas etapas del mismo, define la flexibilidad del texto³⁵.

Particularmente significativa para la complejidad de la entropía de un texto poético, es la discusión sostenida hace unos años en la revista *Limba Română* con relación al último verso del poema *Floarea albastră* de Eminescu. Aunque la controversia gira en torno a una sola palabra —más exactamente, a una parte de palabra—, ya que sólo se trata de elegir entre dos variantes, *totul* ‘todo’ y *totuși* ‘sin embargo’, se ha considerado necesario recurrir a la información poética de todo el poema, desde el primero al último verso³⁶. Lo que confirma, por lo demás, cuán grande puede ser la influencia a distancia entre las palabras de un texto.

8.14. La redundancia

La comparación de dos lenguas desde el punto de vista de sus respectivas redundancias puede encontrar dificultades a causa de las diferencias en la magnitud del alfabeto, del vocabulario o del inventario fonemático o morfemático de ambas lenguas, diferencias que determinan directamente desajustes —de hecho insignificantes— entre las entropías respectivas. Por este motivo, en lugar de la entropía normal de orden n , que hemos notado por medio de H_n y a la que concedemos la denominación de entropía absoluta de orden n , introduciremos en lo sucesivo la entropía relativa de orden n , que notaremos mediante h_n y definiremos como la relación que existe entre la entropía absoluta de orden n y la entropía absoluta de orden 0:

$$h_n = \frac{H_n}{H_0}.$$

35. Cf. ERMILOVA, “Poezia și matematica”.

36. [La última estrofa de *Floarea albastră*, ‘*La flor azul*’, dice así:

<i>Și te ai dus, dulce minune</i>	Y te has ido, dulce maravilla,
<i>Și-a murit tubirea noastră—</i>	Y ha muerto nuestro amor—
<i>Floare-albastră! floare-albastră!...</i>	¡Flor azul! ¡flor azul!...
<i>Totul (totuși) este trist în lume!</i>	¡Todo (sin embargo) está triste en el mundo!].

Cf., en especial, CUCIUREANU, “Versul enigmă”, y NICOLESCU, “Contribuții”.

De este modo queda soslayado el insignificante desajuste sobre el que hablábamos antes. Como $H_n \leq H_0$, resulta que la entropía relativa de orden n está siempre comprendida entre 0 y 1. Si $h_n = 0$, entonces $H_n = 0$ y, por tanto, $H_\rho = 0$ para $\rho \geq n$, es decir las combinaciones de por lo menos n elementos quedan perfectamente determinadas, de manera que su averiguación ya no proporciona información alguna. Si $h_n = 1$, entonces $H_n = H_0$. Entre estas dos situaciones extremas, existe una infinidad de posibilidades para h_n , correspondientes a la infinidad de valores comprendidos entre 0 y 1. Como el caso $h_n = 1$ equivale a una libertad total en el modo de combinarse los elementos, resulta que la diferencia $1 - h_n$ mide el grado de desviación a partir de la libertad combinatoria total y de la igualdad de probabilidad en la aparición de estas combinaciones. Esta diferencia se denomina *redundancia de orden n* y viene notada por la expresión r_n . La redundancia de orden n aumenta a medida que la entropía de orden n decrece y decrece a medida que la entropía de orden n aumenta. De ahí que, como $h_0 \geq h_1 \geq h_2 \geq h_3 \geq \dots \geq h_n \geq \dots$, resulta que

$$r_0 \leq r_1 \leq r_2 \leq r_3 \leq \dots \leq r_n \leq \dots,$$

y, por tanto, la redundancia de orden n aumenta paralelamente con el valor de n .

Vamos a indagar, en lo que sigue, y en una lengua construida por nosotros, cómo las necesidades comunicativas determinan la introducción de restricciones que aumenten de una manera creciente la redundancia de la lengua respectiva y que, en consecuencia, disminuyan la entropía.

Supongamos que exista una lengua (o un código) con cuatro sonidos a , b , e , c . Las palabras de esta lengua estarían formadas por un número cualquiera de sonidos, combinados de cualquier forma; habría, entonces, palabras como b , be , $cabe$, $bbeacbaechebe$, $aecbabeacbeacbecbbacca$, etcétera. Para una lengua de este tipo, la redundancia de orden n será igual a 0, cualquiera que sea el número natural n , ya que toda combinación de sonidos tiene la misma probabilidad de aparición. Inmediatamente nos convenceremos de cuán poco práctica resulta una lengua de esta naturaleza. Si toda combinación puede ser una palabra, hay que admitir que toda combinación tiene significado. Imaginemos que hemos de transmitir un mensaje con ayuda de este código y que utilizamos, como canal, el hilo telefónico. Por muy claramente que marcáramos los límites entre palabras (a base de pausas largas), la comunicación discurriría muy penosamente. En efecto, ¿cómo se captaría una proposición como $bbbcc aacbebb$? Es muy probable que, al otro extremo del cable, se oiga, en lugar de $bbbcc$, $bbbcc$ o $bbbccc$, etcétera, lo que sería molesto, puesto que las combinaciones erróneas tienen también un significado específico y distinto del que se pretende dar con la comunicación correcta. Por otra parte, a causa de la longitud indistinta que pueden tener las palabras de esta lengua, el interlocutor —hasta que no las identifique y reconozca, es decir hasta que no entienda— debe escuchar con

el ánimo tenso para que no se le escape ningún sonido. Un solo sonido pasado por alto promueve otra palabra y, por consiguiente, desvirtúa el mensaje.

Un código así sería muy difícil de utilizar y resultaría extremadamente largo a causa de la dificultad de memorizar palabras casi imposibles de pronunciar. Aun en la comunicación de viva voz, los esfuerzos de los hablantes deberían ser tan grandes y el peligro de pronunciar u oír mal las palabras tan evidente, que el código inventado aquí no correspondería ni de lejos a su función específica: el aseguramiento de la comprensión entre hombres.

¿Qué modificaciones deberíamos aplicar a este código para que se volviera practicable? En primer lugar, dispongamos que la longitud de las palabras no pueda ser indefinidamente grande: p. ej., un mínimo de un sonido y un máximo de diez sonidos³⁷. En tal caso, si el interlocutor oye una sucesión de 11 ó 13 sonidos entre dos pausas, sabrá que se ha producido un error. Pero esto no es suficiente. Hay que añadir nuevas restricciones: exijamos, p. ej., que toda palabra contenga al menos una vez una *a* o una *e*; que no admita combinaciones más largas de dos *c*, una *b*, una *a* o una *e* sucesivas y que entre una *b* y una *c* aparezca siempre una *a* o una *e*. Aplicando estas restricciones, las palabras de la lengua así imaginada tendrán una forma como *e*, *ab*, *beaccaae*, *babecabeb*. Entonces, si nuestro imaginario interlocutor oye una sucesión de tipo **ebc* o **baacbe*, inmediatamente advertirá que el mensaje está equivocado.

Un código de este tipo tiene la ventaja no sólo de permitir el descubrimiento de errores, sino también de afianzar la comprensión del mensaje incluso cuando algunas de sus partes quedan bajo el efecto de distintos ruidos (cf. 10.6 y ss.). Supongamos que de la palabra *Baca* no se ha oído la primera *a* y que, por tanto, aquélla ha quedado reducida a **bca*. El receptor de la comunicación se dará cuenta, sin duda, de que se ha producido un error y reconstituirá la palabra a base de *beca* o *baca* (de acuerdo con la regla según la cual entre *b* y *c* debe haber forzosamente una *a* o una *e*). El contexto le ayudará a decidir también si se trata de *beca* o de *baca*. Ocurre, pues, que la estructura de nuestro código nos permite “adivinar” determinados sonidos: después de *b* sabemos que sigue *a* o *e*, pero no *b* o *c*. La probabilidad de acertar equivale, entonces, al 50%. El sonido que sigue a *b* (*a* o *e*) nos proporciona una cierta información, igual, en este caso, a un bit. Evidentemente, las restricciones aplicadas han aumentado sensiblemente la redundancia de nuestro código, pero este aumento venía impuesto por la necesidad misma de asegurar la comunicación.

Supongamos, ahora, que una palabra empieza por *a*. Después de *a* puede seguir *e*, *b* y *c*. La única posibilidad excluida consiste en que después de *a* siga de nuevo *a*. La probabilidad de acertar resulta, desde luego, menor. Por tanto, también los sonidos *e*, *b* y *c* aportan una determinada información.

37. Este tipo de restricciones caracteriza a toda lengua real, entre otros motivos porque sin ellas surgirían cantidades de palabras muy superiores a las necesidades de comunicación.

Esta información, ¿es mayor o menor que en el ejemplo precedente? Si después de un sonido A sólo puede seguir uno de dos sonidos (x, y), el sonido x proporciona una información más pequeña que cuando después de A puede seguir x, y, w, z . En efecto, si hay que elegir sólo entre dos sonidos, las posibilidades de acertar son mayores que si hay que elegir entre tres (o más) sonidos. Cuando la predicción es más fácil, las variantes a escoger comportan informaciones más escasas. Dicho de otra manera, si hay que optar entre dos variantes se requiere menos información que si hay que optar entre 100 variantes. La información es, como se ha visto, una medida de la libertad de elegir entre dos o más variantes.

Las restricciones enunciadas antes con respecto a las posibilidades de combinación para los cuatro fonemas a, b, c, e constituyen la redundancia de nuestro código. Así, pues, la redundancia limita la posibilidad de aparición de un sonido (en rigor, de cualquier unidad de un código) e implícitamente facilita la predicción de que aparezca el sonido.

Añadamos ahora dos nuevas restricciones que van a aumentar todavía más la redundancia del código: 1) ninguna a puede ir inmediatamente seguida de una e ; entre ambos sonidos debe existir una b (las palabras tipo $*aeb$ quedan excluidas); 2) toda palabra debe terminar con b (no hay palabras tipo $*eba$).

En la nueva textura del código, la seguridad de la transmisión correcta del mensaje se ha acrecentado considerablemente. Si el receptor oye $*aeb$, se dará cuenta del yerro y reconstituirá fácilmente la forma correcta $abeb$. Si oye $*eba$, restablecerá de inmediato la palabra $ebab$.

En seguida observamos que b final o intercalada entre a y e puede ser omitida en la pronunciación o en la escritura, ya que siempre se advierte su ausencia. La presencia o ausencia de b no promueve diferencias de significado y, por tanto, dicho sonido no proporciona información alguna en tanto que se conozca el sonido que lo precede y que lo sucede.

La cantidad de información disminuye con el número de las variantes. Después de a , en el ejemplo de antes, no puede ir más que b , de modo que aparece una sola variante. En posición final de palabra sólo puede figurar una b ; de nuevo, pues, una sola variante. El mínimo de información viene dado por dos variantes (como $H_0 = \log_2 n$, siendo n el número de variantes, en nuestro caso resultará que $H_0 = \log_2 2$, es decir $H_0 = 1$); si hay menos de dos variantes, concretamente una, la información es menor que 1, es decir nula, $H_0 = 0$, ya que el logaritmo de base 2 de 1 es igual a 0.

Las unidades de un código cuya aparición en determinadas condiciones (en un contexto determinado) pueda preverse con seguridad y, por consiguiente, no aportan ninguna información, se denominan *redundantes* en las condiciones respectivas (en el contexto respectivo). Así vemos que redundancia equivale a “información nula” o “aparición totalmente previsible”.

Entre los dos tipos de redundancia (limitación de las posibilidades de apari-

ción de una unidad, por una parte, y completa predicción de la aparición de una unidad, por otra), la diferencia es sólo cuantitativa. En el primer caso, la redundancia es menor y en el segundo, máxima. Como hemos visto, a medida que crece la cantidad de información, la redundancia disminuye y, viceversa, cuando la cantidad de información baja, la redundancia aumenta. En la composición de códigos hay que tener, pues, en cuenta dos requerimientos contrarios: por un lado, la necesidad de comunicar al máximo, de dar la mayor cantidad posible de información; por otro, la necesidad de evitar los errores en la transmisión, de permitir su aquilatamiento y de hacer posible la comunicación aun en las inevitables condiciones perturbadoras que sin duda dificultarán el proceso comunicativo. Esta segunda necesidad reclama la existencia de una cantidad de redundancia suficientemente grande. Se comprende, pues, que la redundancia intervenga en la lengua para intensificar la seguridad de la transmisión.

8.15. La redundancia en las lenguas naturales

En las páginas precedentes, para clarificar nuestro problema, hemos inventado una lengua con cuatro sonidos y con ciertas reglas para su combinación en palabras. Pasaremos ahora a las lenguas reales a fin de demostrar que están formadas a partir de los mismos principios. En realidad, ello es cierto en general; y tanto es así, que muchos códigos se basan en el código fundamental que es la lengua.

La lengua rumana, p. ej., es un código compuesto por un cierto número de fonemas, los cuales no pueden combinarse para formar palabras si no es con arreglo a determinadas reglas. El hecho de que dos o más sonidos no puedan aparecer en secuencia en una lengua dada puede deberse a diversos motivos: en primer lugar, a que la secuencia sería difícil de pronunciar (el concepto de dificultad de pronunciación es relativo; un grupo de sonidos aparentemente impronunciable para los hablantes de una lengua es posiblemente frecuente en otra). En segundo lugar, porque existen en cada lengua ciertas reglas combinatorias como consecuencia de su evolución histórica. Alguna concatenación de sonidos puede no encontrarse en ella a pesar de que los hablantes no ofrecen dificultad en pronunciarla (y la pronuncian con soltura cuando la hallan en una lengua extranjera)³⁸.

En rumano, como en todas las lenguas naturales, existen numerosas reglas combinatorias para los sonidos. Por ejemplo, ninguna palabra puede empezar por *dbl*, *gcr*, etc., ni puede terminar en consonante doble; determinados grupos consonánticos pueden aparecer en posición inicial o medial de palabra, pero no

38. Como se ha dicho antes, sin estas restricciones obtendríamos muchísimas más palabras de las necesarias. Así, p. ej., en una lengua con 36 fonemas se formarían 36 palabras monofonémicas, 1296 de dos fonemas, 46 656 de tres, 1 679 616 de cuatro (es decir 36^1 , 36^2 , 36^3 , etc.).

en posición final (p. ej., *str*). Todas estas reglas constituyen fenómenos de redundancia al nivel de los fonemas, por las cuales se limita la posibilidad de aparición de algunos de ellos. Existen incluso casos de redundancia máxima (previsión total). Así, p. ej., al oír una “pseudo *i* final”, como en *colfi*, ‘colmillos’, sabemos que termina la palabra. (Después de esta *i* sólo puede seguir otra palabra o el final del mensaje.)³⁹ En este caso, la previsión es absolutamente segura.

Durante el habla, tenemos la impresión de que elegimos libremente las palabras y de que decimos lo que nos viene en gana decir. Pero el análisis de la redundancia nos demuestra que sólo una parte del mensaje resulta de nuestra elección; el resto lo decimos casi sin querer, obligados por las reglas del código empleado. (Estas reglas tienen un carácter estadístico, es decir indican lo que ocurre por término medio, lo que es más y menos probable.)

La redundancia de los sonidos depende también del factor semántico. Veamos algunos ejemplos ilustrativos con experiencias simples.

1) Formulamos mentalmente una proposición que empieza por *g*. Pronunciamos esta *g* y pedimos al interlocutor que adivine el siguiente sonido. Como ignora lo que queremos decir, el interlocutor deberá elegir entre las posibilidades que le ofrecen las reglas de redundancia de la lengua (que cualquier hablante conoce, al menos empíricamente): después de *g* puede seguir cualquier vocal o una de las consonantes *l*, *r*. Se comprende, pues, que aunque hay todo el sistema fonemático a disposición, sólo concurren 7 variantes en la elección.

2) Pronunciamos luego la frase *Le han apaleado fuerte con un g* y pedimos al interlocutor que adivine el sonido siguiente. Teóricamente, también aquí hay 7 posibilidades de elección, sólo que en este caso se entrevé bastante bien de qué se trata: la palabra que empieza por *g* debe ser el nombre de un objeto con que alguien pueda ser pegado. Utilizando una cierta dosis de lógica, el interlocutor considerará, pongamos, las palabras *garrote*, *gancho*, *golpete* y *guizque*, que le sugerirán la elección entre *a*, *o*, *i*. Debe, por tanto, escoger un fonema entre tres y ya no entre siete. Si elige entonces *a* y acierta, se hallará ante la alternativa de escoger entre *rr* y *n* y no entre los 24 fonemas del español. Una vez acertado, digamos *rr*, continuará la experiencia pronunciando sin vacilación los fonemas *o*, *t* y *e*, con los que reconstituirá la palabra *garrote*. De ahí surge una conclusión importante: si el mensaje comunicado llega fragmentariamente al receptor *Le han apaleado fuerte con un garr* (de forma que aquí intervenga un ruido que elimine la voz), el resto *-ote* es muy fácil de reconstituir. En otras palabras, en este contexto, *garrote* —la palabra entera— no nos dice más que la parte *garr*, puesto que la parte *ote* es redundante⁴⁰.

39. [A excepción de esta última (cf. también 7.2, nota 6), se trata de restricciones también válidas para el español, mucho más radical, a este propósito, que el rumano].

40. Desde luego, esto no siempre se cumple para la porción /gar/ de la palabra *garrote*. En un contexto [semánticamente mucho más extenso] como *¿Has visto un gar?*, las posibilidades de acierto son menores.

3) Si pedimos al interlocutor que recomponga los sonidos de palabras extranjeras, de una lengua que ignora, la predicción resultará muy dificultosa, tanto al principio como al final de la frase⁴¹.

8.16. Ejemplos y aplicaciones

De los ejemplos aducidos resulta que la redundancia de un fonema es mayor cuando se intuye el significado general de la proposición en que aparece. (De ahí que sea más fácil adivinar los sonidos del final de una frase —donde hay fundadas expectativas— que no los del principio). La redundancia crece, pues, paralelamente con el conocimiento, más o menos vago, del significado contextual.

Las consecuencias del hecho por el cual la aparición (u ocurrencia) de determinados sonidos pueda preverse tienen una gran importancia incluso desde el punto de vista práctico. Efectivamente, ello se refleja en los distintos sistemas de escritura, como lo demuestran los ejemplos siguientes:

1) hay sistemas estenográficos donde apenas se escriben las vocales (debido a que éstas son más redundantes que las consonantes)⁴²;

2) en la escritura de las lenguas semíticas, las vocales no se indican, ya que pueden reconstituirse a partir de las reglas gramaticales y de la intuición del significado general de la frase (de ahí que, a veces, para adivinar las vocales de una palabra árabe escrita, haya que leer la frase hasta el final y luego volver al principio);

3) aunque no conozcamos ningún sistema estenográfico, recurrimos a menudo a abreviaturas basadas en el principio de la redundancia. Así, p. ej., en una frase como *Se ve quién tuvo razón en est. probl.*, las abreviaciones *est. probl.* comportan la misma información que *este problema* (las sucesiones de letras *e-ema* son fácilmente previsibles);

4) en la ortografía rumana, la letra *h* tiene dos valores (cf. *haină* [xâjnə], ‘chaqueta’, frente a *cheie* [kêjye], ‘llave’), lo que no incomoda en absoluto al lector que conozca las leyes de redundancia de la fonología rumana. Conoce, por ejemplo, el hecho (o restricción) de que después del fonema *c* [c] no puede seguir el fonema *h* [x] y, por ello, no leerá esta *h* de *cheie* como la *h* de *haină*. La letra *i* se lee de tres maneras diferentes: *i* “vocálica” (como en *tigri* [tígri], ‘tigres’), *i* “semivocálica” (como en *boi* [bói], ‘bueyes’) y “pseudo *i* final” (como en *elefanți* [elefânti], ‘elefantes’)⁴³. Para el que domina el sistema del rumano,

41. Cf. MEYER-EPLER, “Anwendung”.

42. Cf. PIOTROVSKI, “Problemă”.

43. [Cf. 7.2, nota 6. A despecho de lo que se dice a continuación, hay casos en que debe recurrirse a la sintaxis y al significado del contexto para distinguir ciertas alternativas gráficas; v. gr. *orbi* [órb’] ‘ciegos’, pero [orbi] ‘cegar’].

la diferente lectura de *i* final en las tres situaciones descritas no representa ninguna dificultad. La sucesión de consonante + *r* mojada no está admitida en el sistema fonológico rumano; después de otras consonantes, cuando indica el plural sin artículo, *i* se lee como en *elefanți* y no como en *tigri*; después de vocales, indicando también plural no articulado, *i* se lee siempre como en *boi*. Las restricciones fonológicas y morfológicas, es decir la redundancia, posibilitan una gran economía ortográfica (una sola *i*, en lugar de tres).

La redundancia no sólo existe a nivel de fonemas. La experiencia puede extenderse incluso al dominio morfemático. Así, p. ej., al morfema *băiat*, 'muchacho', sólo puede asignársele uno de los morfemas *ul*, *ule*, *ului* (*băiatul*, 'el muchacho' [Nom., Ac.], *băiatule*, '¡(oh), muchacho!' [Voc.], *băiatului*, 'del/al muchacho' [Gen., Dat.], respectivamente). Después de *veni*, 'venir', puede haber pausa (y seguir otra palabra o terminarse la comunicación) o bien *se*, eventualmente seguido, a su vez, de uno de los morfemas *m*, *și*, *ră* (solo o + *m*, *ți*), *i*, *nd*, *t* (seguido o no de *ă*, *e*), *re* (seguido o no de *a*). Si se da, entonces, un morfema en un contexto, puede predecirse una serie de otros morfemas con certeza. (Si un extranjero pronuncia en rumano algo así como **I-am dat băiat acestuia*, es fácil para los nativos reconstituir *băiatului*, aquí totalmente redundante)⁴⁴. La redundancia morfológica está comprendida en el sistema gramatical de cada lengua, hasta el punto que describir este sistema equivale, en rigor, a describir la redundancia morfológica de la lengua en cuestión.

La redundancia difiere de una lengua a otra. Por ejemplo, en la proposición rumana *Fata mea e leneșă*, *a de mea* y *ă de leneșă* son elementos redundantes (sabiendo que *fata* es un nombre femenino, inmediatamente se hacen predictibles las desinencias de los adjetivos que lo determinan). La versión española de la frase, *Mi hija es perezosa*, resulta menos redundante, en este sentido, ya que la forma *mi* carece de morfema genérico, y todavía menos la versión inglesa, *My girl is lazy*, donde *my* y *lazy* carecen igualmente de desinencias de género⁴⁵. La versión inglesa es, pues, la menos redundante de las tres consideradas.

Si declinamos el sintagma rumano *copilul bun*, 'el niño bueno', obtenemos en genitivo *copilului bun*, 'del niño bueno', en el que se modifica tan sólo la forma del nombre; procediendo de igual modo en alemán, el cambio entre *das gute Kind* y *des guten Kindes* comporta una modificación, por lo demás predictable, tanto en el nombre como en sus determinantes. En este caso, la versión rumana es menos redundante que la alemana⁴⁶.

Un tipo característico de redundancia morfológica resulta de la formación

44. [*I-am dat băiat acestuia* significa, en rigor, 'He dado muchacho a éste', pero ante la rareza de tal significado (salvo en contextos canibalescos), se supone que el extranjero pretende decir *I-am dat băiatului acestuia*, 'Se (lo) he dado a este muchacho', con concordancia de *băiat* y *acest* en dativo].

45. [Hemos interpolado el ejemplo en español].

46. Para otros ejemplos de redundancia, cf. STATI, "Categoriile".

del plural en rumano. Al singular *tigru*, ‘tigre’, se le opone el plural *tigri* (es decir, *i* se opone a *u*); en cambio, al singular *casca*, ‘casco’, se le opone el plural *caști*. *Casca* se compone de cinco fonemas, como *tigru*, pero el plural de *tigru* se distingue del singular por un solo fonema, mientras que *caști* difiere de *casca* en todos sus fonemas a excepción del primero. Para formar el plural, hubiera bastado un solo fonema en este último caso, de modo que el resto constituye redundancia.

De ahí podemos deducir un principio general: si a una diferencia del plano del contenido (del significado) corresponde más de una diferencia en el plano de la expresión (del significante), la realización es *r e d u n d a n t e*. Y al contrario, si a una diferencia de contenido le corresponde una sola diferencia de expresión, la realización es *e c o n ó m i c a*. La experiencia puede extenderse también a la teoría combinatoria de las palabras en el marco de la oración o incluso a la combinación oracional en el marco del período. Así, p. ej., sintagmas como **el niña* o **convengo el hombre* no son admisibles. Ciertas combinaciones están permitidas por la gramática, pero no por la lógica; p. ej., *El cielo verdea bajo el arado de cerillas evaporadas*⁴⁷.

Examinemos ahora un poco las relaciones entre los conceptos de “redundancia” y de “inutilidad”, para evitar su confusión.

Diciendo que *ă* de *fată leneșă*, ‘muchacha perezosa’, es redundante, sobreentendemos que si el mensaje se deteriora en la porción final *ă*, la comunicación no queda dañada, de modo que podríamos prescindir de este morfema sin ningún detrimento: resultaría igualmente informativo decir *fată leneș*. Pero en otro contexto, el mismo morfema se vuelve muy útil: *Popescu este leneșă*, ‘Popescu es perezosa’, donde *Popescu*, como apellido, carece de formas genéricas.

¿Cómo debe interpretarse, desde este punto de vista, *l* como artículo enclítico rumano? [A este respecto, *pom*, ‘árbol’, se opone a *pomul*, ‘el árbol’, para los nombres de género masculino]. Al pronunciar, sin embargo, *pomu*, como ocurre en el habla familiar, en lugar de *pomul*, se ve inmediatamente que se utiliza la forma con artículo, pues se opone igualmente a *pom*, sin artículo. Sucede, entonces, que la añadidura de *l* no aduce ninguna nueva precisión. Pero en palabras como *tigru*, ‘tigre’, o *codru*, ‘oquedal’, sin *l*, la forma articulada y la no articulada coinciden. Este tipo de palabras es, no obstante, escaso en rumano y, por otra parte, concurre además el contexto para distinguir entre *tigru* (sin artículo) y *tigru* (con artículo), del mismo modo que ayuda a distinguir entre el singular *nume*, ‘nombre’, y el plural *nume*, ‘nombres’. Desde luego, *l* resulta, como artículo enclítico, redundante y su uso dispensable. Ahora bien, si desapa-

47. Ciertas combinaciones de palabras son más probables en una lengua que en otra [cf. nota 44]. El lingüista soviético L. R. Zinder da como ejemplo la proposición *Estar sobre el tejado*, más probable, según él, en una lengua hablada en Asia Central que en una lengua europea de la URSS. Cf. ZINDER, “Probabilitatea lingvistică” y MARTINET, *Éléments*.

recieran del rumano demasiados morfemas redundantes para fiarlo todo al factor contextual, la comunicación se volvería en un momento dado equívoca, si no imposible. El empleo de los rasgos redundantes en la morfología de una lengua debe apreciarse en su totalidad. La pérdida de cualquiera de ellos apenas afectaría la comunicación, pero también es verdad que deben existir rasgos de este tipo en cantidad suficiente a fin de cooperar contra el efecto de los ruidos de toda clase.

Para concluir este problema digamos que algunos rasgos morfológicos son redundantes sólo en ciertos contextos, mientras otros lo son (casi) siempre. Ni los unos ni los otros son inútiles, en tanto que fortalecen la seguridad de la comunicación⁴⁸. En el discurrir histórico de las lenguas, los rasgos redundantes, o que se vuelven redundantes, terminan siendo eliminados a veces. Las funciones casuales pueden expresarse igualmente con ayuda de preposiciones; en caso de emplear ambos procedimientos, disminuye el cometido de las desinencias casuales, que pueden alcanzar un máximo de redundancia, hasta que la lengua se desembaraza eventualmente de ellas. Este fenómeno lo han experimentado, entre otras lenguas, el latín tardío [y, con él, las lenguas románicas occidentales, más o menos temprano], el búlgaro, el inglés, etcétera. La desaparición de unos rasgos redundantes puede compensarse en el sistema de la lengua por la aparición de otros. Se ha observado, sin embargo, que existe una tendencia a la reducción gradual de la redundancia: una gramática analítica, p. ej., es menos redundante que una gramática sintética⁴⁹.

8.17. Observaciones finales

En las páginas precedentes no hemos hecho más que esbozar el significado lingüístico de los conceptos de entropía y redundancia. Hasta el presente, los matemáticos, en colaboración con los lingüistas, han estudiado la entropía y la redundancia de diversas lenguas, teniendo más en cuenta la lengua escrita (las letras) que la hablada (los fonemas). Evidentemente, la estructura estadística (y, por tanto, la redundancia) de ambas modalidades no coincide; así ocurre, por ejemplo, en rumano, donde la letra *h* es más frecuente que el fonema *h*, desde el momento en que *h*, como letra, aparece en las combinaciones *che*, *chi*, *ghe*, *ghi*, donde no corresponde al fonema *h*⁵⁰. Para esta misma lengua, la experiencia

48. Por ello, hay que rechazar la idea, sostenida en una serie de tratados lingüísticos, de que la lengua ideal sería una que estuviese completamente desprovista de redundancia.

49. Cf. MILLER, *Language*.

50. [Como en italiano, estas grafías indican la pronunciación oclusiva velar de los grupos respectivos [ke, ki, ge, gi]; frente a la pronunciación palatal de *ce, ci, ge, gi* [ce, ci, je, ji]. En otras condiciones, "h" indica pronunciación fricativa velar sorda [\approx x]].

parece demostrar que se adivina, por término medio, una letra aproximadamente entre tres⁵¹. Sobre textos literarios de Eminescu y de Arghezi, se ha comprobado que la redundancia media es de 10,6%⁵².

La determinación de la entropía y de la redundancia tropieza con grandes dificultades cuando se pasa del nivel fonológico a otros superiores, como el de los morfemas, de las palabras o de las oraciones. A este propósito, los resultados todavía no pueden considerarse satisfactorios⁵³.

Para el rumano, las investigaciones apenas han comenzado. La tarea de los lingüistas consiste en determinar, en colaboración con los matemáticos, la entropía y la redundancia específicas en los distintos compartimentos de la lengua (fonología, morfología, sintaxis). Las cifras así obtenidas sintetizarán los rasgos característicos del sistema de la lengua. Algunos de estos resultados no harán más que precisar lo que se sabe ya a partir de los métodos lingüísticos tradicionales. Pero otros serán completamente nuevos y descubrirán características todavía inadvertidas en la estructura fonológica o gramatical. Una vez conocida la entropía y la redundancia de la lengua, se propiciarán progresos esenciales en la elaboración de una gramática rigurosa y científica, que indique, p. ej., qué unidades no pueden combinarse o se combinan obligatoriamente, qué combinaciones son más probables (y cuánto, exactamente) que otras, y así sucesivamente⁵⁴. A estos beneficios hay que añadir, por supuesto, otros de índole estrictamente técnica: la composición del código para una máquina de traducir, para máquinas informacionales, para la mejora del sistema de comunicaciones a gran distancia, etc.

51. Cf. NICOLAU-SALA-ROERIC, "Observații".

52. Cf. nota anterior, p. 130.

53. Cf. DOLEŽEL, "Význam teorie informace".

54. Desafortunadamente, esto no se ha conseguido todavía, ni siquiera en el plano fonemático, por falta de descripciones estadísticas completas sobre el sistema fonológico rumano.

9. Otros métodos estadísticos en lingüística

9.1. Introducción

Además de las traducciones automáticas, la cibernética ha proporcionado también un nuevo impulso a la lingüística desde el punto de vista cuantitativo con el establecimiento de unas propiedades estadísticas sobre diferentes lenguas. A este respecto, hay que recordar que C. Shannon, ya en la memoria fundamental en que plantea las bases de la teoría matemática de la comunicación, se ocupa de las propiedades estadísticas de la lengua inglesa¹. Más tarde, las mismas investigaciones se han extendido a otras lenguas. En lo que sigue, después de algunas consideraciones generales acerca de los métodos estadísticos, expondremos diversas leyes de este tipo que actúan al transmitirse la información a través del lenguaje.

9.2. Generalidades históricas

La práctica cotidiana nos pone ante hechos o eventos calificables de casuales. Así, si prestamos atención a la última cifra de las matrículas automovilísticas que encontramos en un día, comprobaremos que, por término medio, hallamos tantos números pares como impares.

Los eventos en cuya descripción intervienen medias estadísticas se denominan **eventos estadísticos**. De un modo semejante, cabe definir también las propiedades y las magnitudes estadísticas.

Las leyes que expresan relaciones exactas entre eventos, propiedades o magnitudes parcial o totalmente estadísticas se llaman **leyes estadísticas**. Las leyes estadísticas se identifican por el hecho de que, a medida que disminu-

1. Cf. SHANNON, "Prediction".

ye el número de elementos sobre los que operan, el resultado de la experiencia presenta fluctuaciones paralelamente mayores con respecto a la evolución media descrita por ellas.

Hace mucho que los científicos estudian los problemas relativos al establecimiento de unas leyes que se refieran a los fenómenos casuales. Las primeras investigaciones en esta dirección se deben a los matemáticos Pierre Fermat (1601-1655), Blaise Pascal (1623-1662) y Jacques Bernoulli (1654-1705), quienes analizaron ciertos aspectos de los juegos de azar.

Más tarde se extendió considerablemente el dominio de aplicación de cálculo de probabilidades.

El término “estadística” fue introducido a mediados del s. xviii por Achenwall, a la sazón profesor de Derecho Público en la Universidad de Gottinga. Por estadística, Achenwall entendía “el conocimiento profundizado de la situación (o “status”) relativa y comparada de cada Estado”. Luego, la estadística desbordó los problemas estrictamente demográficos y económicos hasta alcanzar una esfera más amplia de estudio.

En seguida se comprobó que el cálculo de probabilidades encuentra en la estadística un terreno muy propicio para diversos empleos. La introducción de los métodos estadísticos en física promovió resultados muy notables y puso a punto un instrumento de investigación especialmente poderoso. Gracias a ello, la física ha podido tratar los fenómenos de masa, donde intervienen grandes reuniones de objetos y donde no es posible estudiar por separado todos los factores que determinan el comportamiento de estas reuniones o de cada uno de sus elementos. En otras palabras, se ha hecho posible abordar estadísticamente los fenómenos físicos.

En la ciencia, los conjuntos estadísticos han empezado a estudiarse más bien tardíamente: el año 1738, D. Bernoulli explicaba la presión de los gases como un efecto del desplazamiento permanente de las moléculas del gas, las cuales en su movimiento golpean las paredes del recipiente. En aquella época se consideraba que todas las moléculas de un gas tienen la misma velocidad. Esta concepción se encuentra todavía el año 1856 en Krönig, quien, a partir de un modelo mecánico del movimiento molecular en un gas, consigue deducir las leyes de Boyle-Mariotte, Gay-Lussac y Avogadro.

Sin embargo, a J. C. Maxwell corresponde el mérito de haber logrado, en 1860, crear el concepto de distribución estadística de las velocidades de un gas, introduciendo también la distribución que lleva su nombre. La idea básica consiste en que no todas las moléculas de un gas tienen la misma velocidad. En rigor, lo que se denomina “velocidad” en las teorías antiguas relativas a la teoría cinética, no era más que la velocidad media de las moléculas. Pero, en realidad, las moléculas tienen velocidades variables en torno a este valor, unas desplazándose más de prisa y otras más despacio respecto de aquellas que tienen la

velocidad media. Por distribución de las velocidades debemos entender el porcentaje de moléculas que tienen una determinada velocidad, perteneciendo a un determinado gas contenido en un determinado recipiente. Maxwell mostró que en los gases la distribución de las velocidades se produce de acuerdo con una ley que lleva precisamente su nombre.

Las ideas de Maxwell han resultado no sólo justas, sino también extremadamente fecundas al hacer evidente la existencia de unas leyes estadísticas. Más tarde, estos conceptos se han ampliado a otros dominios, entre ellos la lingüística.

En física, partiendo de la teoría cinética de los gases, se han alcanzado éxitos importantes en termodinámica gracias a la aplicación de los métodos estadísticos. L. Boltzmann y J. W. Gibbs crearon la mecánica estadística, que consigue explicar los fenómenos térmicos justamente con ayuda de leyes estadísticas. L. Boltzmann logró explicar también la entropía de un sistema físico igualmente con ayuda de conceptos estadísticos. En su estudio consideraba las partículas componentes del sistema y las probabilidades de que estas partículas se hallaran en determinados estados, p. ej., de energía. Indicando mediante w la probabilidad de que un número dado de partículas se hallen en estado i , Boltzmann demostró que la entropía del sistema viene dada por la fórmula:

$$S = -k N \sum w_i \log w_i,$$

donde k es una constante y N el número total de partículas del sistema considerado.

Parece evidente la identidad entre la expresión de la entropía, tal como la dio Boltzmann para los fenómenos termodinámicos, y la entropía informacional dada por Shannon. Es probable incluso que Shannon se inspirara en los trabajos de Boltzmann, lo que es normal, ya que en ambos casos nos hallamos en presencia de fenómenos estadísticos.

La teoría atómica ha adoptado las nociones de distribución estadística. Así, en el dominio de las partículas elementales se conocen diferentes categorías de partículas que obedecen a distintas distribuciones, como, p. ej., las de Fermi-Dirac o Einstein-Bose, con que se ponen de manifiesto los bosones y los fermiones. También la física ondulatoria está indisolublemente vinculada a la estadística.

9.3. Los métodos estadísticos en lingüística

La introducción de los métodos estadísticos en la Física permitió, como hemos visto, la obtención de unos resultados revolucionarios.

Por analogía, puede decirse asimismo que la introducción del concepto de estadística en la lingüística ha promovido avances muy considerables. En primer

lugar, se introdujo en lingüística la posibilidad de establecer unas determinaciones cuantitativas, lo que permitió obtener el predicamento de exactitud indispensable a toda ciencia contemporánea. Al mismo tiempo, y desde un punto de vista práctico, los métodos estadísticos ponen a disposición de los investigadores un instrumento de análisis excepcionalmente fecundo e indispensable para el descubrimiento de los rasgos característicos de ciertos fenómenos lingüísticos de indole estadística.

Los métodos estadísticos permiten igualmente poner de manifiesto la estricta relación que existe de un modo natural entre habla y cibernética. De ahí que, gracias a la estadística, pueda demostrarse que en el habla corriente actúan diversas leyes que hacen que la transmisión de la información a través del lenguaje se realice conforme a ciertos principios de extremos.

El planteamiento estadístico de los fenómenos lingüísticos puede efectuarse desde niveles muy diferentes.

Así, en primer lugar, puede abordarse de un modo estadístico el problema de los estilos, poniendo en evidencia las características macroscópicas que posibilitan el reconocimiento de una obra literaria como perteneciente a un autor concreto.

A un nivel inferior, cabe efectuar un análisis estadístico sobre la frecuencia de aparición de ciertas palabras en una lengua. Como se sabe, a este aspecto se refiere la ley de Zipf, que corresponde a un principio de mínimo. En esta misma línea, las investigaciones emprendidas sobre la frecuencia de aparición de las sílabas o sobre la longitud media de las palabras —que se mide por el número de sílabas— muestran también la existencia de determinadas leyes estadísticas.

Puede incluso hablarse de la frecuencia de aparición de letras y fonemas. En esta dirección, y aplicándose a textos escritos, pueden investigarse los alfabetos utilizados para establecer, p. ej., los criterios en que se basa el reconocimiento de las letras. Es posible incluso clasificar los alfabetos a partir de la dificultad que plantea —en promedios estadísticos— el reconocimiento de los caracteres escritos en cada caso.

Las averiguaciones de tipo estadístico pueden, desde luego, extenderse a las formas gramaticales.

En los párrafos inmediatos, atenderemos a los principales resultados obtenidos en el dominio del estudio estadístico de las lenguas.

9.4. La ley de Zipf

Uno de los problemas sencillos que admite un planteamiento estadístico se refiere a la frecuencia de las palabras, cuestión generalmente ligada al nombre de Zipf, por haber establecido él una ley muy fácil de presentar (cf. 8.3).

Los historiadores dicen, no obstante, que la ley conocida por el nombre de Zipf habia sido descubierta anteriormente por Condon e, independientemente, por Estoup. En esencia, esta ley indica lo siguiente. Si se considera un texto de una lengua dada y se cuentan las palabras contenidas en él, se comprueba que la frecuencia de aparición de estas palabras es diversa, en el sentido de que las distintas palabras aparecen con una frecuencia variable, unas más a menudo que otras.

Existe la posibilidad de clasificar estas palabras en orden decreciente con arreglo a sus frecuencias de aparición. Si se nota mediante n el rango de la palabra que ocupa el lugar n en el conjunto de las palabras ordenadas de acuerdo con su frecuencia de aparición y mediante f_n la frecuencia respectiva de aparición, entonces se advierte experimentalmente que subsiste la siguiente relación:

$$f_n = \frac{P}{n},$$

donde P es una constante.

Evidentemente, el rango 1 se refiere a la palabra más frecuente; el rango 2 a la palabra cuya frecuencia sigue inmediatamente a f_1 en el conjunto de las frecuencias ordenadas según sus valores decrecientes, etc.

Más tarde, B. Mandelbrot trató de dar una interpretación estratégica a esta ley, empíricamente expuesta por Condon, Estoup y Zipf². Para ello, Mandelbrot supone que la información total que se transmite por medio de palabras está fijada de antemano, pero no las probabilidades p_n correspondientes a la aparición de cada una de las palabras, ya que la persona que las emite queda libre de distribuir la información entre las R palabras que posee.

Se sabe que la cantidad total de información media por palabra equivale a:

$$H = -\sum p_n \log_2 p_n$$

y que el número de las palabras potenciales es R .

Según Mandelbrot, la mejor distribución para las frecuencias de las palabras es aquella para la cual se cumple un determinado principio de mínimo. Considera, así, que la emisión de cada palabra está asociada a un determinado coste, cuyo valor medio es

$$\sum p_n j_n.$$

Mandelbrot no precisa cuál es la naturaleza de este coste, tal vez psicológico, fisiológico o de otra índole, eventualmente la suma de todos ellos. Postula, en todo caso, que la mejor distribución para la frecuencia de las palabras será aquella para la cual el valor del coste medio por palabra sea mínimo.

2. "Estratégica", en el sentido de la moderna teoría matemática de la "estrategia de los juegos".

Utilizando el método de los multiplicadores de Lagrange, Mandelbrot consigue, a partir del principio variacional global aducido, llegar a la ley canónica local

$$p_n = P e^{-Bj_n}$$

que admite asimismo la forma

$$p_n = PM^{-Bj_n} = P(m+n)_M^{-B},$$

donde $[x]_M$ es el número cuyo logaritmo equivale al entero más pequeño mayor que $\log_M x + j_0$.

De un modo simplificado, esta ley canónica se convierte en

$$p_n = P(m+n)^{-B}.$$

Se observa inmediatamente que la ley de Zipf aparece como un caso particular de esta ley más general, dentro de la cual supone

$$m = 0, \quad B = 1.$$

La ley simplificada de Mandelbrot concuerda con datos experimentales procedentes de textos de distintas lenguas.

Hay que subrayar el hecho de que, en la demostración de Mandelbrot, corresponde un papel decisivo a la magnitud H , la cual es, en realidad, la entropía de las categorías de palabras agrupadas según su probabilidad de aparición.

9.5. La distribución de las palabras en la lengua rumana y en otras lenguas

En lo que concierne al rumano, las investigaciones demuestran que también aquí las palabras ofrecen distintas frecuencias de aparición. Vale la pena destacar el hecho de que, desde hace mucho, B. P. Hasdeu había probado, en diversas lenguas, que la frecuencia de aparición de las palabras tiene valores muy distintos en función de las palabras utilizadas. Demostró que la estadística efectuada por Cihac para el rumano ni era correcta ni reflejaba como se debe el carácter románico de la lengua, ya que las palabras de su diccionario tienen valores de circulación —es decir, frecuencias de aparición— muy diferentes entre sí³.

Investigaciones más recientes emprendidas en el Instituto de Lingüística de Bucarest han confirmado por completo la exactitud de las tesis de B. P. Hasdeu.

Por lo que se refiere a la verificación de la ley de Zipf, merecen citarse las estadísticas realizadas por E. Graeser y P. Şuteu, cuyos resultados se reflejan, respectivamente, en las figuras 30 y 31. En sus conclusiones, se ratifican igual-

3. [En su recuento sobre el diccionario —y no sobre los textos—, A. de Cihac (1825-87) había llegado a la conclusión de que las palabras rumanas de procedencia eslava son mucho más numerosas que las de origen latino].

mente los juicios de Hasdeu sobre el origen latino del fondo léxico principal del rumano⁴.

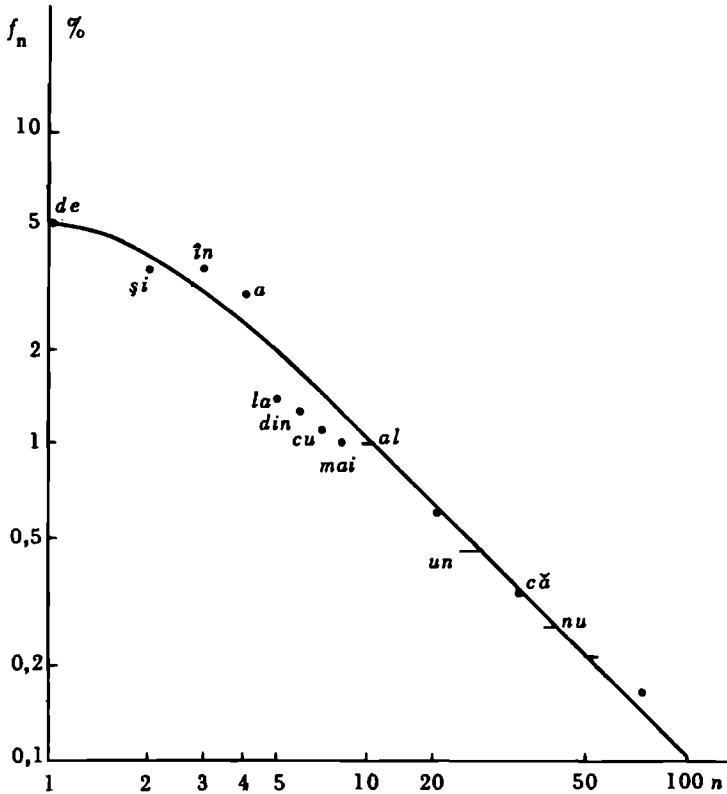


Figura 30

En la figura 31 se representa tanto la familia de curvas correspondientes a la ley de Mandelbrot —tomando m por parámetro—, como la distribución de puntos correspondientes a la lengua rumana. Se observa una buena concordancia entre los datos teóricos y los datos experimentales para valores de n comprendidos entre 3 y 600 aproximadamente.

Para el rumano, los valores de las constantes deben tomarse así:

$$p = 0,03, \quad m = 1,5, \quad B = 1.$$

El ejemplo precedente permite comprobar que la ley de Zipf actúa también en rumano.

4. Cf. ȘUTEU, "Observații asupra frecvenței".

Es interesante subrayar el hecho de que a partir de este mismo ejemplo puede demostrarse fácilmente una de las vías normales por la cual progresa la ciencia: después de establecer empíricamente una ley tomando como base el material experimental, esta ley queda explicada teóricamente con ayuda del aparato matemático.

También es interesante mencionar que B. P. Hasdeu advirtió intuitivamente que la mayor potencia circulatoria se registra en especial entre las palabras más cortas. Para ello dio algunos ejemplos del inglés, pero el fenómeno vale asimismo para el rumano [y para el español].

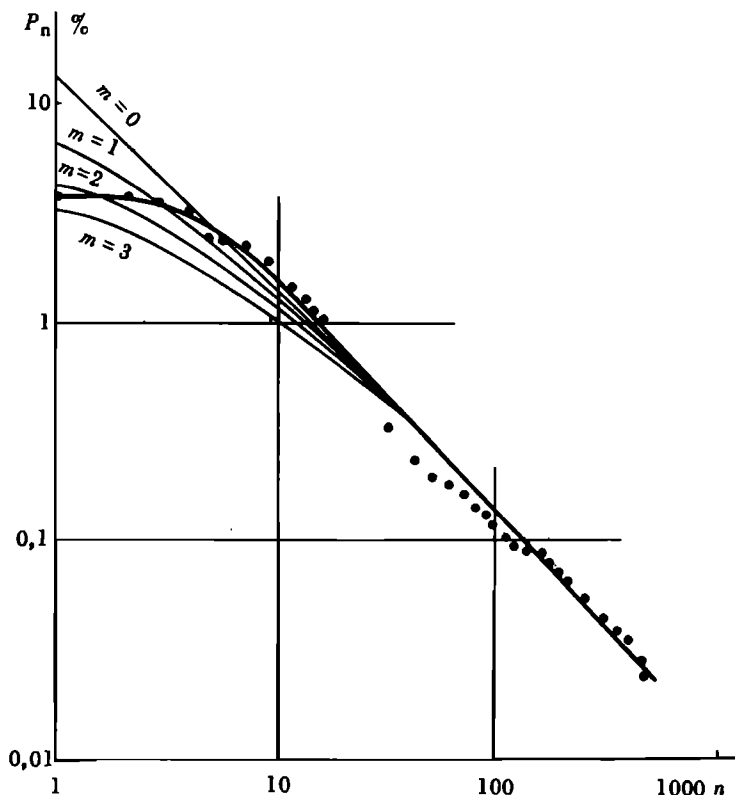


Figura 31

A medida que han evolucionado las técnicas para la elaboración electrónica de los datos, se han incrementado asimismo los métodos para el estudio del vocabulario de diferentes lenguas con ayuda de calculadores. Para el ruso hablado, N. P. Vakar utilizó en realidad el texto de ciertas obras dramáticas admitiendo, de paso, que estos textos transcribían bastante fielmente la lengua hablada⁵.

A título informativo, vale la pena presentar algunos detalles sobre el análisis estadístico del ruso. Las primeras investigaciones las llevó a cabo Harry H. Josselson, en 1953. Allegó 506 044 palabras a partir de 134 fragmentos literarios distribuidos sin uniformidad durante un período de 118 años (1830-1948). El lote analizado comprendía, porcentualmente, un 59% de prosa, 7% de drama, 20% de periodismo y 14% de crítica literaria. Steinfeldt, por su parte, reunió 400 000 palabras de 350 muestras obtenidas de la literatura infantil soviética (75%) y de novelas y piezas teatrales (25%), algunas de ellas también para niños. Es interesante y significativo notar que, comparando las primeras 100 palabras de las listas compuestas por estos dos investigadores, tan sólo 35 son comunes, lo que demuestra cuán importante es precisar el corpus que va a analizarse. Vakar utilizó, a su vez, un lote de 200 actos, elegidos entre 93 piezas de teatro. Ningún acto contenía menos de diez páginas impresas. Como control, se tomó además otro lote de 5 000 palabras a partir de muestras de 100 palabras extraídas de la página 8 a lo largo de 50 actos.

Cuando se realiza una experiencia de este tipo sobre el análisis estadístico del vocabulario hablado, es imprescindible definir las unidades básicas, en este caso las palabras. Entre lexicógrafos y gramáticos no hay consenso sobre todas las formas. Por ello, incluso distintos diccionarios de una misma lengua insertan palabras diferentes bajo una sola entrada o viceversa, como en el caso de *davat*, *davatsia*, *dat* y *datsia*, que pueden aparecer bajo una o cuatro rúbricas en distintos diccionarios oficiales de la lengua rusa.

Los estudios de Vakar han logrado establecer las primeras 360 palabras del vocabulario común, lo que permite evaluar los límites en que varía un tal vocabulario utilizado en una conversación corriente.

Por lo que se refiere al francés, J. M. Cotteret y René Moreau han analizado el vocabulario del general De Gaulle⁶. El experimento se realizó sobre un corpus de 62 471 palabras pronunciadas por aquél durante el período que va de 1958 a 1965. Se consideró como palabra todo grupo de letras comprendidas entre dos espacios ortográficos y, con ello, se omitió el análisis automático de las formas, su tratamiento ulterior, etcétera. Evidentemente, la definición adoptada es muy clara y admite una fácil aplicación por parte del calculador, pero es también discutible. Un análisis posterior, efectuado sobre una quinta parte de los discursos tratados, demostró que la relación entre el número de las formas flexivas y el número de las palabras es bastante estable, con una variación entre 1,17 y 1,23 y un valor medio de 1,19. El vocabulario establecido por el calculador contenía 6 195 palabras-forma, lo que correspondía a más de 4 000 palabras.

[Las distintas experiencias llevadas a cabo durante los últimos años han per-

5. Cf. VAKAR, *A Word Count*.

6. Cf. COTTERET-MORFAU *Vocabulaire*.

mitido poner de manifiesto ciertos problemas que exigen una discusión general de los criterios metodológicos. En primer lugar, el de la identificación, ya apuntada, de la palabra como entidad básica del recuento. Para hacernos una idea algo más aproximada del asunto, imaginemos el peligro que se corre, sin una delimitación semántica previa, de homologar conjuntamente homonimias tales como *este*, en oposición a *ese* y *aquel*, y *este*, en oposición a *oeste*, cuyos índices frecuenciales difieren, sin duda, muy ampliamente. Como se sabe (cf. 2.20, 8.2, 8.16), no hay lengua natural que no plantee serios inconvenientes en la relación, e interferencia, de sus distintos planos de análisis gramatical.

Otra fuente de problemas deriva de la selección de los materiales a examinar, donde concurren factores espaciales (variantes dialectales), socioculturales (coloquialismos, expresiones obscenas, etc.), temporales (periodos largos o cortos, con posibles cambios culturales bruscos), así como otros que presentan igualmente una gran incidencia en el resultado final. Entre ellos, p. ej., el de la búsqueda de una variedad estilística cabal dentro de lo que cabe llamar el uso corriente de la lengua, muy difícil de establecer, por cierto, cuando se opera sólo con materiales escritos. Piénsese, a este propósito, en los géneros literarios y en la dificultad de su distinción: poesía, cuento infantil o narrativa de ficción científica, para citar tan sólo algunos especialmente capaces de plantear dificultades cuando menos semánticas. Y junto a la acotación cualitativa de las muestras hay que aludir asimismo al problema correlativo de su acotación cuantitativa, importante tanto para la economía del proceso como para la fiabilidad de sus conclusiones].

Por lo demás, los resultados sobre distintos ámbitos lingüísticos han venido a demostrar que no siempre se aplica la ley de Zipf (cf. 9.4). Cuando se opera con textos variados, sí se cumple. Pero si, p. ej., en el análisis del ruso nos limitamos a los pronombres, los datos de Vakar permiten deducir que es mucho más correcta la ley

$$f_n = A 2^{-kn}$$

que nosotros mismos hemos indicado para la frecuencia de las letras en los textos o en el habla⁷. El mismo resultado se obtiene para los textos de De Gaulle. Por el contrario, el amplio vocabulario que proponen los soviéticos para los textos de electrónica, escritos en inglés, nos conducen a la ley de Zipf.

9.6. La longitud media de las palabras

Las leyes estadísticas también funcionan, desde luego, en otros niveles lingüísticos. Por ello, cabe estudiar, como decíamos antes, el problema de la longitud

7. Cf. NICOLAU, *Omni informacional*.

de las palabras, expresando esta longitud en número de sílabas. En esta dirección discurrían las primeras investigaciones de Chebanov, que demostraron que, en las lenguas indoeuropeas, la distribución de las palabras según el número de sílabas tiene lugar de acuerdo con la ley de distribución estadística de Poisson⁸. Estudiando un gran número de textos, Chebanov determinó la frecuencia $P(n)$ de aparición de las palabras con n sílabas. Mostró que si se nota mediante i la longitud media de la palabra, obtenida por la fórmula

$$i = \sum nP(n),$$

entonces, entre $P(n)$ y la longitud media subsiste la relación

$$P(n) = \frac{\nu^{n-1}}{(n-1)!} e^{-\nu}$$

donde

$$\nu = i - 1.$$

Es sabido que la distribución de tipo Poisson es un caso límite de la distribución binomial y que también se denomina ley de los eventos raros.

Examinando 127 lenguas diferentes de la familia indoeuropea, Chebanov obtuvo en todos los casos resultados semejantes. Más aún, demostró que, en general, desde este punto de vista, la estructura estadística ha variado muy poco en el curso del tiempo. Así, atendiendo especialmente una gran cantidad de textos en alemán, a partir de la Biblia de Ulfilas, comprobó que la estructura estadística general de la lengua se ha modificado muy poco a lo largo de 1200 años. El mismo Chebanov advirtió que esta misma estabilidad en la estructura estadística aparece también en algunos dialectos franceses.

Idéntica propiedad estadística ha sido confirmada por estudios minuciosos sobre la obra *Ab urbe condita* de Tito Livio, tanto al considerarla en su totalidad como en párrafos relativamente cortos del orden de algunos centenares de palabras.

También para el rumano existen investigaciones análogas sobre textos escritos en verso y en prosa y de índole tanto literaria como científica, cuyos resultados se indican en el cuadro 26.

Para los textos científicos se han escogido materiales sin fórmulas matemáticas, ya que éstas son difíciles de interpretar desde nuestro punto de vista. Por ello, se ha recurrido a textos gramaticales o jurídicos (una 'Guía de puntuación' y el 'Código civil', respectivamente).

En cuanto a los textos literarios, el análisis se ha verificado sobre *Luceafărul*, 'El lucero', de Eminescu y algunos poemas de Arghezi, para la poesía, y sobre *Hanu-Ancuței*, 'La venta de Ancutsa', de Sadoveanu, para la prosa.

8. Cf. CHEBANOV, "O podchineni".

Autor y texto	Número de sílabas						Número total de sílabas
	1	2	3	4	5	6	
EMINESCU, <i>Luceăfarul</i>	58,0	26,3	11,9	3,4	0,3	—	1 255
ARGHEZI, <i>Poezii</i>	52,5	30,0	13,6	3,1	—	—	1 914
SADOVEANU, <i>Hanu-Ancuței</i>	45,5	28,4	18,6	6,68	0,70	—	388
*** <i>Îndreptar de punctuație</i>	36,8	21,5	22,2	14,1	5,0	1,1	1 137
*** <i>Codul civil</i>	40,0	22,2	16,0	15,8	4,85	—	1 077

Cuadro 26. Frecuencia porcentual de las palabras con diferente número de sílabas

Hay que notar, en primer lugar, el hecho de que, en general, puede decirse que los textos literarios ofrecen resultados que corresponden plenamente a los que anteriormente obtuvo Chebanov. Para los textos poéticos, y en concreto para los dos casos estudiados, la longitud media de las palabras equivale a 1,6 sílabas aproximadamente. En *Hanu-Ancuței*, este valor se eleva a 2,15 sílabas, semejante al del cuento *Soacra cu trei nurori*, 'La suegra con tres nueras', de I. Creangă⁹.

Para los textos científicos, sin embargo, la longitud media aumenta hasta 2,2 para el *Codul civil*, y hasta 2,34 para el *Îndreptar de punctuație*.

La diferencia entre los valores extremos resulta bastante grande. El mismo Chebanov había observado ya que, dentro de la misma lengua, las palabras acusan una longitud variable; p. ej., en ruso, esta longitud oscila entre las 2,4 y las 2,75 sílabas por palabra.

Ahora bien, en relación con los textos analizados, cabe hacer otra observación aún más interesante. En efecto, los puntos representativos que corresponden a los textos literarios se sitúan relativamente bien sobre las curvas teóricas de Chebanov. En cambio, en los textos científicos surge una anomalía. Así, el punto representativo que corresponde a las palabras de una sílaba aparece situado sobre la curva que pertenece a las palabras de dos sílabas. En la figura 32, donde los puntos van acompañados por la inicial del autor o del texto respectivo, se aprecian estos desplazamientos que afectan, asimismo, a los puntos correspondientes a las palabras de dos y tres sílabas.

Cabe extraer una conclusión inmediata en el sentido de que la lengua literaria, y en especial la poética, emplea, por lo general, palabras más cortas que las que aparecen en los textos científicos, cosa, por lo demás, normal. En la lengua

9. [En unas experiencias similares, realizadas por la Srta. K. Biermann en Barcelona, sobre textos literarios en prosa escritos en inglés (E. HEMINGWAY, *For Whom the Bell Tolls*), en castellano (P. BAROJA, *César o nada*) y en catalán (M. RODOREDA, *La Plaça del Diamant*), se obtuvieron valores medios de 1,15 - 1,89 y 1,53 sílabas por palabra, respectivamente].

poética se utilizan, en primer lugar, palabras del fondo léxico básico de la lengua, y como éste está formado por palabras de gran circulación, tales palabras tienen, por lo común, una longitud muy pequeña.

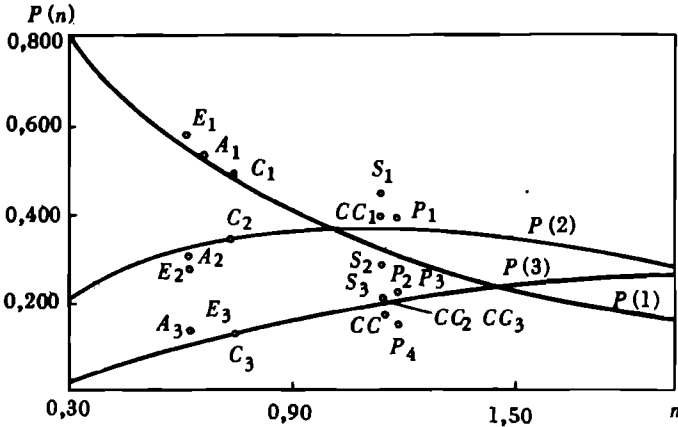


Figura 32

En los textos científicos, por el contrario, aparecen conceptos especiales, ajenos al fondo léxico básico. Como las combinaciones fonemáticas compatibles con las leyes de la lengua han sido utilizadas para la creación de las palabras del fondo básico, las nuevas palabras del vocabulario científico han de resultar forzosamente más y más largas. Este resultado manifiesta una vez más la estricta relación que existe entre lingüística y estrategia: como en el lenguaje corriente es necesario que las informaciones se transmitan lo más rápidamente posible, de un modo espontáneo el fondo básico se compone especialmente de palabras cortas.

La falta de concordancia entre los resultados experimentales y los que se desprenden de la teoría de Chebanov, en el caso de los textos de prosa científica, demuestra que las leyes estadísticas que actúan en estos textos no coinciden con las leyes que actúan en el lenguaje normal y en el poético, debido a que la estrategia adoptada en la elección de las palabras es distinta.

En estos estudios se ha calculado la longitud de las palabras sobre el número de sílabas y no sobre el número de letras, por cuanto la sílaba constituye una unidad fonética dotada de una individualidad precisa, mientras que el fonema representa una unidad meramente teórica en el marco de la sílaba. A este respecto, las vocales aparecen como casos particulares de sílabas.

9.7. Teoría de Fucks

Las diferencias que acusan el lenguaje literario y el científico son tratables en el seno de otra teoría, elaborada por W. Fucks¹⁰.

Este autor postula una tesis según la cual toda lengua viene representada por un determinado punto representativo, cuya posición en el plano está determinada por dos coordenadas cartesianas: una que corresponde a la longitud media de la palabra, expresada en número de sílabas, y la otra que representa la entropía determinada a partir de la frecuencia de aparición de las palabras, según el número de sílabas.

De acuerdo con esta teoría, todos los puntos que representan las distintas lenguas se inscriben sobre una curva regular, de tal manera que toda lengua resulta caracterizada, en rigor, por un solo parámetro: el número medio de sílabas de sus palabras respectivas.

En lo que atañe al rumano, hay que reconocer que, en general, los puntos representativos que se refieren a los casos que hemos analizado se sitúan bastante correctamente sobre esta curva (cf. fig. 33), de modo que, desde este punto de vista, puede afirmarse que la teoría de Fucks también describe con acierto los fenómenos lingüísticos de la lengua rumana.

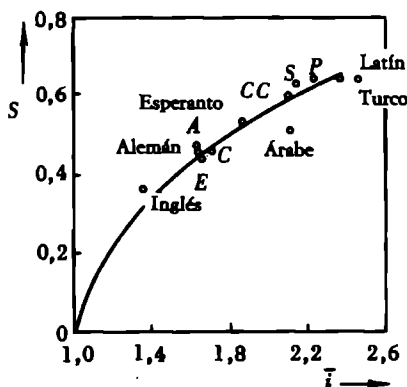


Figura 33

Pero es evidente que toda la riqueza de fenómenos correspondientes a una lengua no puede quedar reproducida por un solo punto en un diagrama.

Nótese que para el rumano, los puntos correspondientes a los distintos tipos textuales se agrupan en dos zonas en el diagrama de Fucks: una que pertenece a los textos literarios (Arghezi = (A), Creangă = (C), Eminescu = (E)), y la otra,

10. Cf. Fucks, "Mathematical theory".

a los textos científicos y, en general, escritos en prosa (Código civil = (CC), Guía de puntuación = (P) y Sadoveanu = (S)).

Mercede indagar una interpretación estratégica para la ley de Fucks. Conforme a esta teoría, la entropía aumenta continuamente en consonancia con la longitud media de las palabras, medida en sílabas. El aumento de la entropía indica una mayor indeterminación en el conjunto de las palabras clasificadas según el número de sus sílabas. El paso a un conjunto de palabras con una longitud media mayor se corresponde con una mayor posibilidad de elegir una palabra de una determinada longitud y, por tanto, corresponde implícitamente a una información media mayor.

También con respecto a la aparición de los distintos fonemas o letras en un texto dado cabe emprender investigaciones análogas.

9.8. Leyes estadísticas relativas a letras y fonemas

Puede intentarse, en primer lugar, disponer una ley semejante a la de Zipf, que actúe sobre fonemas y letras.

Las investigaciones efectuadas en este sentido han llegado a la conclusión de que, para las letras y fonemas, puede enunciarse la ley:

$$f_n = A2^{-nk}, \quad (1)$$

donde f_n es la frecuencia de aparición de la letra que ocupa el rango n en el conjunto de las letras ordenadas según los valores decrecientes de su frecuencia de aparición; A y k indican dos constantes características de la lengua considerada. Evidentemente, el número 2 no tiene un valor esencial para esta ley, ya que sólo se refiere a que la evaluación de la información se expresa en bits.

Esta ley se ha verificado para las siguientes lenguas: rumano escrito (fig. 34), francés escrito (fig. 35), y hablado (fig. 36), italiano (fig. 37), español (fig. 38),

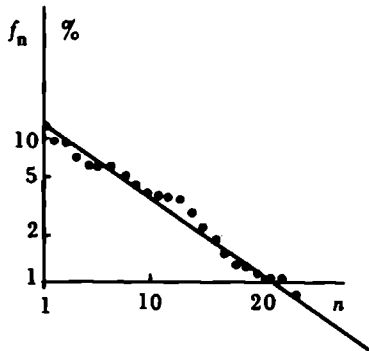


Figura 34

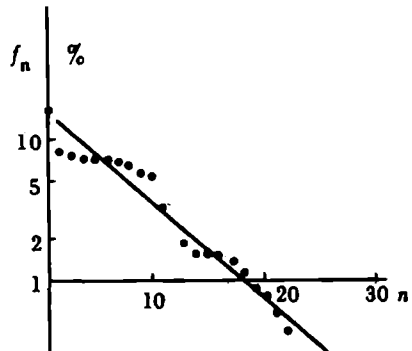


Figura 35

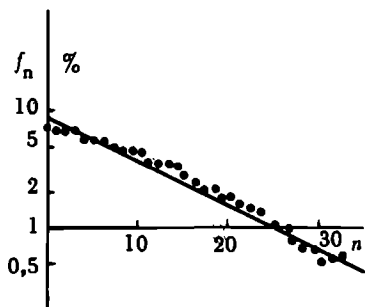


Figura 36

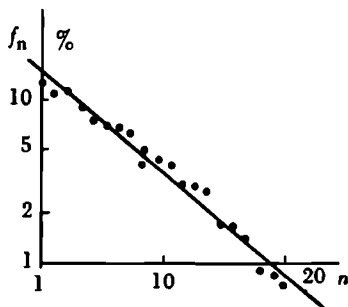


Figura 37

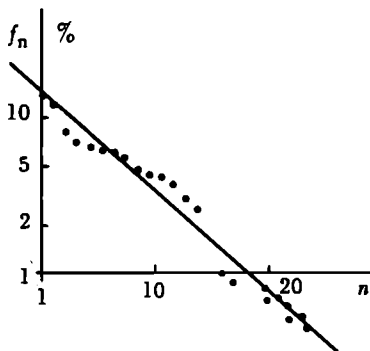


Figura 38

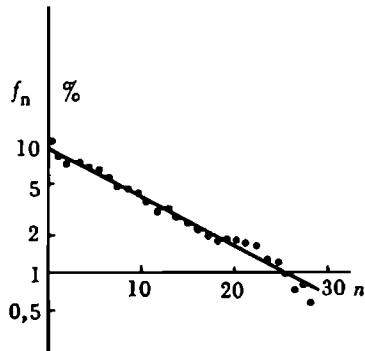


Figura 39

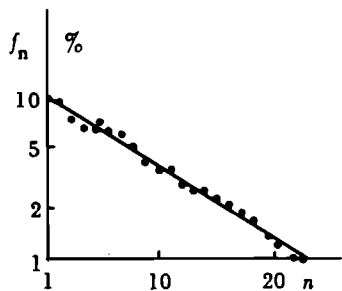


Figura 40

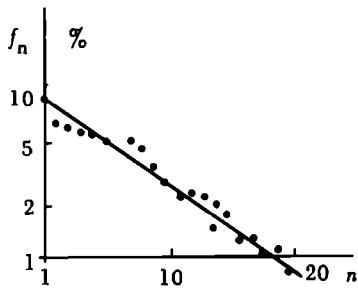


Figura 41

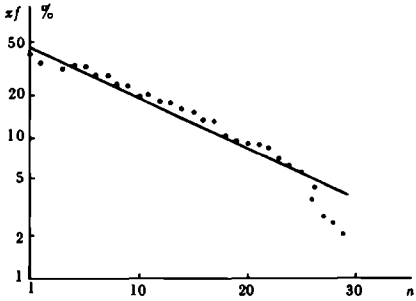


Figura 42

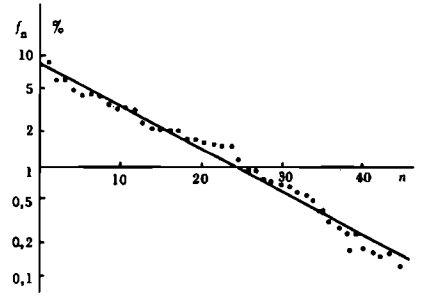


Figura 43

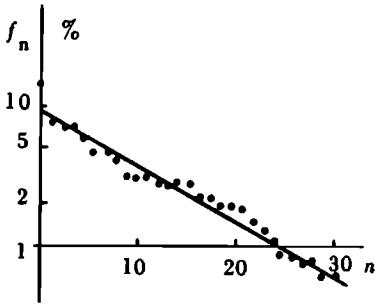


Figura 44

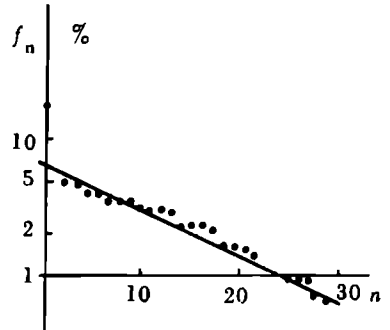


Figura 45

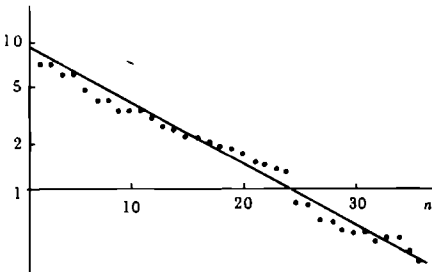


Figura 46

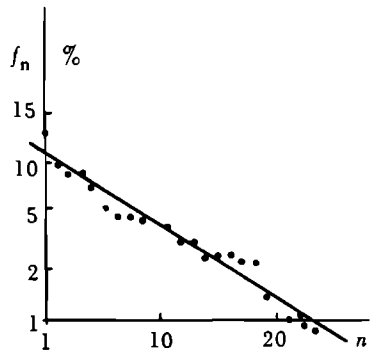


Figura 47

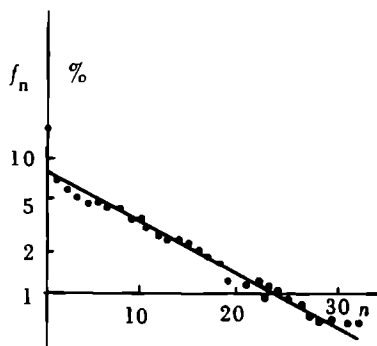


Figura 48

ruso (fig. 39), alemán (fig. 40), inglés escrito (fig. 41) y hablado (fig. 42), hindi (fig. 43), malayo (fig. 44), marathi (fig. 45), telego (fig. 46), tamil (fig. 47), kannada (fig. 48), sueco, búlgaro, checo, hebreo, griego clásico, hawaiano, portugués, flamenco, servocroata, etc.

Todas estas figuras representan gráficamente las frecuencias de aparición para algunas de las lenguas mencionadas, donde la línea continua indica la ley teórica propuesta.

Se observa que en las lenguas que no emplean una escritura fonética —como, p. ej., en francés y en inglés— la concordancia entre la ley y los datos de observación es sensiblemente mejor cuando el gráfico se refiere a fonemas que a letras. Esta ley puede verificarse también cuantitativamente observando que, si nos remitimos a todas las N letras de un alfabeto, puede escribirse

$$\sum_1^N f_n = 1$$

Teniendo en cuenta la ley (1) y esta relación, y añadiendo la correspondiente progresión geométrica, se halla:

$$A \frac{1 - 2^{-Nk}}{1 - 2^{-k}} 2^{-k} = 1 \quad (2)$$

Esta relación puede simplificarse atendiendo al hecho de que mientras la constante k tiene un valor pequeño, el número N de las letras del alfabeto es suficientemente grande para que el producto Nk tenga un valor mayor de 1. Se deduce, entonces, que

$$A + 1 \approx 2^k.$$

Para todas las lenguas estudiadas, esta relación es verificable con errores generalmente despreciables. Véase en el cuadro 27 los valores de A y k para algunas de las lenguas estudiadas.

Lengua	<i>A</i>	<i>k</i>
Marathi	0,075	0,096
Kannada	0,092	0,16
Francés hablado	0,10	0,13
Inglés hablado	0,10	0,137
Hindi	0,10	0,16
Malayo	0,11	0,126
Telego	0,11	0,13
Ruso	0,11	0,15
Alemán	0,12	0,16
Inglés escrito	0,12	0,17
Rumano	0,13	0,18
Tamil	0,13	0,174
Italiano	0,16	0,21
Español	0,17	0,21

Cuadro 27. Tabla de los valores de *A* y *k*

Con ayuda de la relación (1) también puede calcularse la entropía de las distintas lenguas. En este caso:

$$\begin{aligned}
 H &= -\sum f_i \log f_i = -\sum f_i (\ln A - ki) = \\
 &= -\log A \sum f_i + \sum ki f_i = -\ln A + k \sum i f_i
 \end{aligned}$$

El último término puede calcularse con una fórmula conocida, con lo que se obtiene:

$$H = -\log (2^k - 1) + k \frac{2^{-k} (1 - 2^{-Nk})}{(1 - 2^{-k})^2} - k \frac{N2^{-(N+1)k}}{1 - 2^{-k}}$$

Como *N* y *k* tienen valores semejantes para todas las lenguas, es natural que también *H* los tenga en todos estos casos.

Si se trata de interpretar la ley (1) desde un punto de vista estratégico, puede considerarse la ley estadística de tipo

$$C = -\log_2 p,$$

donde *p* es la probabilidad de aparición de una señal en un mensaje, y *C* el coste de la transmisión de esta señal a través de un código óptimo. La ley que proponemos puede adoptar la equivalencia

$$nk = -\log_2 p_n + \log_2 A.$$

Por analogía, resulta que el término *nk* mide el coste de la producción del fonema o de la letra respectiva.

Es probable que la ley (1) corresponda a la utilización eficiente de los códigos binarios en ausencia de ruido.

9.9. Relación entre número de palabras nuevas y número total de palabras

Otro problema estudiado consiste en hallar la relación que existe entre el número m de unidades léxicas distintas y el número total n de unidades léxicas que aparecen en un texto dado.

Examinando una cierta cantidad de textos, se ha llegado a la conclusión de que estos dos números — m y n — no son independientes entre sí, sino que guardan una relación

$$m = an^a,$$

donde a y α son coeficientes constantes.

Esta ley se comprueba perfectamente para textos que contengan entre 10 y 100 palabras, como puede observarse en las figuras anexas, referidas a un texto de divulgación científica en rumano (fig. 49), al soneto xviii de Shakespeare (fig. 50), a la *Hamburgische Dramaturgie* de Lessing (fig. 51), a *Rusland* y *Ludmila* de Pushkin (fig. 52), a *Les illuminations* de Rimbaud (fig. 53) y a textos de Rafael Obligado (fig. 54).

Por lo que atañe a textos más extensos, se ha examinado tan sólo *Hanu-Ancuței*, 'La venta de Ancutsa', de M. Sadoveanu, cuyos resultados aparecen en la figura 55. Como es natural, en estos textos largos, los coeficientes no pueden permanecer constantes.

Para el caso examinado, la ley puede presentarse en forma de:

$$\begin{aligned} \log m &= \log a_E + \alpha_E \log n \\ E &= [\log_{10} n], \end{aligned}$$

donde $[x]$ representa la parte entera de x . También esta ley ofrece, a nuestro parecer, una interpretación estratégica.

En cuanto al número de las palabras de una lengua, se plantea de un modo natural el problema de si este número es o no finito. Las primeras investigaciones parecían sugerir el hecho de que es efectivamente finito. Así, los textos corrientes empleaban alrededor de 1 500 palabras, frente a las 5 000 — 6 000 que utilizaban aproximadamente los textos literarios, a excepción, en todo caso, de Shakespeare, que usa un vocabulario de varias decenas de miles de palabras.

Hoy la situación ha variado completamente, en el sentido de que el vocabulario científico es extraordinariamente rico. Sólo en el ámbito de la química se emplean más de 1 000 000 de términos (o nombres de los distintos compuestos conocidos). De hecho, cada dominio científico dispone de un vocabulario propio bien desarrollado. Y evidentemente, todas las palabras del vocabulario científico presentan un número de sílabas sensiblemente mayor que las correspondientes al fondo léxico básico de la lengua.

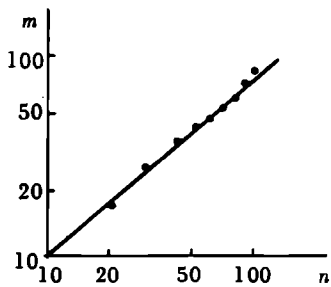


Figura 49

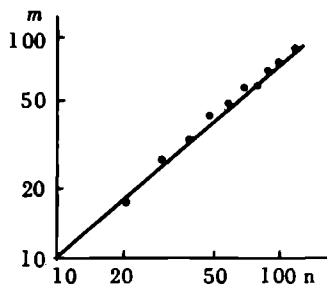


Figura 50

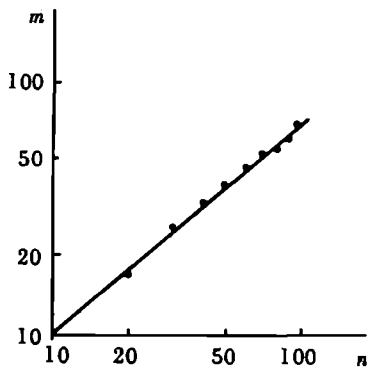


Figura 51

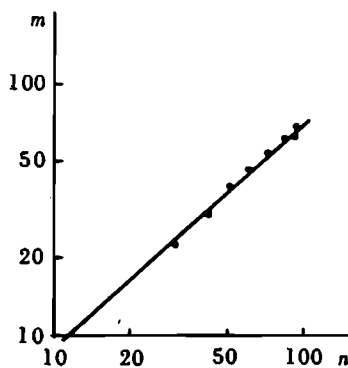


Figura 52

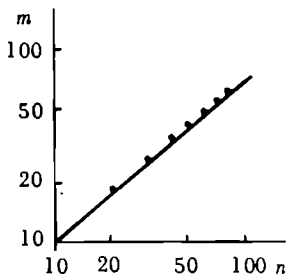


Figura 53

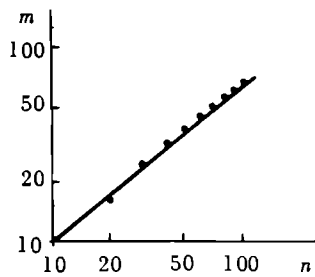


Figura 54

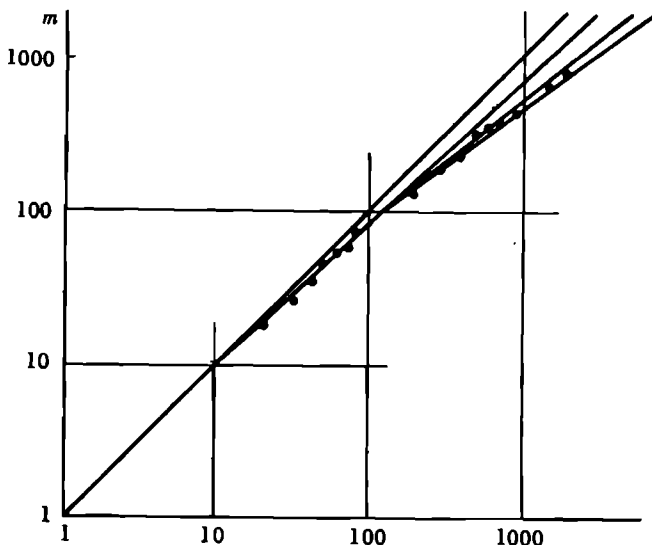


Figura 55

9.10. Métodos de correlación

Puede estimarse la entropía de una lengua considerando tan sólo la longitud n del alfabeto respectivo y admitiendo que todas las letras tienen la misma probabilidad de aparición. De este modo, se fija un límite superior para el valor de la entropía: $H_0 = \log n$. Pero se alcanza una evaluación más exacta si se tiene en cuenta la distinta frecuencia de aparición de cada letra aunque se haga abstracción de las relaciones mutuas entre ellas: $M_1 = -\sum p = \log p$.

Ahora bien, a medida que se tienen en cuenta estas relaciones también aumenta la precisión para determinar la entropía. Cuanto mayor es el número de letras consideradas en sus relaciones mutuas, tanto más disminuye el valor de la entropía, debido a que el texto en cuestión aparece más organizado. Así, pues, este tipo de indagación referido a la entropía se halla estrictamente vinculado a la necesidad de disponer de tablas estadísticas sobre las frecuencias de aparición de los grupos de letras.

Para el rumano, C. Sala ha determinado las frecuencias de aparición en grupos de dos letras en diversos textos. En la bibliografía especializada se advierten, por lo común, tanto las frecuencias absolutas de aparición, como la frecuencia relativa de asociación de una letra con las demás¹¹.

11. Cf. NICOLAU-SALA-ROGERIC, "Observații".

Respecto a la determinación del valor de la entropía, conviene retener el hecho de que, en teoría, es necesario utilizar grupos de letras cada vez más amplios hasta llegar a un grupo de longitud m tal que el paso al grupo de longitud $m + 1$ modifique el valor de la entropía en una magnitud más pequeña que la que se haya fijado anteriormente como límite. Para determinar la distancia máxima en la que se hace sensible la relación entre letras, puede recurrirse a diversos procedimientos. Uno consiste en analizar los textos y en establecer la frecuencia de aparición de los grupos de m letras. Otro se basa en la correlación.

Sobre un texto dado, cabe asociar a cada letra un determinado valor numérico, sea v_i , donde i es el rango que ocupa la letra respectiva en el alfabeto en cuestión. Un texto se presentará, pues, en forma de $\prod_1^N v_{ij}$, siendo N su longitud.

Para determinar la función de correlación del texto, correspondiente a un desplazamiento de d letras, se considera el sistema:

$$\begin{matrix} v_{i_1}, & v_{i_2}, & \dots, & v_{i_d}, & \dots, & v_{i_j} \\ v_{i_1}, & v_{i_2}, & \dots, & v_{i_{1+d}}, & v_{i_{2+d}}, & \dots, & v_{i_{j+d}} \end{matrix}$$

después de lo cual se efectúan los productos sobre la vertical, empezando por v_{i_d} y continuando hasta v_{i_N} . Se considera luego la suma

$$C_d = \sum_{i_d}^{i_N} v_{ij} v_{ij+d}.$$

Para determinar la distancia máxima en la que aún se manifiesta la relación entre letras, hay que determinar el valor mínimo de d , para el cual el producto de correlación C_d tiene un valor independiente de d , sea LC , donde $L = N - d$ es el número de los productos de correlación.

Pero C puede obtenerse también mediante recursos teóricos. Si se consideran las letras que tienen probabilidades de aparición independientes entre sí, entonces, después de una letra dada de valor v_i , puede aparecer cualquier otra letra, con el valor v_j . En este caso, en el producto de correlación las letras pueden agruparse así:

$$C = \sum_{i=1}^n p_i v_i \sum_{j=1}^n p_j v_j.$$

Ahora bien, $\sum p_j v_j$ tiene un valor bien determinado. A su vez, las primeras letras pueden agruparse según sus valores, por lo que se llega a

$$C = \left(\sum p_i v_i \right)^2.$$

Mencionemos el hecho de que el problema de la autocorrelación también plantea otro problema matemático, relacionado con la posibilidad de determinar aquella distribución de valores v_i que, atendiendo a las probabilidades de aparición p_j , lleve a la más rápida convergencia del producto de correlación.

9.11. Lenguaje científico

A diferencia de los textos literarios, los textos científicos se caracterizan por el uso de fórmulas especiales. En ciertos casos, se trata de fórmulas matemáticas, pero pueden encontrarse también fórmulas químicas, esquemas electrónicos, diseños de aparatos, etcétera. Estos elementos presentan, sin duda, un valor informacional que debe determinarse.

Desde el punto de vista estadístico, el problema de los textos matemáticos requiere que se establezca con precisión cuáles son los elementos funcionales mínimos que intervienen en las diversas fórmulas consideradas o, dicho en otros términos, hay que sentar el alfabeto empleado para la escritura de los textos en cuestión. Evidentemente, una fórmula matemática puede descomponerse en signos elementales, de forma que siempre es factible evaluar la frecuencia de aparición y la entropía de un texto matemático dado.

A partir del análisis estadístico de unos textos de lógica matemática, se han extraído los resultados que se indican en el cuadro 28.

A 16,19	B 13,8	· 7,6	⊃ 7,0	⊃ 5,8	∨ 5,6
¬ 16,6	⊢ 7,6	⊂ 7,0	~ 6,3	& 5,7	

Cuadro 28. Frecuencia de los signos de un texto de lógica matemática

La fórmula de la entropía da, en este caso:

$$H_0 = 3,4594 \qquad H_1 = 2,6355$$

lo que supone una redundancia relativa de

$$R = 23,8 \%$$

Es difícil, sin embargo, interpretar este tipo de resultados debido a que en textos matemáticos de índole diferente se utilizan alfabetos al menos parcialmente distintos. En los textos corrientes de análisis aparece muy a menudo el signo “=” y la letra *d*. En trabajos de lógica matemática no aparecen; en cambio, surgen otras notaciones: \vee , $\&$, \neg , \sim , \vdash , que nunca aparecen en los textos de cálculo diferencial e integral. Queda, pues, en pie el problema de determinar el valor informacional de las fórmulas matemáticas.

* * *

Para ciertas disciplinas, los esquemas tienen un valor excepcional. Si bien existen tratados de mecánica sin figuras, no cabe imaginar, en cambio, un tratado de electrónica sin esquemas. Se plantea, entonces, el problema de determinar el valor informacional que comportan los esquemas de este tipo.

Un primer método posible consiste en considerar los esquemas como si estuvieran formados por elementos: dispositivos electrónicos de circuito (tubos electrónicos, diodos, transistores, etc.), resistencias, bobinas, condensadores, y así sucesivamente. Según esto, y allegando un gran número de esquemas, es posible calcular las frecuencias de aparición de los distintos elementos y, partiendo de aquí, las entropías de cada esquema.

En el cuadro 29, se ofrecen los resultados obtenidos del examen de una cantidad de esquemas de rectificadores estabilizados con elementos semiconductores. Se comprueba que:

$$H_0 = 3,1699,$$

$$H_1 = 2,3842,$$

$$R = 24,9\%$$

Cabe, no obstante, la posibilidad de hacer un estudio basado en otros criterios, considerando en primer lugar todos los esquemas posibles formados por un número dado n de elementos.

El método más generalizado consiste en hacer abstracción total de la índole de los elementos a fin de considerar tan sólo el número de bornes que aparecen en el montaje. En otras palabras, se ignoran los elementos como tales desde el momento que se hallan en el esquema sólo por sus bornes. Luego se consideran todas las conexiones posibles entre estos elementos. Y como cada conexión une dos bornes, resulta que en un esquema en que existan n bornes hay un máximo de C_n^2 conexiones.

Ahora bien, como que cada una de estas conexiones puede existir o no, cabe decir que cada una de ellas puede tener dos valores posibles. Resulta, entonces, que el número máximo de esquemas eléctricos que se pueden formar, partiendo de un número de elementos que presenta en total n bornes, es

$$N_{Max} \equiv 2^{C_n^2}.$$

<i>R</i>	33,25	<i>L</i>	2,12	<i>Tr</i>	14,95	<i>S</i>	4,12	Bat	0,4
<i>C</i>	15,78	<i>D</i>	28,2	StV	1,77	Re	0,2		

Cuadro 29. Tabla de la frecuencia de los símbolos en esquemas de rectificadores de tensión con semiconductores. (*R* = resistencias, *C* = capacidad, *L* = bobina, *D* = diodo, *Tr* = transformador, StV = rectificador de tensión, *S* = fusible, Re = relé, Bat = batería)

Este valor representa el límite superior para el número de esquemas que pueden formarse a partir de m elementos que tienen en total n bornes.

Si se admite que todos estos esquemas son equiprobables, la cantidad de información que comporta cada uno de ellos equivale a

$$H_0 = \log_2 N_{Max} = C_n^2 = \frac{n(n-1)}{2} .$$

Para ilustrar este punto, consideremos un oscilador LC como en el esquema principal de la figura 56a, donde no se han tenido en cuenta todos los elementos que intervienen en el esquema real sino tan sólo los más necesarios para la comprensión del principio por el que funciona el montaje. En este esquema aparece un tubo electrónico (3 bornes), dos bobinas (2×2 bornes) y un condensador (2 bornes). En total, el esquema comprende $3 + 4 + 2 = 9$ bornes:

$$C_9^2 = \frac{9 \times 8}{2} = 36$$

y, por tanto,

$$H_0 = 36 \text{ bits.}$$

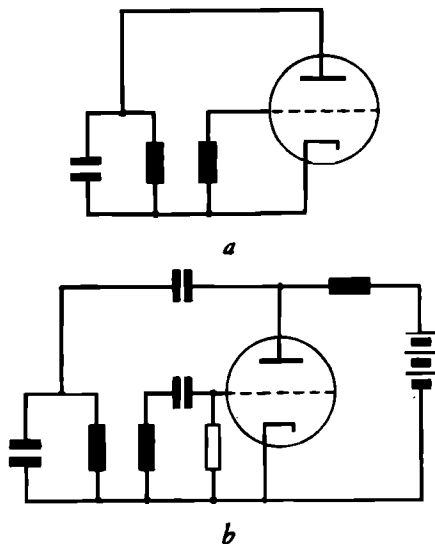


Figura 56

Pero el esquema real contiene más elementos. Para nuestro oscilador hay que recurrir a la figura 56b, donde aparecen los siguientes elementos: un tubo electrónico (3 bornes), tres bobinas (3 × 2 bornes), tres condensadores (3 × 2 bornes), una batería (2 bornes) y una resistencia (2 bornes). En total, aparecen en el esquema 3 + 6 + 6 + 2 + 2 = 19 bornes; es decir $C_{19}^2 = \frac{19 \times 18}{2} = 171$. Así, el valor de H_0 se eleva a

$$H_0 = 171 \text{ bits.}$$

Este valor entrópico es muy grande. En rigor, teniendo en cuenta la naturaleza real de los elementos, muchas de las conexiones posibles desaparecen por inútiles, de modo que hasta el número mismo de los esquemas posibles disminuye considerablemente.

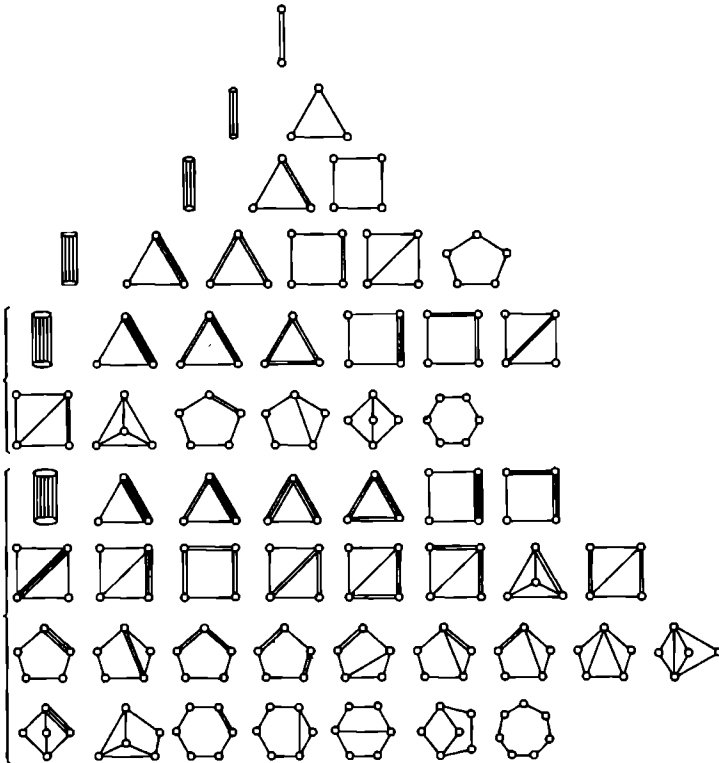


Figura 57

Un sencillo ejemplo aclarará esta idea. Sean dos elementos bipolares, cada uno de ellos con dos bornes. En total se distinguen cuatro bornes y, por consiguiente, el número máximo de esquemas imaginables equivale a

$$2C_4^2 = 2^6 = 64.$$

Pero, en realidad, si se consideran las conexiones posibles, se llega a un solo circuito posible, justamente al circuito en serie.

Los distintos esquemas posibles, de acuerdo con los principios generales enunciados en lo que sigue, quedan expuestos en la figura 57.

En caso de considerar elementos direccionales —como, p. ej., los diodos—, el número de montajes posibles crece, tal como se desprende de la figura 58.

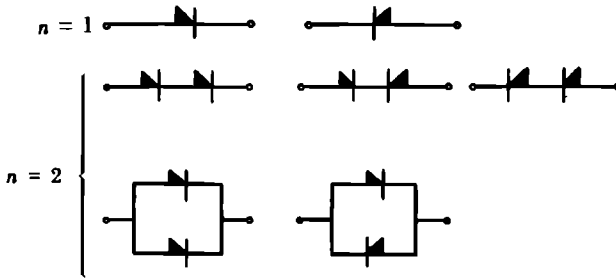


Figura 58

Evidentemente, en todo cuanto venimos exponiendo se ha examinado tan sólo el caso de los esquemas cerrados y compuestos por un total de n bornes. Pero el problema puede plantearse también para esquemas que presenten una cierta cantidad de bornes de acceso: dipolos, tripolos, cuadripolos, etc.

Si se considera una red con m nódulos, entonces desde un borne exterior pueden establecerse m conexiones con esta red. Para p bornes exteriores son posibles m^p conexiones de este tipo, es decir

$$N_{Max} = m^p$$

redes con p bornes de acceso, teniendo el resto de la red m nódulos.

Hay que precisar que, en este caso, entendemos por nódulo el punto donde se encuentran diversos elementos de la red o la extremidad libre de un elemento que tenga relación eléctrica con la red. Una resistencia sólo puede tener un borne libre, pero una bobina puede tener dos, en una red.

Estos resultados son aplicables a determinados casos reales. Consideremos, en primer lugar, una red formada por varias categorías de elementos, cada uno

de ellos con dos bornes: es el caso de las redes pasivas RLC. Entre los elementos de un esquema de este tipo pueden establecerse diversas conexiones. El número máximo de esquemas formables con elementos que tengan en total $2n$ bornes es

$$N_{Max} = 2^n(2^n - 1).$$

Pero, en realidad, el número de esquemas distintos que cabe formar con elementos no directivos es mucho menor. En la figura 57 se representan los esquemas formados por un máximo de siete elementos indicados por medio de líneas.

Volviendo al N_{Max} conviene observar que en el caso de los esquemas eléctricos intervienen tres categorías de elementos, pues en cada lado de la red puede aparecer una resistencia, una bobina o un condensador. Se averigua, así, que el número total de esquemas formados por estas tres categorías de elementos y que tengan n elementos —entre los cuales no haya acoplamientos mutuos— equivale a

$$N_{Max} = 3^{2^n(2^n - 1)}.$$

Los esquemas que incluyen además tubos electrónicos o transistores pueden estudiarse del mismo modo.

Ahora bien, el problema de los esquemas electrónicos puede también estudiarse de una manera un poco diferente. En efecto, todo esquema puede ser representado por una sucesión de símbolos gráficos que indiquen las conexiones existentes entre los elementos del montaje. Así, los tubos electrónicos se representarán por medio de letras, los bornes de acceso por números, etc.

Para ejemplificarlo, observemos el esquema de la figura 59. Como sólo contiene un tubo electrónico, lo llamaremos T . En general, en un esquema donde

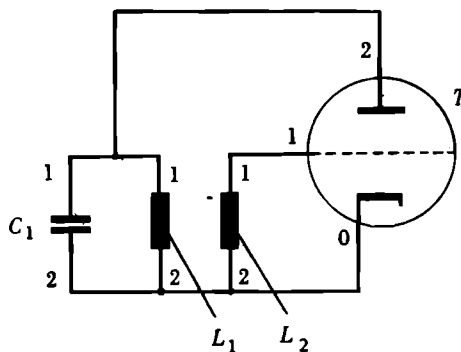


Figura 59

aparezcan n tubos, se nota cada uno de ellos a base de T_i ($i = 1, \dots, n$). La naturaleza del tubo (diodo, triodo, hexodo, etc.) se especifica por un número colocado detrás de i . Así, p. ej., si T_2 es un triodo, se escribirá T_{23} . Cabe introducir, además, la siguiente notación para los bornes del tubo T_{23} : cátodo T_{230} , rejilla T_{231} , ánodo T_{232} . De un modo análogo se consignan las inductancias (L_1 y L_2) y sus bornes (L_{11} , L_{12} , etc.). Se procede igualmente con los condensadores y las resistencias. Queda por especificar la conexión entre bornes, para lo cual basta con escribir entre paréntesis los bornes conectados entre sí. Por ejemplo, la conexión de la rejilla del tubo T_1 con el borne 1 de la bobina L_1 se escribe

$$(T_{131}, L_{11}).$$

En cuanto a las inductancias mutuas, se procede igualmente, incluyendo dentro de un paréntesis las bobinas entre las que se forma un acoplamiento magnético y especificando mediante un bit suplementario el signo de M . Se consignará, entonces, $(L_1 L_{20})$ para un acoplamiento positivo entre L_1 y L_2 , y $(L_1 L_{21})$ para el mismo acoplamiento pero con valor negativo.

Así, todo esquema electrónico puede representarse a base de una sucesión de letras, cifras y paréntesis. De modo que, p. ej., el esquema de la figura 57 adquirirá la forma

$$\begin{aligned} &(T_{132} L_{11}) (L_{11} C_{11}) (L_{12} C_{12}) \\ &(L_{22} T_{130}) (L_{21} T_{131}) (L_{12} T_{130}) \\ &(L_1 L_2, 1). \end{aligned}$$

De esta manera, cabe expresar cualquier esquema a través de las letras de un alfabeto finito, que admite, por tanto, la transmisión a partir de señales binarias. A este propósito merece subrayarse el problema de simplificar las fórmulas que expresen esquemas: ¿con qué procedimientos es posible reducir al mínimo la fórmula de un esquema? Como el esquema va expresado a base de las letras de un alfabeto, se infiere que esta fórmula debe ser interpretable por los métodos estadísticos usuales.

A este respecto, es necesario precisar que, a veces, de los esquemas interesa tan sólo su aspecto topológico. Así sucede, p. ej., cuando se estudian las condiciones de oscilación en un oscilador o el funcionamiento de un nivel de amplificación. En los esquemas reales, conviene, sin embargo, precisar tanto los valores de las piezas como el tipo de los elementos activos de circuito utilizados. Ello se consigue fácilmente mediante tablas anexas al esquema topológico o al texto equivalente, en las cuales se indiquen los valores de las piezas, los tipos de elementos activos de circuito, etc.

También estas tablas ofrecen un valor informacional evaluable, desde el momento en que el valor de cada pieza corresponde a una opción efectuada a partir del conjunto de los valores posibles mediante un cálculo que contemple el hecho de que la opción siempre se realiza dentro de un conjunto finito cuyos elementos

disponen de valores convencionales. Así, p. ej., para una resistencia se escoge un valor del conjunto de valores convencionales, para un tubo se elige un determinado tubo de la categoría de los tubos electrónicos, y así sucesivamente.

De un modo semejante puede plantearse el problema de la entropía que se obtiene en un semáforo o, en general, en las señales de tráfico. En cuanto al semáforo hay que observar que la interpretación informacional de los resultados de la experiencia no es siempre fácil. En este caso, existen tres valores posibles para el resultado de la experiencia: rojo, verde y ámbar. Sea, entonces, un semáforo situado en la intersección de dos calles, una de ellas con tráfico menos denso. El semáforo favorecerá la circulación de la calle principal. El ciclo completo comprende las siguientes fases: rojo, ámbar, verde, ámbar y sus repeticiones sucesivas. Para la vía principal, el verde permanece iluminado cuatro unidades de tiempo, el ámbar una unidad cada vez y el rojo sólo dos unidades. Calculado en tiempo, un ciclo dura ocho unidades de tiempo. La probabilidad de aparición de un color en el semáforo viene dada por la relación entre la duración de este color y la duración total del ciclo:

$$\text{rojo } p_r = \frac{4}{8} = \frac{1}{2}; \text{ verde } p_v = \frac{2}{8} = \frac{1}{4}; \text{ ámbar de rojo a verde } p_a = \frac{1}{8}; \text{ ámbar de verde a rojo } p_a = \frac{1}{8}.$$

Resulta, entonces, que para la calle principal la entropía de este campo de hechos es igual a:

$$\begin{aligned} H &= - \sum p_i \log_2 p_i = - \frac{1}{2} \log \frac{1}{2} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{8} \log \frac{1}{8} - \frac{1}{8} \log \frac{1}{8} \\ &= 1,750 \text{ bits/señal} \end{aligned}$$

En realidad, estas señales no son independientes entre sí, ya que tras una señal roja o una verde aparece siempre una ámbar. La única indeterminación surge en el caso de una señal ámbar y sólo cuando ambas alternativas (rojo o verde) son equiprobables.

La entropía H , calculada, nos da la información media que obtenemos al considerar todas las posibilidades. Pero como los eventos producidos por el semáforo no son aleatorios, después de considerar un máximo de dos situaciones dejamos de tener información. Así, pues, H mide la información que proporciona, en promedio, la primera señal del semáforo.

* * *

El interés informacional por los textos científicos se halla apenas en sus comienzos, pero constituye un dominio donde, con seguridad, se alcanzarán resultados importantes.

10. Lengua y código

10.1. Introducción

La afirmación según la cual la lengua es un código se ha repetido hasta la saciedad tanto para apoyarla como para combatirla. Pero si examinamos con atención los argumentos de sus adversarios, comprobaremos que ninguno de ellos duda de la existencia, en la lengua, de ciertas modalidades de código, de manera que lo que más bien combaten es la pretensión de *reducir* la lengua a un código y, con ello, de disolver la teoría de la lengua en la teoría de los códigos. En efecto, la lengua es un código específico, formado de un modo natural y que ha evolucionado en estrecho contacto con el desarrollo de la sociedad. Pero esta circunstancia no sólo disminuye, sino que, por el contrario, acrecienta la importancia de los estudios que se aplican a los aspectos de la lengua que son comunes a cualquier código. La lengua constituye un proceso particular de codificación informacional, pero, para comprender su carácter específico, hay que delimitar escrupulosamente las modalidades comunes a cualquier proceso de comunicación. En este sentido se establece en qué medida se aprovecha el estudio de la lengua de los métodos y resultados de la teoría de la información y, más en general, de la cibernética. Las experiencias realizadas en esta dirección revisten un especial interés para la teoría de la lengua, por cuanto arrojan una nueva luz sobre las relaciones entre la acción de los factores internos (intralingüísticos) y la acción de los factores externos (extralingüísticos).

10.2. El proceso de codificación

Toda información discurre por lo que se llama, en teoría de la información, un *canal de transmisión*, representado por un cable, una conexión radiofónica, la voz humana, etcétera. Para transmitir una información, hay que

codificarla de antemano, es decir hay que representarla mediante determinados símbolos adaptados a las propiedades físicas del canal de transmisión que empleemos. El proceso de codificación consiste precisamente en este tratamiento sobre el modo de representar la información. El nuevo sistema de símbolos al que se pasa constituye, entonces, el código adoptado, el cual puede valorizarse tan sólo a base del conocimiento de la correspondencia entre los dos sistemas, el que representa inicialmente la información y el final.

En el proceso de codificación que se halla en la base del lenguaje, los conceptos, las ideas y, en general, toda la actividad del pensamiento humano son representados de un modo asequible al canal fonador del hombre. El proceso de codificación consiste aquí en el paso de los elementos de orden semántico a los de expresión o, en la terminología de F. de Saussure, en el paso del plano del contenido al plano de la expresión. La correspondencia entre contenido y expresión no se ha formado, sin embargo, por vía artificial, a partir de una deliberación consciente, como ocurre en muchos otros procesos de codificación, sino como resultado de una larga evolución en el marco del desarrollo social. Entre los elementos de contenido y los elementos de expresión no existe, en general, un vínculo lógico, interno, sino un vínculo necesario, resultado de un devenir histórico-social (cf. 1.1).

Completamente distintos son los códigos utilizados en telegrafía (p. ej., el código Morse), pues se deben a una construcción humana deliberada en la que se tenía en cuenta un uso bien determinado.

10.3. La noción matemática de código

Sea A un conjunto finito de elementos denominados, pongamos por caso, *letras*. A constituye un *alfabeto*. Una sucesión finita de elementos de A (de modo que un mismo elemento pueda aparecer más de una vez en la sucesión) se denomina *palabra* o *texto* en el alfabeto A . Sea $L(A)$ un conjunto de palabras en el alfabeto A y digamos que es una *lengua* en el alfabeto A .

Si E es un conjunto tal que exista una aplicación f biunívoca de E sobre $L(A)$, el trío $\{E, L(A), f\}$ define una codificación: E es el conjunto que se codifica, $L(A)$ es el código y f la regla de codificación.

Es importante el caso particular en que E sea una lengua. He aquí cómo se desenvuelve la codificación en esta circunstancia:

Sean dos alfabetos A y B . Sean $L(A)$ y $L(B)$ lenguas en los alfabetos A y B . Supongamos que existe una aplicación biunívoca de $L(A)$ sobre $L(B)$ que llamaremos f . El trío $\{L(A), L(B), f\}$ constituye una codificación, donde $L(A)$ es

la lengua a codificar, $L(B)$ es el código y f la regla de codificación.

Una sucesión finita de palabras en el alfabeto A forma un mensaje. Si no nos interesa codificar toda la lengua $L(A)$, sino tan sólo la parte que interviene en un determinado mensaje, entonces, notando mediante A_M el conjunto de las palabras que intervienen en M , consideraremos el trío $\{A_M, f(A_M), f/A_M\}$, donde $f(A_M)$ es la imagen del conjunto A_M mediante la aplicación f , y f/A_M es la restricción de f en A_M . Este trío define una codificación del mensaje M . Desde luego, puede definirse directamente una codificación del mensaje M sin considerar de antemano una codificación de la lengua en su totalidad.

Es posible que obtengamos $A = B$, de modo que la codificación mantenga el mismo alfabeto, aunque cambiando las palabras.

La codificación que realiza la lengua no satisface rigurosamente las condiciones que acabamos de exponer. En efecto, E es el conjunto de los elementos de contenido, A es el alfabeto de la lengua, $L(A)$ es su vocabulario, pero f no es una aplicación biunívoca de E sobre $L(A)$, debido, entre otras razones, a la homonimia y a la sinonimia.

El paso de la lengua hablada a la lengua escrita constituye una codificación de tipo $\{L(A), L(B), f\}$, donde A es el conjunto de los fonemas y B el de las letras. Pero tampoco aquí se cumple la biunivocidad de la aplicación f .

10.4. Los códigos binarios y su empleo en lingüística

Existe una enorme cantidad de clases de códigos. Muy utilizados son los llamados códigos binarios, donde se emplea un alfabeto compuesto sólo por dos elementos distintos, normalmente representados por las cifras 0 y 1. Un código de este tipo contiene, pues, un número indeterminado de sucesiones de forma 0 1 1 1 0 0 1 0. El número de cifras puede variar de una a otra sucesión, pero existen códigos muy importantes en que el número de cifras es el mismo para todas las sucesiones.

El uso preponderante que tienen los códigos binarios es fácil de explicar. Ocurre que muchísimas veces los canales de transmisión sólo son capaces de transmitir impulsos (eléctricos, magnéticos, acústicos, ópticos o de otra índole); en tales condiciones, la ausencia de impulso significa 0 y su presencia, 1. Así, pues, toda secuencia de impulsos / no-impulsos se refiere a una combinación de cifras 1 y 0. El mismo carácter binario aparece también en la máquina electrónica de cálculo. Una máquina de esta clase comporta, en su construcción, un gran número de tubos electrónicos, cada uno de los cuales admite tan sólo dos estados: que pase, o no, corriente por él. Podría decirse que un ordenador electrónico no comprende cualquier lengua, sino que se limita a la que emplea exclusivamente dos símbolos (0 y 1). De ahí que los datos que deban introducirse en la máquina requieran una codificación previa a base de un código binario.

Merece observarse que, en la lingüística estructural, se emplean códigos binarios en distintos problemas de fonología o de gramática a fin de expresar de un modo sintético, y sugestivo los resultados de un análisis efectuado a un superior nivel de abstracción. No se trata de una casualidad, sino de un hecho que ilustra el gran valor que tienen en lingüística las oposiciones con carácter binario. Ya Ferdinand de Saussure había llamado la atención sobre la importancia del binarismo en nuestra ciencia. En fonología, la primera concepción sistemática a partir de oposiciones binarias se debe a E. C. Cherry, M. Halle y R. Jakobson¹. Como ilustración presentamos, en el cuadro 30, el código binario que representa el sistema consonántico del inglés².

	p	b	m	f	v	k	g	t	d	θ	ð	n	s	z	c	ʃ	ʒ	ʒ
Vocálico	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Consonántico	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Grave	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
Compacto	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	1	1	1
Estridente	0	0	0	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1
Nasal	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Continuo	0	0	0	1	1	0	0	0	0	1	1	0	1	1	0	0	1	1
Sonoro	0	1	1	0	1	0	1	0	1	0	1	1	0	1	0	1	0	1

Cuadro 30

En la base del código expuesto se hallan las siguientes oposiciones binarias: vocálico / no-vocálico, consonántico / no-consonántico, grave / no-grave [o agudo], compacto [o denso] / no-compacto [o difuso], estridente / no-estridente [o mate], nasal / no-nasal [u oral], continuo / no-continuo [o interrumpido], sonoro / no-sonoro [o sordo]. La cifra 0, situada en la intersección de la columna de una consonante y la fila de un rasgo acústico indica la ausencia de este rasgo en la consonante en cuestión [o la presencia del contrario en la oposición correspondiente]. Por el contrario, la cifra 1 indica la presencia del rasgo en la consonante respectiva. Así, p. ej., la cifra 0 que aparece en la intersección de la columna de *t* y la fila del rasgo de nasalidad consigna la naturaleza no-

1. Cf. CHERRY-HALLE-JAKOBSON, "Logical description".

2. Cf. HALLE, "Role of simplicity", p. 90; [MULJAČIĆ, *Fonología*, y CHOMSKY-HALLE, *Sound Patterns*].

nasal [u oral] de *t*; la cifra 1 que se halla en la intersección de la columna de *z* y la fila del rasgo de continuidad indica la naturaleza continua de *z*, etc.

Las cifras 0 y 1 constituyen aquí el alfabeto del código, mientras que las sucesiones de 0 y 1 asociadas a las distintas consonantes constituyen las palabras del código. El número de cifras que aparecen en una palabra del código equivale a la longitud de esta palabra. Tal como vemos, cada palabra tiene una longitud de ocho en el código aducido, cuyo carácter binario se refleja en el hecho de que se compone de un alfabeto formado por dos elementos.

La descripción de Hjelmslev para el sistema casual representa, a su vez, un notable ejemplo del empleo de códigos binarios en la gramática. Más tarde, muchos otros autores han utilizado también criterios análogos en la descripción de sistemas casuales particulares³.

En la traducción automática se emplea igualmente un código binario. Del alfabeto de la lengua que se traduce se pasa al alfabeto formado por las cifras 0 y 1. Si, p. ej., se trata de la lengua rumana, a cada letra se le asigna de momento un número comprendido entre 1 y 26 (equivalente al número de letras que componen el alfabeto en rumano). Luego, se escribe con base 2 cada uno de estos números mediante sucesiones de cifras 0 y 1. Como 2^4 es menor que 26 y 2^5 , mayor, resulta que toda letra se representa a través de una sucesión formada a lo sumo por cinco cifras 0 ó 1. Así, la letra *z*, que ocupa el lugar 26 en el alfabeto, queda rescrita por la sucesión 11010, ya que:

$$26 = 1 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0.$$

Un código de este tipo, construido con base binaria, es decir que utilice dos símbolos fundamentales y forme luego combinaciones con repetición entre ellos, es conocido desde hace mucho en telegrafía con el nombre de "sistema Baudot". Como en los casos precedentes, en el sistema Baudot se emplean combinaciones (con repetición) de cinco símbolos formadas a partir de dos símbolos fundamentales, que dan lugar, pues, a $2^5 = 32$ combinaciones.

10.5. Desciframiento de un texto codificado

Volvamos, por el momento, a la representación de las letras con la ayuda de los números que van de 1 a 26.

Supongamos que se nos da un texto escrito a base de un código numérico (o cifrado). ¿Qué posibilidades hay de descifrarlo? Para contestar a esta suerte de preguntas hay que contar con la probabilidad de aparición de cada letra. Algunas investigaciones recientes han allegado las frecuencias, expresadas en

3. Cf. HJELMSLEV, "Catégorie"; PADUCHEVA, "Ob opisani padezhnoi", y SHAUMIAN, "Lingvisticcheskie voprosy".

porcentajes, que aparecen en el cuadro 31 para las 26 letras que se emplean en rumano⁴:

E	11,6	O	4,54	î	1,47	X	0,276
I	10,8	S	3,94	Ț	1,34	J	0,24
A	10,15	C	3,56	Ș	1,015	H	0,071
R	7,78	D	3,44	V	0,96		
N	6,87	A	3,43	B	0,9		
U	6,35	P	2,98	Z	0,78		
T	5,95	M	2,54	G	0,64		
L	5,04	F	1,95				

Cuadro 31

Precisamos que, en estos resultados, la frecuencia de la letra H se ha establecido sin tomar en consideración su presencia en los grupos CHE, GHE, CHI, GHI, los cuales han sido evaluados por separado con otros grupos de letras⁵. Algunos de estos resultados figuran en el cuadro 32.

CI	1,53	CHI	0,23	CI	0,22	GHI	0,134
----	------	-----	------	----	------	-----	-------

Cuadro 32

Es suficiente ahora disponer los grupos de dos cifras cada uno con arreglo a su frecuencia. La identificación de cada grupo se hará comparándolo con la tabla de antes. Así, p. ej., un grupo de rango 5 en la lista formada de este modo corresponderá a la letra N, pues N tiene este rango en la tabla del cuadro 31.

Hay que observar, sin embargo, que todo lo dicho vale sólo para cuando el texto codificado sea suficientemente largo. En este caso, las posibles discordancias pueden remediarse a base de un estudio directo, examinando el contexto.

4. Cf. CONSTANTINESCU-CONDREA-NICOLAU. *Teoria informației*.

5. [Cf. 8.17, nota 50].

Existen códigos no contruidos sobre una base binaria. Así ocurre en el código utilizado en el sistema telegráfico Morse, donde aparecen, como símbolos fundamentales, líneas, puntos e intervalos entre líneas y puntos. No es difícil demostrar, no obstante, que también este código obedece a una concepción binaria.

El problema de construir una codificación eficiente adquiere la mayor importancia para la transmisión de los mensajes lingüísticos. En efecto, un proceso de transmisión de este tipo se compone esencialmente de una operación de codificación, la operación propiamente dicha de transmisión (sometida a ciertas perturbaciones) y la operación de descodificar o de identificar el mensaje transmitido. Hay ruidos a lo largo de estas operaciones que pueden desfigurar el mensaje. Se plantea, entonces, el problema de construir códigos que no sólo posibilite el descubrimiento de los errores, sino también su corrección. Se trata, en este caso, de los llamados *códigos correctores*. En lo que sigue consideraremos algunos ejemplos de códigos que descubren y corrigen errores.

Sean 0 y 1 los símbolos fundamentales del código. Sobre ciertas sucesiones normalmente formadas a base de estas cifras supongamos que todas las sucesiones observadas se componen de un número l constante de cifras. Estas sucesiones se denominarán *palabras del código* y l será la longitud de las palabras. Así, los códigos que aducimos en lo sucesivo serán códigos binarios y con palabras de la misma longitud.

10.6. Distancia entre dos palabras

Llamaremos *distancia entre dos palabras* al número de lugares que en las palabras aparezcan con cifras diferentes. Por ejemplo, suponiendo que $l = 4$, las palabras 1011 y 1001 están a una distancia igual a 1 porque sólo difieren por la cifra que se halla en tercer lugar, mientras que las palabras 0010 y 1111 están a una distancia de 3, ya que difieren por las cifras que se hallan en los lugares 1, 2 y 4. El término "distancia" parece justificado si tenemos en cuenta las siguientes propiedades que recuerdan la distancia habitual entre dos puntos: α) la distancia de una palabra a otra es igual a 0 si, y sólo si, las dos palabras coinciden; β) la distancia entre dos palabras no depende del orden por el que se consideren ambas palabras, y γ) dadas tres palabras x, y, z , la distancia de x a y no supera la suma de las distancias de x a z y de z a y .

10.7. Códigos que descubren errores

Aunque transmitamos correctamente una determinada palabra del código, puede suceder que, a causa de las deficiencias técnicas del canal de transmisión o de

lo que, en general, se denomina ruido, la palabra en cuestión sea recibida erróneamente, como en el caso, p. ej., de que en lugar de un 0 transmitido se reciba un 1, o a la inversa. Se plantea, entonces, el problema de descubrir y de paliar los eventuales errores a que está expuesto todo mensaje.

Diremos de un código, de las características expuestas antes, que descubre un error simple (único) si, cualquiera que sea este error único, la aparición del mismo determina la transformación de una palabra del código en una palabra que no pertenezca al código. Con ayuda del concepto de distancia que hemos introducido en el apartado anterior, es muy fácil caracterizar los códigos que cumplen la condición de detectar errores únicos:

Para que un código descubra un error único es necesario y suficiente que la distancia entre dos palabras arbitrarias del código sea superior a 1.

En efecto, admitamos que un código descubra un error único y supongamos, por reducción al absurdo, que este código contenga dos palabras a una distancia igual a 1. Para fijar mejor las ideas, digamos que estas palabras son 1010 y 1011. Consideremos ahora el error que se produce al sustituir la cifra 0 en el cuarto lugar de la palabra 1010 por la cifra 1. Advertimos que este error único determina la transformación de una palabra del código en otra palabra del mismo código, en contradicción con la hipótesis.

Supongamos ahora que la distancia entre dos palabras arbitrarias del código sea superior a 1. Si una palabra x que contuviera un error único se transformara en otra palabra y del mismo código, entonces la distancia entre las palabras x e y equivaldría a 1, contra el supuesto que acabamos de asumir. Así, pues, el código descubre un error único.

Un ejemplo de código que descubre un error único estaría formado, pongamos por caso, por palabras de tipo 1111, 1001, 0110. Si, p. ej., la palabra 1111 quedara deformada a base de 1110, este resultado queda automáticamente identificado como erróneo, ya que 1110 no pertenece al código.

La proposición de más arriba puede generalizarse si se incluye la siguiente definición: un código descubre n errores si toda sucesión de, por lo menos, n errores transforma las palabras del código en palabras que no pertenezcan al mismo. Tiene lugar, entonces, la siguiente propiedad:

Un código descubre n errores si, y sólo si, dos palabras arbitrarias del código se hallan a una distancia superior a n .

10.8. Códigos correctores

Hasta aquí hemos caracterizado códigos que descubren errores. Pero, claro está, no basta descubrir los errores si no cabe corregirlos. Diremos que un código

corrige cualquier error único si un tal error determina la transformación de una palabra que pertenece al código en otra palabra que no pertenece al código, pero que se encuentra, frente a una sola palabra dada del código, más cerca de ella que de cualquier otra palabra del mismo código. Esta única palabra del código, frente a la cual la palabra errónea se halla a una distancia mínima, será precisamente la palabra correcta que debe sustituir a la errónea. Sea, p. ej., un código formado por las palabras 00101, 11001 y 10110. Supongamos ahora que la última palabra queda deformada a base de 10111, con lo que se forma una expresión que no pertenece al código. Entre las palabras que componen el código, la más cercana a 10111 es justamente 10110. En efecto, la distancia entre ambas es igual a 1, mientras que la distancia entre la palabra errónea y la palabra 00101 es de 2, y respecto de la palabra 11001, de 3. Por lo tanto, la palabra errónea 10111 debe sustituirse por 10110. Puede demostrarse que:

Un código corrige todo error único si, y sólo si, la distancia entre dos palabras arbitrarias del mismo es superior a 2.

Dejamos a cargo del lector la verificación de este principio.

Diremos que un código corrige n errores si toda secuencia de, por lo menos, n errores transforma una palabra cualquiera del código en otra palabra que no pertenezca al código, pero que se halle, a su vez, frente a una palabra dada y única del código, más cerca de ella que de cualquier otra palabra del mismo código. Puede demostrarse, entonces, que:

Un código corrige n errores si, y sólo si, la distancia entre dos palabras arbitrarias del mismo es superior a $2n$.

Es fácil comprobar que, a medida que aumentan las distancias entre palabras, el número de éstas que contiene el código disminuye o permanece estacionario. Así, p. ej., si consideramos dos códigos con palabras de longitud igual a 4 y si en el primero las distancias entre las palabras son mayores que 2 y en el segundo estas distancias son mayores que 3, entonces ninguno de los dos códigos puede contener más de dos palabras, digamos 1111 y 0000 (o bien 1111 y 0001, en el primer código, y 1010 y 0101, en el segundo). Ahora bien, un código en que las distancias entre palabras se sometan tan sólo a la restricción de ser mayores que 1 puede contener más de dos palabras; p. ej., 0000, 1111, 0011 y 1100*. Para que un código detector o corrector admita el mayor número posible de palabras, lo más indicado es que las distancias entre palabras alcancen el valor mínimo posible. Así, p. ej., en un código detector de n errores se tomarán las palabras a una distancia de $n + 1$, y en un código corrector de n errores las palabras adoptarán una distancia de $2n + 1$.

* En el original rumano estas líneas contenían una inexactitud, que pudo ser subsanada gracias a la observación del traductor (Nota del autor).

10.9. El método de las verificaciones de paridad

Para la detección y corrección de los errores se emplea muy a menudo el llamado método de las verificaciones de paridad. El caso más simple corresponde al descubrimiento de los errores simples. En todo sistema práctico de transmisión siempre existe una probabilidad finita de aparición para un error simple, probabilidad muy pequeña, por lo común. La probabilidad de que aparezca un error doble, triple, etc. es todavía menor. De ahí que, en determinadas condiciones técnicas, podamos tener una seguridad suficiente de que en cada una de las palabras recibidas no aparezca más que un error simple, a lo sumo. En estas condiciones, y con vistas al descubrimiento, a cada palabra transmitida se le asigna una cifra 0 ó 1, denominada *cifra de verificación*, de tal manera que el número total de cifras 1, en la palabra nueva, sea par. Por ejemplo:

101100, cifra de verificación: 1,

100010, cifra de verificación: 0.

En el primer caso, transmitimos la palabra 1011001. Cualquier error que sobrevenga, el número de cifras 1 en la palabra recibida dejará de ser par, con lo que el receptor inferirá de inmediato que la palabra contiene un error. Nótese que así sólo es posible descubrir errores, ya que el receptor ignorará, en todo caso, *cuál* de las cifras está equivocada y, por tanto, no podrá subsanar el error. Evidentemente, existen aún tres variantes, en el sentido de que al añadir la cifra de verificación puede procederse a fin de que el número total de cifras 1 sea impar o que el número total de cifras 0 sea par o impar. Esto depende de la convención establecida por el emisor y el receptor de mensajes. Una vez comprobado que la verificación de paridad queda satisfecha, se borra la cifra de verificación, puesto que no contiene ningún valor informacional, aparte de asegurar que la palabra respectiva ha sido recibida correctamente.

10.10. Verificaciones de paridad para la corrección de errores simples y el descubrimiento de los dobles

Veamos ahora qué métodos existen no sólo para detectar, sino también para corregir los errores simples. Adoptemos el caso más sencillo en que las palabras del código contengan una sola cifra, *A*. Utilizaremos, además, dos cifras de verificación, *B* y *C*, de tal modo que

$$A + B \text{ y } A + C$$

sean pares. El que recibe la palabra calcula las dos sumas: si sólo una de ellas es impar, infiere que la cifra de verificación es errónea. Si, p. ej., $A + B$ es impar

y $A + C$, par, entonces la cifra B está equivocada. Si ambas sumas resultan impares, entonces la cifra A está equivocada. Procediendo así se sabe cuál es la cifra errónea.

Consideremos ahora el supuesto de un código con palabras de cuatro cifras cada una, A, B, C, D , para las cuales tomamos tres cifras de verificación, E, F, G . Las cifras de información A, B, C, D admiten arbitrariamente los valores 0 y 1, pero las cifras de verificación están determinadas por el emisor de la palabra de tal modo que las sumas

$$A + B + C + E, \quad A + B + D + F, \quad A + C + D + G$$

sean pares.

El receptor de la palabra calcula inmediatamente las tres sumas. Existen las siguientes posibilidades:

- 1) si las tres son pares, no hay ningún error;
- 2) si sólo una de las sumas es impar, la cifra de verificación de la suma respectiva es errónea;
- 3) si sólo dos sumas son impares, el error se halla en la cifra común a las sumas impares y ausente en la par, es decir si las dos primeras son impares, el error está en B , y
- 4) si las tres sumas son impares, la cifra equivocada es un término común a las tres sumas; en nuestro caso, A .

Para ilustrar en concreto este punto, supongamos que se recibe la palabra 1101100. Las primeras cuatro cifras son de información y las tres últimas de verificación. Ocurre, entonces, que:

$$A + B + C + E = 1 + 1 + 0 + 1 = 3,$$

$$A + B + D + F = 1 + 1 + 1 + 0 = 3,$$

$$A + C + D + G = 1 + 0 + 1 + 0 = 2,$$

luego, la cifra equivocada es la que se halla en la posición 2 y la palabra se escribirá correctamente a base de 1001.

Puede procederse igualmente disponiendo de 11 cifras de información y cuatro cifras de verificación. El método es general y puede demostrarse que mediante k cifras de verificación es posible verificar palabras que contengan un máximo de $2^k - k - 1$ cifras de información. No insistiremos, sin embargo, sobre el método general.

Examinemos ahora el problema de descubrir los errores dobles. La hipótesis de base establece que no pueden aparecer más que errores simples o dobles.

Partimos del código corrector de errores simples construido en el apartado anterior. A este código le añadimos una cifra suplementaria a fin de que el nú-

mero total l de cifras sea par. En la recepción de una palabra pueden darse las siguientes alternativas:

- 1) todas las verificaciones de paridad, es decir las del código inicial y la suplementaria, quedan satisfechas y, por lo tanto, no hay error;
- 2) algunas de las verificaciones de paridad iniciales no quedan satisfechas, ni tampoco la final; entonces, es que hay un error simple en la parte inicial de la palabra y, en consecuencia, es susceptible de ser corregido;
- 3) la última verificación de paridad no queda satisfecha y si, en cambio, las iniciales; resulta, entonces, que la cifra suplementaria es errónea, y
- 4) la última verificación de paridad queda satisfecha, pero no algunas de las iniciales; se trata, en este caso, de un error doble.

Este código permite la corrección de los errores simples, pero sólo descubre, sin corregirlos, los errores dobles.

Se observará que, en contraste con la tendencia a la economía —que reclama la máxima reducción en el número de las cifras componentes de cada palabra—, aquí se añaden cifras. Las cifras de verificación son inútiles desde el punto de vista informacional, pero, en cambio, permiten aprovechar las propiedades de detección y corrección.

10.11. Una interpretación geométrica

A menudo se recurre a una elegante interpretación geométrica para los códigos autocorrectores. La idea consiste en identificar una palabra, constituida por una cadena de n cifras binarias, con el punto del espacio de n dimensiones que tiene estas cifras como coordenadas. De esta manera, las palabras del código forman un subconjunto del conjunto de los vértices del cubo unitario del espacio de n dimensiones (cf. fig. 60).

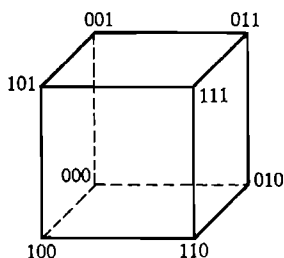


Figura 60

En el conjunto formado por los 2^n vértices del cubo se introduce una distancia del siguiente modo: la distancia entre los dos vértices es igual a la distancia que hay entre las palabras asociadas a cada vértice. Esta última se ha definido, en el apartado 10.6, como igual al número de posiciones en que las dos palabras se distinguen recíprocamente.

Se advierte fácilmente que la distancia entre dos vértices no es más que el número mínimo de aristas que deben recorrerse para pasar de un vértice al otro. Con la distancia así definida, el conjunto de los vértices del cubo forma un espacio métrico.

Cabe así ilustrar e interpretar geoméricamente los teoremas de los apartados 10.7 y 10.8 sobre las condiciones en que un código descubre o corrige un determinado número de errores.

Un código binario es un subconjunto del conjunto de todos los vértices del cubo-unidad del espacio de n dimensiones (siempre es posible disponer que todas las palabras del código tengan la misma longitud n). Un error simple consiste en el cambio de una sola coordenada, de modo que transforma un vértice cualquiera del cubo en otro vértice situado a la distancia 1 del inicial. Un error doble transforma un vértice en otro situado a la distancia 2, y así sucesivamente.

Supongamos ahora que todos los vértices que representan los símbolos del código se encuentran a una distancia mutua de 2, como mínimo. En estas condiciones, todo error simple transforma un punto cualquiera del espacio métrico asociado al código en un punto que no pertenece a este espacio métrico. Así, pues, el código permite el descubrimiento de los errores simples.

Admitamos ahora que todos los puntos del código se hallan a una distancia mutua de 3, por lo menos. En este caso, todo error simple transformará un punto de código en un punto que no pertenece al código, pero que se hallará más cerca del punto inicial que de cualquier otro punto del mismo código. Luego, el código permite tanto el descubrimiento como la corrección de los errores simples.

Procediendo de un modo semejante, se comprueba que si la distancia mínima entre los puntos del código es igual a 4, el código permite corregir los errores simples y detectar los errores dobles; si la distancia mínima es igual a 5, podrán corregirse ya los errores dobles, y así sucesivamente.

En este sentido, la tarea de encontrar los códigos detectores y correctores de errores es paralela a la de encontrar determinados subconjuntos de los conjuntos de vértices del cubo unitario n -dimensional, subconjuntos cuyos puntos se encuentren suficientemente apartados entre sí. Es típico el siguiente problema: hallar, en el conjunto de los vértices del cubo unitario de n dimensiones, cuál es el número máximo de puntos que pueda contener un subconjunto tal que la distancia entre dos de sus puntos cualesquiera sea por lo menos igual a d (siendo d un número natural). Una vez contestado este problema, hay que construir

efectivamente este subconjunto maximal. Se trata, por lo demás, de un problema que aún no ha sido resuelto por completo.

10.12. Código, grupo, algoritmo

Otra vía de acceso para el problema de los códigos detectores y correctores de errores se basa en la teoría de los grupos finitos.

El conjunto formado por los elementos 0 y 1 puede organizarse como un grupo abeliano \mathcal{C}_1 con la ley de composición $0 + 0 = 0$, $0 + 1 = 1 + 0 = 1$, $1 + 1 = 0$ (la llamada adición módulo 2)⁶. El conjunto B_n de las sucesiones de n cifras binarias cada una forma, evidentemente, un grupo abeliano isomorfo con el producto directo \mathcal{C}_1^n . Un código con n posiciones se denomina *código grupal* si forma un subgrupo de B_n . Se ha demostrado que existe identidad entre los códigos basados en verificaciones de paridad, por una parte, y los códigos grupales, por otra.

Del mismo modo, se utilizan los métodos de la teoría de la representación de los grupos a fin de obtener códigos optimales.

Existen aun otras ramas de la matemática aplicables a la teoría de los códigos. Parece que la teoría de los semigrupos es un marco muy natural para la teoría de los códigos. El cálculo de probabilidades y la estadística matemática ocupan un puesto central en la problemática sobre los códigos y en los métodos que abordan esta problemática. Se han puesto de manifiesto analogías interesantes entre la teoría de los códigos y ciertas teorías de la lógica matemática⁷. También aquí es forzoso recordar los vinculos estrechos que hay entre la teoría de los códigos y la teoría de los algoritmos. Cosa natural, por lo demás, ya que la codificación se reduce a una transformación de un conjunto de palabras en otro conjunto de palabras, de modo que esta transformación será algoritmizable en tanto que pueda representarse como una aplicación del tipo que interviene en la definición de algoritmo normal (en el sentido de Markov).

10.13. Código y entropía

Sea un alfabeto A formado por m elementos. Supongamos que hemos codificado un determinado texto con ayuda de un código sobre el alfabeto A . Supongamos asimismo que conocemos la estructura estadística del texto y, en particular, sus entropías de distintos órdenes $H_0, H_1, \dots, H_n, \dots$ incluso la entropía de orden ∞ , que viene dada por la expresión $H = \lim_{n \rightarrow \infty} H_n$ (cf. 8.9 y ss.). Como se sabe, suce-

6. [Cf. 3.18, nota 36].

7. Cf. APOSTEL-MANDELBROT-MORF. *Logique*.

de que $H_n \leq H_0$ para todo n , de manera que $H \leq H_0$. Admitamos, para simplificar, que el alfabeto en que se ha escrito inicialmente el texto considerado es idéntico al alfabeto A del código. Por ello, la entropía de orden 0 del texto inicial es la misma que la entropía de orden 0 del texto codificado y, por tanto, igual a $\log_2 m$.

Ocurrirá, entonces, que:

$$H \leq \log_2 m.$$

Es natural codificar del modo más breve, es decir a base de palabras de longitud mínima cuando se encuentren con la mayor frecuencia posible en el texto considerado, y de longitud mayor para las que aparezcan más raramente. De esta forma se consigue una sensible reducción del tiempo de transmisión del mensaje codificado. Toda palabra C del texto inicial tiene, en este texto, una determinada probabilidad de aparición $p(C)$. Notemos por medio de $l(C)$ la longitud de la palabra en que se transforma C al codificarse. Notando por medio de s la longitud de C , la relación

$$\frac{l(C)}{s}$$

puede considerarse como el coeficiente de contracción de la palabra C . La esperanza matemática de esta relación, definida a base de

$$\mu_s = \frac{\sum p(C)l(C)}{s}$$

dónde la adición se hace según todas las palabras C de longitud s , representa la **contracción media** para las palabras de longitud s . La magnitud

$$\mu = \lim_{s \rightarrow \infty} \sup \mu_s$$

se denomina el coeficiente de contracción del texto considerado, mediante la codificación adoptada. Esta magnitud mide la reducción del texto debida a la codificación. Puede demostrarse que, en los casos más normales, la sucesión $\{\mu_s\}$ es convergente, de forma que el límite superior que define a μ es un límite propiamente dicho.

Se plantea, de un modo natural, el siguiente problema: ¿cuál es el valor mínimo del coeficiente de contracción que puede obtenerse mediante una codificación lo más adecuada posible, suponiendo conocida y bien determinada la estructura estadística del texto? Una vez determinado este valor, hay que construir el código que lo realice. En teoría de la información existe el siguiente teorema, que brinda una respuesta casi completa a los dos problemas formulados,

en el caso de que el texto considerado sea asimilado a una cadena de Markov simple, de un determinado tipo⁸:

Si la entropía del texto inicial es igual a H , entonces el margen inferior del coeficiente de contracción μ , tomada según todos los códigos posibles, equivale a

$$\frac{H}{\log_2 m} = \left(\frac{H}{H_0} \right)$$

(recordemos que m es el número de los elementos del alfabeto A).

Como se ve, para establecer el margen inferior de las reducciones posibles de un texto mediante la codificación basta conocer su entropía y el número de elementos que componen el alfabeto. La relación aducida constituye lo que se llama, en teoría de la información, la entropía relativa del texto considerado. Del hecho por el cual $H \leq \log_2 m$, resulta que la entropía relativa se halla comprendida entre 0 y 1.

10.14. Algunas ilustraciones y analogías lingüísticas

Una interesante aplicación lingüística de los códigos binarios se debe a Dénes Varga en lo que atañe a la descripción de las distintas posibilidades de combinación sintagmática de los casos en el marco del sintagma nominal ruso⁹. (Tal como aprecia el mismo Varga, el método tiene validez general.) Se considera, a este propósito, un código binario donde cada palabra presenta una longitud igual a 12. Las palabras del código se obtienen del siguiente modo: a cada palabra del ruso se le asocia una sucesión de 12 cifras binarias. Las primeras seis cifras binarias corresponden a los casos de singular y las seis restantes a los de plural. Los casos se ordenan así: nominativo (N), acusativo (A), genitivo (G), dativo (D), instrumental (I) y preposicional (P). En cada posición se pone la cifra 0 ó 1 en función de que el caso respectivo se refiera, o no, a la palabra rusa considerada. Así, p. ej., a la preposición *s*, 'con' (en relación con su regente) se le asocia la siguiente palabra del código:

$$\varphi_1 = 0 \ 1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 1 \ 0.$$

La cifra 1 de la tercera posición indica que la preposición *s* puede construirse con el caso genitivo singular; a su vez, la cifra 0 de la séptima posición indica que la preposición no puede construirse con el nominativo plural. Con arreglo al mismo principio, a la palabra *náshei*, 'de, a, con, ...nuestra', se le asocia la siguiente palabra del código:

$$\varphi_2 = 0 \ 0 \ 1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0.$$

8. Cf., p. ej., JINCHIN, "Poniatie entropii".

9. Cf. VARGA, "Morphological analysis".

En efecto, el pronombre *náshei* es compatible con el genitivo singular, dativo singular, instrumental singular y preposicional singular, pero no con otros casos. En fin, al nombre *fábriki*, 'de la fábrica', se le asocia la siguiente palabra del código:

$$\varphi_3 = 001000110000.$$

Definimos ahora la operación \wedge (multiplicación lógica), aplicado a las palabras del código, de la siguiente manera: la palabra $\varphi \wedge \psi$ tiene, en una posición dada, la cifra 1 si, y sólo si, tanto φ como ψ tienen, en la misma posición, la cifra 1. El significado de esta operación es inmediato: la palabra $\varphi \wedge \psi$ tiene la cifra 1 exactamente en aquellas posiciones en que φ y ψ son compatibles en el ámbito de un sintagma nominal. Entonces, si, p. ej.:

$$\varphi_1 \wedge \varphi_2 = 001010000000,$$

es que *s* y *náshei* sólo son compatibles en el genitivo singular y en el instrumental singular. Y si:

$$\varphi_1 \wedge \varphi_2 \wedge \varphi_3 = 001000000000,$$

es que en el sintagma nominal *s náshei fábriki*, 'con nuestra fábrica', el pronombre y el nombre se hallan en genitivo singular.

La noción de distancia definida en la teoría de los códigos (cf. 10.6) es ampliamente aplicable desde el punto de vista lingüístico. Así, volviendo al código binario del sistema consonántico inglés del apartado 10.4, cabe observar que, en general, las oposiciones más susceptibles de neutralizarse corresponden a las que se ejercen sobre fonemas situados a una distancia recíproca de 1 (en cuanto al código en cuestión). En condiciones de distancia 1 se hallan, p. ej., *p* y *b*, *f* y *v*, *t* y *d*. Si, entonces, recordamos que la existencia, en un código, de palabras que se hallen a una distancia de 1 impide la posibilidad de descubrir los errores simples, resulta que la introducción de los archifonemas sirve para facilitar el descubrimiento de los errores simples, ya que identifica, en determinadas posiciones, los fonemas que pueden confundirse. Cuanto mayor es la distancia entre fonemas, tanto más reducida es la posibilidad de que se neutralice su oposición.

Refiriéndonos igualmente a un fenómeno de sincretismo, en este caso gramatical, examinemos el problema de la homonimia morfológica de los adjetivos calificativos del rumano. Desde el punto de vista morfológico, estos adjetivos pueden describirse mediante un código binario formado por palabras de longitud 4. A cada forma adjetiva le asociamos una o más sucesiones de cuatro cifras binarias cada una del siguiente modo: en primera posición pondremos 1 ó 0 según que la forma respectiva esté en singular o en plural; en la segunda posición escribiremos un 1 si la forma se halla en caso directo [nominativo o acusativo] o un 0 si se halla en caso oblicuo [genitivo o dativo]; en la tercera posición pondremos un 1, para el masculino, o un 0, para el femenino, y en la

cuarta posición pondremos un 1 si va en forma determinada [con artículo determinado] o un 0 si va en forma no determinada [sin artículo determinado]. A una forma como *frumosul*, 'lo hermoso', se le asocia una sola palabra del código: 1111. A una forma como *frumos*, 'hermoso', se le asocian dos palabras: 1110 y 1010. A una forma como *mici*, 'pequeños, -as', se le asocian cuatro palabras: 0110, 0010, 0100 y 0000. A una forma como *cumsecade* [adjetivo formado a base de *cum se cade*, 'como se debe', con valor de 'honesto', 'adecuado'], se le asocian ocho palabras: 0001, 0011, 0101, 1001, 1101, 0111, 1011 y 1111. Observemos que, en consecuencia, entre las palabras asociadas a *frumos* la distancia es de 1; entre las asociadas a *mici*, de 2, y entre las asociadas a *cumsecade*, de 3¹⁰. En este sentido, la noción de distancia proporciona un criterio de clasificación para los distintos tipos de homonimia morfológica, la cual resultará más pronunciada cuanto mayor es la distancia máxima correspondiente.

Algunos fenómenos de la fonología o de la gramática admiten una explicación a la luz de la tendencia que la lengua manifiesta hacia el mejor cumplimiento posible de la función de descubrir y corregir los errores. El fenómeno de la concordancia es precisamente una ilustración de esta tendencia, ya que desempeña la función de la cifra de control que permite la detección de un error. Si hemos de comunicar diversas formas nominales, cabe convenir la adición, como "cifra de control", de una forma adjetiva, de modo que la verificación de paridad consistirá, en este caso, en el hecho de que la forma adjetiva debe concordar con la forma nominal en género, número y caso. Al captar un mensaje a base de **pom frumosi*, '*árbol hermosos', nos damos cuenta de que contiene un error, que, por otra parte, no podemos corregir, debido a que no sabemos a cuál de los dos términos afecta (el mensaje transmitido podía ser tanto *pom frumos*, 'árbol hermoso', como *pomi frumosi*, 'árboles hermosos'). Para que la lengua llegue a ser no sólo un código detector, sino también corrector de errores, el fenómeno de la concordancia es insuficiente, pues hay que recurrir a un análisis contextual más complejo.

Debemos observar, sin embargo, que contra lo que ocurre en los códigos matemáticamente contruidos —donde cabe, en principio, descubrir y corregir cualquier tipo de error—, la lengua cumple el mismo cometido de una manera muy compleja, no desprovista de molestas vacilaciones. La "homonimia inso-

10. [Recapitulemos: como se ha dicho (cf. 7.2, nota 5), los adjetivos rumanos presentan diversos paradigmas, que esta notación binaria cubre en todas sus posibilidades. Aquí no se prevé, p. ej., el género neutro, puesto que la concordancia con nombres neutros se hace en masculino para el singular y en femenino para el plural. En los ejemplos aducidos, *frumos* tiene un paradigma regular de cuatro terminaciones (masc. + sing., plur.; fem. + sing., plur.) y, como es ambiguo respecto del factor casual, requiere dos palabras del código; *mici*, 'pequeño', tiene tres terminaciones (*mic*, masc. sing.; *mică*, fem. sing.; *mici*, masc. y fem. plur.) y en su forma *mici* es ambiguo respecto a los factores género y caso, cuyas dobles alternativas producen las cuatro palabras del código (2²); a su vez, *cumsecade* tiene una sola desinencia y, como sólo está marcado en cuanto al factor determinación, el resto produce 8 variantes posibles (2³).

portable” de la que hablan algunos lingüistas constituye precisamente un callejón sin salida de este tipo. Aunque, por lo común, la concordancia descubre, como hemos visto, los errores simples, a veces también falla. Así sucede en un sintagma de tipo *învățătoare tenace*, ‘maestra(s) tenaz(ces)’, donde es imposible precisar si se trata de formas en singular o en plural. Todo el problema de la resolución sintáctica de la homonimia puede abordarse como un caso particular de otro problema más general, referido al modo en que la lengua desempeña la función de código detector y corrector de errores.

11. El reconocimiento automático de las formas

11.1. Generalidades

El reconocimiento de las formas es un problema que interesa a todos cuantos se ocupan de las cuestiones generales de la cibernética.

Desde el punto de vista cibernético, se consideran en primer lugar los objetos susceptibles de ser reconocidos a través de un número finito de notas. Si cada nota puede reconocerse en un intervalo finito de tiempo, entonces es que también el objeto de este tipo puede reconocerse en un tiempo finito. En caso de que el número de notas sea infinito, pero finita la duración del reconocimiento de cada una de ellas, ya no podrá, entonces, reconocerse el objeto en un intervalo finito de tiempo.

Conviene observar que existe la posibilidad de expresar incluso cualidades mediante números. Por lo que se refiere a los colores, p. ej., la situación es muy fácil de explicar. Toda señal luminosa monocromática queda caracterizada por dos indicios: la intensidad y la longitud de onda, magnitudes cuantitativamente expresables. Pero también en el caso de las señales luminosas complejas cabe aplicar esta propiedad, puesto que todo color admite una descripción como combinación adecuada de señales correspondientes a los tres colores fundamentales y, por tanto, a base de seis números.

En lo que atañe a las formas espaciales, la geometría analítica ofrece justamente la posibilidad de clasificarlas a partir de determinados indicios cuantitativos. Así, p. ej., las secciones cónicas pueden clasificarse en hipérbolas, parábolas, elipses, círculos o rectas. Si sólo se consideran las secciones cónicas, basta conceder un número a cada una de estas formas características. En una segunda etapa, sólo hay que añadir, a lo sumo, dos nuevos parámetros (los ejes) para determinar exactamente la forma que debe clasificarse. Evidentemente, este método puede utilizarse para cualquier colección finita de tipos a partir de los cuales se clasifiquen determinados objetos.

Existen, no obstante, otras posibilidades. Por ejemplo, las curvas algebraicas pueden clasificarse en virtud de los $n + 1$ coeficientes que intervienen en el polinomio de grado n que representa la curva. Si una magnitud geométrica en un espacio con n dimensiones viene dada paramétricamente por n polinomios de grado m , a lo sumo, entonces es que la magnitud viene dada por nm parámetros numéricos.

Antes de pasar al problema particular del reconocimiento de las formas lingüísticas, nos parece indicado hacer resumidamente una presentación general de la cuestión.

Reconocer un objeto significa establecer sus indicios característicos. Dos objetos que no presenten ningún indicio diferente son, desde este punto de vista, el mismo objeto, en tanto que no pueden distinguirse. Pero si estos dos objetos ocupan dos posiciones espaciales diferentes, también esta particularidad puede constituir un indicio de clasificación y, por tanto, de reconocimiento. El problema del reconocimiento debe ponerse, pues, en relación con las posibilidades efectivas de clasificar y con las posibilidades del analizador que se emplee con este propósito.

Sea I_i uno de los indicios de reconocimiento. Un objeto dado está caracterizado por los valores v_i que presentan sus indicios característicos I_i , donde $i = 1, \dots, N$. El caso más simple es aquél en que cada uno de los indicios de reconocimiento ofrece sólo dos valores posibles, digamos 0 y 1. Puede considerarse que un objeto es identificado, en estas condiciones, por una fórmula lógica en la que se indican los valores v_i que tienen todos sus indicios de reconocimiento. Si se escribe

$$\prod_{i=1}^N x_i^{v_i}, \quad (1)$$

y se adopta la convención de que

$$x^a = \begin{cases} 1 & x = a \\ 0 & x \neq a \end{cases}$$

de modo que x y a tengan sólo el valor 0 ó 1, entonces toda combinación de indicios, para la cual la expresión (1) tenga el valor 1, corresponde a un objeto. Para un objeto dado se toma $a_j = 1$, si el objeto presenta la cualidad correspondiente a x_j .

En caso en que se tenga en cuenta más de cerca la realidad, cabe introducir también criterios probabilísticos. En estas condiciones, los objetos vienen caracterizados por notas que pueden adoptar no sólo dos valores posibles —digamos 0 y 1—, sino una variación eventualmente continua. Puede considerarse que las n notas características, los n indicios de reconocimiento, son los n ejes de coordenadas de un espacio abstracto E , n -dimensional. En este espacio E , cada ob-

jeto determina un punto dado. Se denomina *categoría de objetos* al conjunto de todos los puntos que se agrupan en un sector definido de este espacio.

Sean dos subespacios I_j y I_k , comprendidos ambos en E , que determinen cada uno un objeto dado. Para poder reconocer de un modo unívoco los objetos, es necesario que los conjuntos I_j y I_k sean disjuntos:

$$I_j \cap I_k = \emptyset \quad (2)$$

para todo $j \neq k$ y donde \emptyset es el conjunto vacío.

La relación (2) es necesaria, ya que, en caso de que los dos conjuntos no fuesen disjuntos, el reconocimiento de los objetos no sería ya unívoco. En efecto, si

$$I_j \cap I_k \neq \emptyset,$$

entonces, para un elemento

$$x \in I_j \cap I_k$$

no se puede afirmar si pertenece al subespacio I_j o al subespacio I_k , por lo que el reconocimiento deja de ser unívoco.

Esta teoría general puede aplicarse también a las señales lingüísticas. En lo que sigue presentaremos el problema derivado del reconocimiento no sólo de las formas sonoras, sino también de las formas gráficas, en el ámbito lingüístico.

11.2. Reconocimiento de las vocales. Espectros

En lo que atañe al reconocimiento automático de las señales fónicas del habla, el caso más simple se refiere a las vocales. Para comprender los métodos utilizados en los aparatos de reconocimiento automático de señales vocálicas, conviene que introduzcamos de antemano algunas cuestiones relativas a los llamados espectros de señales.

Consideremos una señal sinusoidal

$$a(t) = A \operatorname{sen}(2\pi ft + \varphi),$$

donde A es la amplitud de la señal, f la frecuencia, t el tiempo y φ la fase. La magnitud $\omega = 2\pi f$ se llama *pulsación*. Esta señal es periódica en el sentido de que tras un tiempo $T = 1/f$, llamado *período*, la señal vuelve al mismo valor.

Se denomina *función periódica* a una función $f(t)$ que tiene la propiedad de que

$$f(t + T) = f(t)$$

para todo valor del argumento t . El menor número $T \neq 0$ que verifica esta igualdad se denomina período.

Hay un teorema, debido a Fourier, que demuestra que toda función periódica que verifique ciertas condiciones conocidas por el nombre de **condiciones de Dirichlet**, puede considerarse como resultado de la superposición de un número suficientemente grande de funciones sinusoidales. Las condiciones de Dirichlet son las siguientes:

- 1) $f(t)$ es finita;
- 2) a cada t le corresponde un solo valor de $f(t)$;
- 3) el número total de máximos, mínimos y discontinuidades de primera especie, en el curso de un período, es finito.

Hay que mencionar que, si $f(t)$ es discontinuo en un punto cualquiera t_0 , la suma de la serie en el punto en cuestión es igual a la media aritmética de los límites a la derecha y a la izquierda del punto t , es decir

$$\frac{1}{2} \left[\lim_{t \rightarrow t_0^+} f(t) + \lim_{t \rightarrow t_0^-} f(t) \right] = f(t)$$

Obsérvese que estas condiciones quedan verificadas por todas las señales encontradas en la realidad.

Para una tal función puede escribirse:

$$f(t) = \sum_0^{\infty} A_n \text{sen} (n \omega t + \varphi_n),$$

donde el componente de frecuencia más baja ($f = \omega/2\pi$) se denomina **fundamental**, frente a los demás componentes de frecuencia nf , denominados **armónicos**.

Por lo común, la función se expresa tan sólo a base de las amplitudes de sus distintos componentes, y se ignoran, así, sus fases respectivas debido a que el oído humano no registra las fases de las señales que componen un sonido complejo. La totalidad de las amplitudes de las distintas frecuencias de una señal periódica constituye el **espectro** de esta señal. En la figura 61 se ofrece una señal periódica y su constitución en el fundamental y los armónicos.

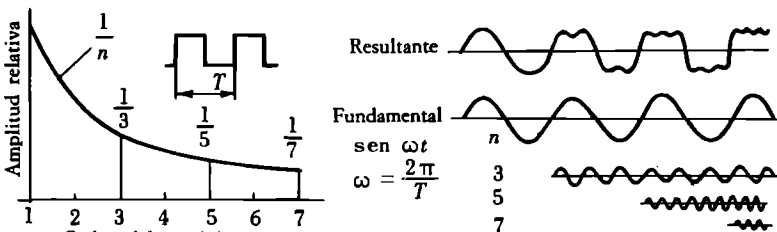


Figura 61

Fácilmente se observa que el espectro de una señal periódica está formado por elementos discretos y que la frecuencia de cada componente es un múltiplo de la frecuencia fundamental.

En realidad, no existen señales periódicas, por cuanto una señal periódica tiene una duración desde $-\infty$ hasta $+\infty$. De hecho, sin embargo, toda señal tiene una duración finita.

Incluso para las señales no-periódicas hay otro teorema de descomposición. Indica que una señal no-periódica puede reconstituirse asimismo a partir de señales sinusoidales, pero con la diferencia de que en este caso el espectro es continuo, en el sentido de que para todo componente existe otro componente con una frecuencia no importa cuán cercana a la anterior. Ahora bien, si la señal no-periódica está constituida por una sucesión finita de señales, que se repitan periódicamente n veces en el intervalo considerado, entonces se demuestra que a medida que crece n , el espectro continuo de la señal respectiva tiende a transformarse en un espectro discreto. Esta situación se produce en el caso de las señales vocálicas. En la figura 62a, hemos representado, tal como aparece en el oscilógrafo catódico, el registro de la variación de presión del aire para una vocal durante una experiencia realizada a baja velocidad. A una mayor velocidad, se obtiene la imagen de la figura 62b, donde se aprecian los elementos que, al repetirse un número de veces suficientemente grande, dan la señal en su totalidad.

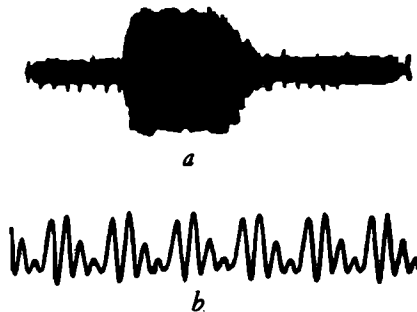


Figura 62

Si la señal $s(t)$ de la fig. 62a viene expresada por la ecuación $s(t) = a(t)$, entonces la señal de la fig. 62b queda expresada por la ecuación:

$$v(t) = m(t) a(t),$$

donde $m(t)$ representa la llamada *en v o l t u r a* de la señal, es decir indica el modo en que varía en tiempo la *a m p l i t u d* de la señal elemental $a(t)$ (figura 63). Como para una vocal $a(t)$ se repite un número de veces suficientemente

grande, resulta que las señales vocálicas pueden considerarse dotadas, con aproximación, de un espectro discreto.

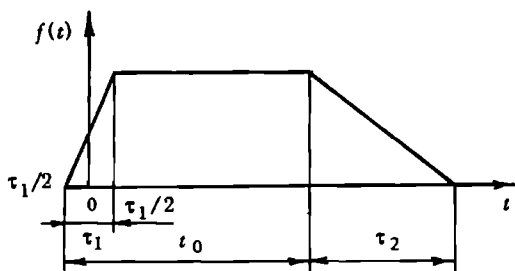


Figura 63

Hay que consignar que, en general, cuando se conoce la ecuación de una función, digamos $f(t)$, puede determinarse su espectro $F(\omega)$ a partir de la transformación de Fourier:

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) \exp(j\omega t) dt.$$

La función $F(\omega)$ describe la variación en frecuencia de las amplitudes que tienen los componentes de la función $f(t)$ y, por tanto, es el espectro de $f(t)$.

Por espectro corriente de la señal $f(t)$ se alude a la función:

$$F_c(\omega, t_0) = \int_{-\infty}^{t_0} f(t) \exp(j\omega t) dt,$$

donde t_0 es un momento dado, eventualmente el tiempo corriente. Para $t_0 \rightarrow \infty$, $F_c(\omega, t_0) = F(\omega)$. Si t_0 es suficientemente grande, $F_c(\omega, t_0)$ da una abundante información sobre $F(\omega)$ y puede considerarse que, en este caso, el espectro corriente coincide con el espectro propiamente dicho.

En la actualidad existen aparatos que efectúan automáticamente el análisis de las señales sonoras. Estos aparatos permiten obtener imágenes como la que vemos en la figura 64, donde en un eje se representa la frecuencia y en el otro la amplitud de las señales del espectro para un momento dado, elegido sobre el transcurso del registro.

El examen de estas imágenes correspondientes a vocales lleva a la conclusión de que la energía se concentra especialmente en determinadas porciones del espectro, denominadas *formantes*.

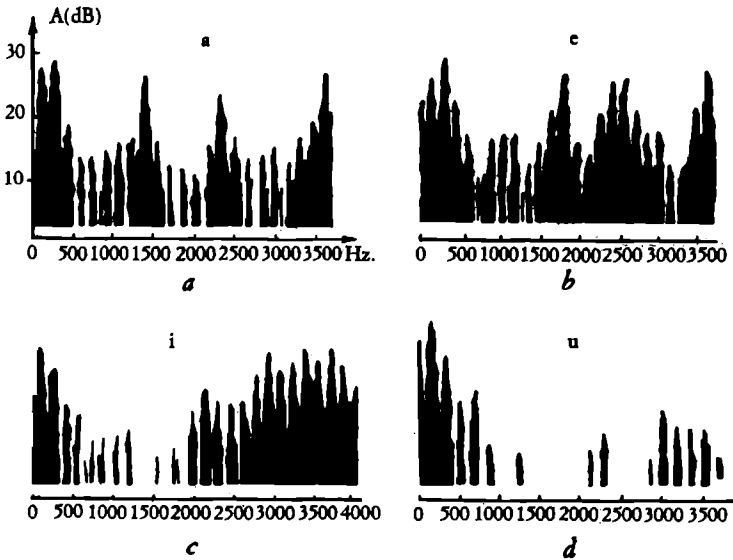


Figura 64

11.3. Aparatos que reconocen las señales vocálicas

Los aparatos capaces de reconocer vocales se construyen sobre la base de los principios generales expuestos, en el apartado 11.1, sobre la teoría del reconocimiento de las formas. Uno de los principios empleados consiste en el reconocimiento de los sonidos a partir de los formantes (en la elección, como indicios de reconocimiento, de los formantes característicos).

La señal acústica se analiza gracias a un número n de filtros. Recordemos que los filtros son dispositivos que seleccionan señales u objetos con arreglo a determinados criterios. En el caso presente, se trata de filtros eléctricos, cada uno de los cuales caracterizado por una frecuencia propia dada f_i . A la salida del filtro se obtiene una señal proporcional al componente de frecuencia f_i del espectro de la señal aplicado a la entrada. En realidad, no es que los filtros permitan el paso de una sola frecuencia, sino de todas las frecuencias que se hallan en una determinada banda de frecuencias en torno a f_i .

Cuando ante un micrófono conectado a la entrada de los filtros se pronuncia una vocal cualquiera, la totalidad del espectro de la vocal respectiva es analizado y a la salida de los filtros se obtiene una señal dada. Se ha comprobado que

las señales resultantes pueden dividirse en dos grandes categorías. En algunas de ellas, la señal de salida presenta valores grandes: ello significa que el filtro respectivo está situado en la banda de uno de los formantes característicos de la señal vocálica analizada. En otros filtros, por el contrario, la señal es muy pequeña. Ello significa que el filtro respectivo se halla situado con su frecuencia característica en una región donde la señal analizada no tiene ningún formante.

Evidentemente, para que el análisis sea concluyente es necesario que la banda que debe analizarse esté dividida en un número n de franjas suficientemente grande, de tal modo que, a su vez, cada franja sea suficientemente estrecha. En estas condiciones puede cuantificarse la distribución de las señales a la salida del filtro, atribuyendo el valor 1 a las señales de valor grande, y 0 a las señales de valor pequeño. De esta manera, cabe representar las vocales a base de n cifras, cada una de ellas con valor 0 ó 1. Si se ordenan los números de acuerdo con un criterio preestablecido, p. ej., por el orden de los filtros en las bandas de frecuencia analizada, entonces el número resultante formado por n cifras nos facilita una indicación sobre los formantes que componen la señal analizada.

Por lo que se refiere al rumano, dividiendo la gama de frecuencias que van de 0 a 4 500 hz en nueve bandas de 500 hz, se obtiene la distribución de formantes del cuadro 33, donde se indica mediante 1 la presencia de formante y mediante 0 su ausencia.

	500	1 000	1 500	2 000	2 500	3 000	3 500	4 000	4 500
<i>a</i>	0	1	1	0	0	1	0	1	0
<i>o</i>	1	1	0	0	0	1	0	1	0
<i>u</i>	1	1	0	0	0	0	0	0	0
<i>e</i>	1	0	0	0	1	1	1	0	1
<i>ă</i>	1	0	1	0	0	1	0	1	1
<i>î</i>	1	0	0	1	1	1	0	1	0
<i>i</i>	1	0	0	0	1	0	1	1	0

Cuadro 33

Recurriendo al método propuesto al principio del presente capítulo, la vocal *ă* puede expresarse a base de la fórmula lógica:

$$x_1^1 x_2^0 x_3^1 x_4^0 x_5^0 x_6^1 x_7^0 x_8^1 x_9^1$$

donde x_i es la variable booleana asociada a la señal de la salida del filtro F_i .

Los estudios de análisis Fourier, practicados sobre las vocales rumanas han demostrado que cada una de estas vocales se caracteriza bien aun tomando en consideración sólo tres formantes. Se han examinado también, desde el punto de vista estadístico, las vocales *i e a o u* y *ă î*, y se ha llegado a la conclusión de

que, para estas vocales, los formantes principales F_1 , F_2 y F_3 presentan los valores medios indicados (en hz) en el cuadro 34¹.

Vocales	F_1	F_2	F_3
<i>i</i>	317	2 225	2 940
<i>u</i>	339	850	2 507
<i>î</i>	350	1 600	2 550
<i>e</i>	398	1 934	2 754
<i>o</i>	413	984	2 957
<i>ă</i>	500	1 500	2 750
<i>a</i>	703	1 278	2 622

Cuadro 34

Las vocales se han dispuesto según la frecuencia media del primer formante. A partir de este análisis espectrográfico es posible proyectar un aparato que reconozca las vocales utilizando el esquema indicado en la figura 65.

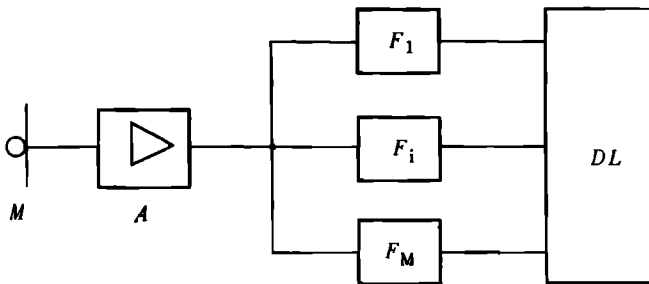


Figura 65

La señal se emite frente a un micrófono M , seguido de un amplificador A ; desde aquí la señal llega a los filtros F_i , seguidos del dispositivo lógico DL . Este último cuantifica en primer lugar las señales de la salida del filtro, atribuyendo un valor de 1 a las demás. En función de los valores obtenidos, cierra el contacto en un determinado circuito electrónico, que señaliza la vocal pronunciada frente al micrófono. En cuanto a la parte lógica, debe tener en cuenta los datos expuestos en el cuadro 33 y expresados gráficamente en la figura 66.

Por lo que se refiere al reconocimiento de las consonantes, el problema se hace mucho más complicado. Debemos recordar el hecho de que en la bibliografía especializada se indican también otros métodos para el reconocimiento automático de los fonemas, que nada tienen que ver con los métodos basados en el análisis de Fourier, de los cuales sólo se retienen los valores de la amplitud.

1. Cf. ŞUTEU, "Observații asupra structurii acustice", y AVRAM, "Vocale".

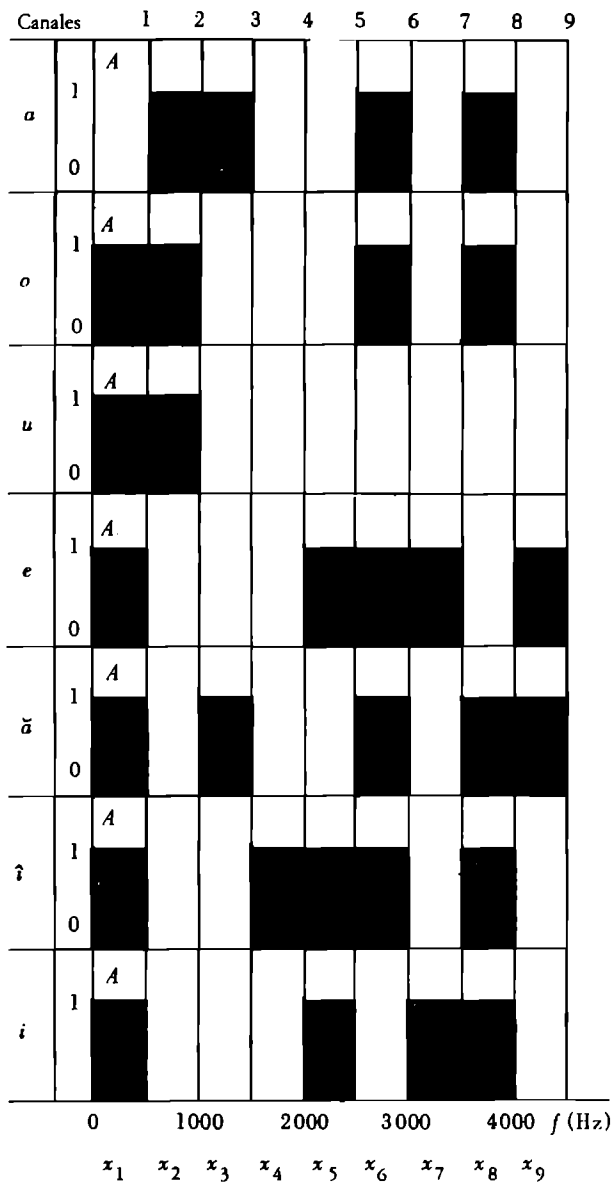


Figura 66

Recientemente, los trabajos en esta dirección han progresado mucho en relación con el problema de comprimir el espectro utilizado en el habla con vistas a la transmisión a distancia, en un solo circuito telefónico, del mayor número posible de conversaciones. En este sentido, se han realizado importantes estudios que han permitido crear aparatos que transforman el habla reduciendo mucho el espectro empleado. Uno de estos aparatos es, p. ej., el llamado Vocoder. En lo que sigue hablaremos en torno a esta cuestión.

11.4. La síntesis de las señales vocálicas

Con respecto a la síntesis de las señales vocálicas, el procedimiento habitual resulta de alguna manera inverso al de análisis.

Para toda señal a sintetizar se determina, en primer lugar, su espectro en cuanto a los formantes característicos. Luego basta que en un difusor se introduzcan corrientes con frecuencias iguales a las de los formantes y amplitudes proporcionales a las amplitudes de los formantes, a fin de que el altavoz reproduzca una señal acústica con las características sonoras de la vocal deseada. Las experiencias efectuadas hasta el presente demuestran que si se utilizan tres formantes se obtiene ya una aproximación acústica perfectamente válida en las señales sonoras.

Para comprender el principio de funcionamiento de una clase de sintetizadores, es muy indicado seguir el mismo proceso fisiológico que emplea el hombre para producir sonidos articulados. Las cuerdas vocales emiten señales que contienen un espectro muy rico. Estas señales presentan una forma no regular, pero periódica. A lo largo de su trayecto el aparato fonador, que hace las veces de filtro dotado para una determinada frecuencia, modifica la forma del espectro resultante. De un modo análogo se actúa en la síntesis de las señales vocálicas.

En las figuras 67 y 68 se representan dos variantes posibles para un sintetizador de señales vocálicas. En uno de los montajes se parte de un generador de ondas rectangulares, que son señales que se caracterizan por un espectro especialmente rico. Esta señal compleja llega a tres filtros dispuestos paralelamente (fig. 67). Cada filtro extrae del espectro de la onda rectangular un determinado

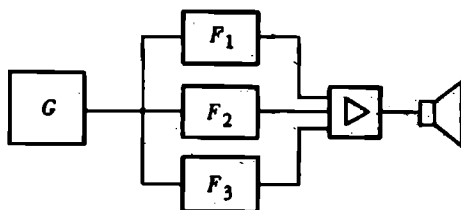


Figura 67

componente, para que la suma de los tres componentes se aplique a continuación al altavoz.

En el esquema de la figura 68, los filtros están dispuestos en serie. En principio, el funcionamiento es idéntico al del ejemplo anterior.

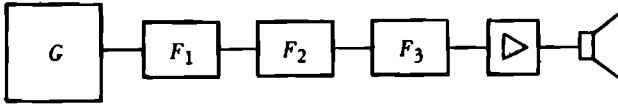


Figura 68

En los laboratorios de la Cátedra de Radiocomunicaciones del Instituto Politécnico de Bucarest se ha construido un aparato que permite no sólo el reconocimiento automático de las vocales rumanas, sino también su síntesis, empleando los esquemas de las figuras 65 y 66. Las experiencias realizadas han allegado resultados satisfactorios, pese a que sólo se han utilizado tres filtros regulables, es decir recurriendo a tres únicos formantes para la síntesis.

Como en el habla corriente existen determinadas proporciones de amplitud entre los componentes de los formantes, se han elaborado sistemas especiales que realizan de un modo automático las señales complejas, donde los formantes guardan la relación deseada.

Para especificar el espectro instantáneo con una precisión suficiente, que satisfaga las necesidades del oído, basta con controlar sólo seis variables: S_1 , la intensidad de la excitación fricativa; S_2 , la amplitud de la excitación laríngea; S_3 , la frecuencia fundamental de la excitación laríngea, y f_1, f_2, f_3 , las tres frecuencias de los formantes.

La experiencia demuestra que la disposición relativa de la amplitud de los formantes no es esencial en el proceso de síntesis para las señales vocálicas. La misma experiencia demuestra asimismo que es suficiente adoptar un sistema todavía más simple, en el que se controlen tan sólo tres magnitudes: S_1 , la intensidad de la excitación fricativa, y f_1, f_2 , las dos frecuencias de los dos formantes más graves; se ignoran, por tanto, los formantes de frecuencia más alta y la excitación laríngea. La frecuencia de los formantes está comprendida entre los 200 y los 1 200 hz para el primero y entre los 1 000 y los 2 400 hz para el segundo (fig. 69a).

El esquema del aparato de habla sintética realizado a partir de estos principios queda indicado en la figura 69b. Allí, la excitación fricativa viene producida por un generador eléctrico de ruido. Los formantes f_1 y f_2 se realizan con ayuda de unos circuitos sintonizados. Las tensiones U_1 y U_2 se regulan mediante potenciómetros.

Para producir sonidos semejantes a las señales vocálicas es necesario que simultáneamente se varíen las dos tensiones de control dentro de una determinada relación, para lo cual se recurre a un dispositivo mecánico adecuado. En

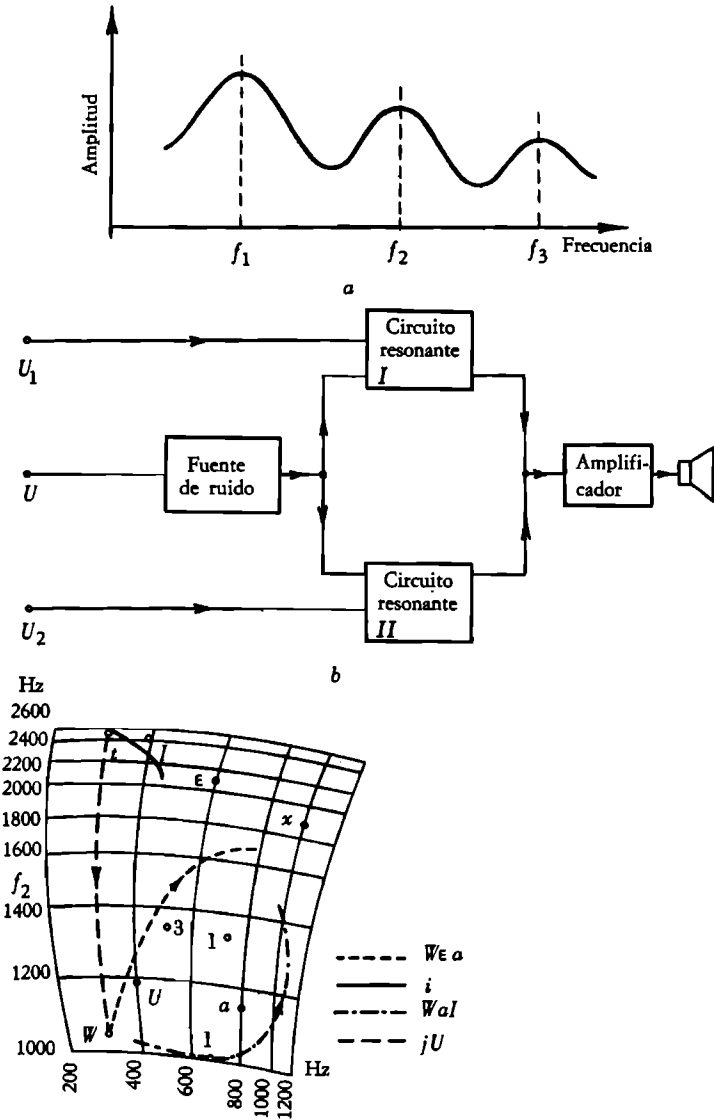


Figura 69

la figura 69c, se indica el movimiento que debe seguir el dispositivo de mando de este aparato con objeto de que produzca ciertas señales vocálicas. A base de las líneas trazadas en el diagrama se consigna el tipo de desplazamiento del dispositivo a fin de producir ciertas palabras formadas sólo por vocales y generadas con este aparato. En general, se trata de palabras inglesas.

Las pruebas de inteligibilidad efectuadas con esta máquina demuestran que, por lo común, se obtiene un 70% de identificaciones correctas con sonidos aislados y casi el 100% de identificaciones correctas con palabras y frases cortas, compuestas casi por vocales solamente.

Digamos, sin embargo, que la síntesis de las señales vocálicas está, en la actualidad, prácticamente resuelta con la ayuda de instalaciones complejas que comprenden generadores sinusoidales y un calculador digital.

Las experiencias efectuadas en los Laboratorios Bell han probado la posibilidad de obtener, a base de fichas perforadas que contienen indicaciones relativas a las señales vocálicas que deben sintetizarse, unas emisiones muy semejantes al habla humana corriente, en las cuales no se omite ni el timbre de base ni la composición tonal de frases. Se grabó, con este procedimiento, un disco con la recitación, por parte de la máquina, del célebre monólogo de Hamlet y una canción tradicional inglesa (*Daisy, Daisy*) cantada por el conjunto de máquinas cibernéticas de aquel laboratorio.

11.5. El reconocimiento de los caracteres escritos

El reconocimiento de los caracteres plantea una problemática informacional diferente. En la bibliografía especializada aparecen ya numerosos datos relativos a los programas propuestos con este fin. En algunos casos, se trata de métodos topológicos, en otros, de métodos basados en analogías eléctricas, etcétera². Más recientemente, han adquirido un especial desarrollo los métodos fundados en el uso de los ordenadores digitales, para cuya comprensión estableceremos, de momento, algunas ideas básicas.

Reconocer un objeto significa compararlo con los distintos elementos del repertorio (ing., *alphabet*) y dar cuenta de las diferencias entre los dos miembros de la comparación. Esto puede llevarse a cabo con facilidad cuando se trata de elementos estandarizados, a base de considerar juntos los elementos centrados y evaluando las diferencias recíprocas entre los dos, el modelo y el elemento que debe reconocerse. La diferencia es determinable a partir de una fórmula como

$$D = \int_{\Omega} |f(\bar{r}) - g(\bar{r})| dr,$$

2. Cf. HIGHLEYMAN, "Analog methods".

donde f representa la ecuación del carácter que debe reconocerse, g la ecuación del modelo con que se compara y Ω el dominio de proyección de los elementos. Se considera que f coincide con aquel elemento del repertorio para el cual la diferencia —calculada por la fórmula precedente— es mínima.

Hasta el presente, se ha admitido que los indicios de reconocimiento son conocidos. A menudo, sin embargo, esto no es así. Desde el punto de vista cibernético, el problema consiste, en este caso, en determinar justamente los elementos que permiten la identificación.

Como de costumbre, en estas condiciones se hacen intentos sucesivos, eligiendo en primer lugar los elementos de identificación según la experiencia anterior del que dirige las investigaciones. Se determina así un espacio n -dimensional de los elementos de identificación, con arreglo al cual se considera el conjunto de los elementos que deben identificarse, respectivamente de los modelos en cuestión. Si se ha hecho correctamente la elección de los indicios de reconocimiento, entonces cada modelo adoptado ofrecerá una agrupación de los valores de los indicios de reconocimiento, de modo que se obtenga una partición del dominio entero E en m subdominios, caracterizados por la propiedad

$$I_j \cap I_k = \emptyset,$$

lo que eliminará las ambigüedades de identificación.

Si no se satisface esta condición, significa que la elección de los indicios no se ha efectuado correctamente y que deben elegirse, a partir de la experiencia acumulada, otros indicios de reconocimiento. Este problema guarda vínculos estrechos con la ejemplificación del modo de formar nociones y con la taxonomía matemática, es decir con la teoría de la clasificación.

Al mismo tiempo, conviene subrayar el hecho de que en la partición del dominio E , el número de los subdominios debe ser igual al de los modelos que constituyen el repertorio con arreglo al cual se efectúa la identificación.

En la actualidad existen aparatos perfeccionados que reconocen bien formas sonoras, bien formas gráficas. Los progresos alcanzados en el terreno de los aparatos que reconocen formas sonoras han permitido construir artefactos que, conectados a una máquina de escribir, escriben un texto a partir de un dictado a velocidad normal. En cuanto a las formas gráficas, hay actualmente máquinas que pueden “leer” hasta varios centenares de letras por minuto. Se emplean para la introducción de datos en los cálculos electrónicos rápidos.

Todo ello prueba una vez más el gran desarrollo registrado en los más diversos dominios gracias a la cibernética.

En el dominio de la taxonomía, P. Constantinescu ha introducido el original método del análisis “clusterial”.

Bibliografía e índices

Bibliografía

- AJMANOVA, O. S.; MELCHUK, I. A.; PADUCHEVA, E. V. y FRUMKINA, R. M.: *O točinij metodaj issledovania iazyka (O tak nazyvalnoi "matematičeskoj lingvistike")*, [Sobre los métodos exactos en la investigación de la lengua (la llamada "lingüística matemática")], Moscú, 1961. Versión inglesa: *Exact Methods in Linguistic Research*, Berkeley y Los Ángeles, 1963.
- AJMANOVA, O. S. y MIKAELAN, G.: *The Theory of Syntax in Modern Linguistics*, La Haya, 1969. [ALARCOS LLORACH, E.: *Fonología española*, Madrid, 4.ª ed., 1971].
- ALEKSEIEV, P. M. y PIOTROVSKI, R. G. (eds.): *Chastotnoi anglo-russkii slovar minimum po elektronike*, [Pequeño diccionario anglo-ruso especial para la electrónica], Moscú, 1971.
- [ALVAR, M.: *Estructuralismo, geografía lingüística y dialectología actual*, Madrid, 1969].
- ANDREEV, N. D.: "Modelirovanie iazyka na baze ego statističeskoj i teoretiko-mnozhestvennoj struktury", [La modelación de la lengua a partir de su estructura estadística y teórico-pluralista]. *Tezisy Soveshchianii po Matematičeskoj Lingvistike*, Leningrado, 1959, pp. 15-22.
- ANDREEV, N. D.; IVANOV, V. V. y MELCHUK, I. A.: "Nekotorye zaméčania i predloženia otnositelno raboty po mashinnomu perevodu v S.S.S.R.", [Algunas notas y propuestas relativas a los problemas de traducción mecánica en la URSS], 4, 1960, pp. 3-24.
- ANDREEV, N. D. y ZINDER, L. R.: "Osnovnye problemy prikladnoj lingvistiki", [De los problemas fundamentales de la lingüística aplicada], *Voprosy Iazykoznanija*, 4, 1959, pp. 3-19.
- APOSTEL, L.; MANDELBROT, B. y MORF, A.: *Logique, langue et théorie de l'informaton*, Paris, 1957.
- APRESIAN, J. D.: "O poniatij i metodaj strukturhoi leksikologuii", [Sobre los problemas y métodos de la lexicología estructural], Moscú, 1962.
- APRESIAN, J. D.: *Idei i metody sovremennoj strukturnoj lingvistiki (Kratkii očerok)*, [Ideas y métodos sobre la lingüística estructural contemporánea (Breve resumen)], Moscú, 1966. Traducción inglesa: *Principles and Methods of Contemporary Structural Linguistics*, La Haya, 1973. Traducción francesa: *Éléments sur les idées et les méthodes de la linguistique structurale contemporaine*, Paris, 1973.
- ARGHEZI, T.: *Versuri*, [Versos], Bucarest, 1960.
- AVRAM, A.: "Despre dialectologia structurală", [Sobre la dialectología estructural], *Limba Română*, 11, 6, 1962.

- AVRAM, A.: "Vocalele românești /ă/ și /i/ din punct de vedere acustic", [Las vocales rumanas /ă/ y /i/ desde el punto de vista acústico], *Studii și Cercetări Lingvistice*, 14, 2, 1963, pp. 165-77.
- AVRAMESCU, A.: "Geometric model for selection of hierarchic levels of descriptors", *FID/RI, Committee Meeting*, Moscú, 24-7 febrero, 1970.
- BABITSKI, K. I.: "K voprosu o modelirovanii struktury prostogo predložhenia", [Sobre la modelación de la estructura de la frase simple], *Problemy Strukturnoi Lingvistiki*, Moscú, 1962, pp. 115-29.
- BACH, E.: *An Introduction to Transformational Grammars*, Nueva York, 1964.
- [BÁEZ SAN JOSÉ, V.: *Introducción crítica a la gramática generativa*, Barcelona, 1975].
- BĂLĂNESCU, T.: "Studiu matematic comparativ al unor variante ale baladei *Miorița*", [Estudio matemático comparativo de unas variantes de la balada "Mioritsa"], *Revista de Etnografie și Folclor*, 18, 2, 1973.
- BALLY, CH.: *Linguistique générale et linguistique française*, Berna, 1950.
- BAR-HILLEL, Y.: "Recursive definitions in empirical sciences", *Proceedings of the 11th Intern. Cong. of Philosophy*, 5, 1953, pp. 160-5.
- BAR-HILLEL, Y.: "A quasi-arithmetical notation for syntactic description", *Language*, 29, 1953, pp. 47-58.
- BAR-HILLEL, Y.: "Three remarks on linguistics fundamentals", *Word*, 13, 1957, pp. 323-35.
- BAR-HILLEL, Y.: "The present status of automatic translation of languages", Appendix II en *Advances Computers*, vol. 5, Nueva York, 1960.
- BAR-HILLEL, Y.: "The future of machine translation", *Freeling the Mind, Articles and Letters from Times Literary Supplement during March-June*, 1962, pp. 32-7.
- BAR-HILLEL, Y.: "Some recent results in theoretical linguistics", *Logic, Methodology and Philosophy of Science, Proceedings of the 1960 Intern. Congr.*, Standford, California, 1962, pp. 551-7.
- BAR-HILLEL, Y.: *Language and Information. Selected Essays on their Theory and Application*, Jerusalén, 1964.
- BAR-HILLEL, Y.: "Dictionaries and meaning rules", *Foundations of Language*, 3, 1967, pp. 409-14.
- BAR-HILLEL, Y.; PERLES, M. y SHAMIR, E.: *On Formal Properties of Simple Phrase-structure Languages*, Applied Logic Branch, Technical Report n. 4; Jerusalén, 1960.
- BAR-HILLEL, Y. y SHAMIR, E.: "Finite-state languages. Formal representations and adequacy problems", *Bulletin of the Research Council of Israel*, 8F, 1960, pp. 155-66.
- BAR-HILLEL, Y.; GAIFMAN, C. y SHAMIR, E.: "On categorial and phrase-structure grammars", *Bulletin of the Research Council of Israel*, 9F, 1960, pp. 1-16.
- BATOG, T.: "Logiczna rekonstrukcja pojęcia fonemu", [La reconstrucción lógica del concepto de fonema], *Studia Logica*, 11, 1961.
- BATOG, T.: "Critical remarks on Greenberg's axiomatic phonology", *Studia Logica*, 12, 1961, pp. 195-205.
- BATOG, T.: *The Axiomatic Method in Phonology*, Londres, 1967.
- BECKMANN, P.: *The Structure of Language. A New Approach*, Boulder, Colorado, 1972.

- BEKTAEV, K. B.: *Statistika rech, 1957-1972*, [Estadística del habla, 1957-72], Alma-ata, Institut Iazykoznania AN Kazajkoi SSR.
- BELETSKI, M. I.: "Neskolko modelei iazyka", [Algunos modelos de lengua], *Doklady na Konferentsii po Obrabotke Informatsii Mashinnomu Perevodu i Avtomaticheskomu Chiteniu Teksta*, Moscú, 1961.
- BELEVITCH, V.: *Langage des machines et langage humain*, Bruselas, 1956.
- BENZÉCRI, J. P.: *Cours de linguistique mathématique* (Tesis), Rennes, 1964.
- BERG, A. I.: "Despre unele probleme ale ciberneticii", [Sobre algunos problemas de la cibernética], *Probleme de Filozofie*, 5, 1960, p. 68.
- BIERWISCH, M.: "Über den theoretischen Status des Morphems", *Studia Grammatika*, 1, 1962, pp. 51-89.
- [BIERWISCH, M.: *El estructuralismo. Historia, problemas, métodos*, Barcelona, 3.^a ed., 1974].
- BLOCH, B.: "A set of postulates for phonemic analysis", *Language*, 24, 1948, pp. 3-46.
- BLOCH, B. y TRAGER, G. L.: *Outline of Linguistic Analysis*, Baltimore, 1942.
- BLOOMFIELD, L.: "A set of postulates for the science of language", *Language*, 11, 1926, pp. 153-64.
- BLOOMFIELD, L.: *Language*, Baltimore, 1933. Traducción española: *Lenguaje*, Lima, 1964.
- BOIANGIU, D.; NICOLAU, E. y NIȚĂ, C.: *Instrumente, dispozitive și mașini matematice*, [Instrumentos, dispositivos y máquinas matemáticas], Bucarest, 1963.
- BOLOCAN, G.: "Unele caracteristici ale studiului publicistic al limbii române literare", [Algunas características del estudio publicístico de la lengua rumana literaria], *Studii și Cercetări Lingvistice*, 12, 1, 1961, pp. 37-51.
- Bolshaia Sovetskaia Entsiklopedia*, [Gran Enciclopedia Soviética], vol. 2, s. v. *algoritm*. Traducción rumana en *Analele Româno-sovietice, Seria Matematica-Fizică*, 2, 1956, pp. 122-3.
- BOOTH, A. D.: "The practical realization of machine translation", *Methods*, 8, 1956, pp. 23-33.
- BOOTH, A. D.; BRANDWOOD, L. y CLEAVE, J. P.: *Mechanical Resolution of Linguistic Problems*, Nueva York, 1958.
- BOTTENBRUCH, H.: "Structure and use of ALGOL-60", *Journal of the Association for Computing Machinery*, 9, 2, 1962.
- BOWDEN, B. V. y PITMAN, I. (eds.): *Faster than Thought. A Symposium on Digital Computing Machines*, Londres, 1953.
- BRAINERD, B.: *Introduction to the Mathematics of Language Study*, Nueva York, 1972.
- BRAINERD, B.: *Weighing Evidence in Language and Literature. A Statistical Approach*, Toronto, 1975.
- BRATCHIKOV, I. L.: "Nekotorye teoremy formalnoi morfologii", [Algunos teoremas de morfología formal], *Konferentsii po Obrabotke Informatsii Mashinnomu Perevodu i Avtomaticheskomu Chiteniu Teksta*, Moscú, 1961.
- BRATCHIKOV, I. L.; FITIALOV, S. J. y CHEITIN, G. S.: "O strukture informatsii dlia mashinnogo perevodu", [Sobre la estructura de la información en la traducción mecánica], *Konferentsii po Obrabotke Informatsii Mashinnomu Perevodu i Avtomaticheskomu Chiteniu Teksta*, Moscú, 1961.

- BRILLOUIN, L.: *Science and Information Theory*, Nueva York, 1956.
- BUNIAKOVSKI, V. I.: "O vozmozhnosti vvedeniia opredelitelnykh mer doveriia k rezul'tatam nekotorykh nauknabliu datelnykh i preimushchestvenno statistiki", [Sobre la posibilidad de dirigir unas medidas teóricas definidas que lleven a los resultados de datos científicos con prioridad estadística], *Sovremennik*, 3, 1847, p. 48.
- BURTON, N. G. y LICKLIDER, J. C.: "Long-range constraints in the statistical structure of printed English", *American Journal of Psychology*, 68, 4, 1955.
- [BUSTOS, E. de: "Un nuevo recuento del vocabulario español", *Filología Moderna*, 25-6, 1966-7 (separata)].
- CAMION, P.: "Analyse algébrique élémentaire du critère de Lecerf-Ihm", *Rapport GRISA, Euratom*, 3, 1960, pp. 1-17.
- CANTINEAU, J.: "Le classement logique des oppositions", *Word*, 11, 1955, pp. 1-9.
- CARNAP, R.: *Logical Syntax of Language*, Nueva York, 1938.
- CARNAP, R.: *Meaning and Necessity*, Chicago, 1956.
- CARROLL, J. B.: "An operational model for language behavior", *Anthropological Linguistics*, 1, 1959, pp. 37-54.
- [CERDA, R.: "Probabilidad y determinación etimológica", *Revue Roumaine de Linguistique*, Bucarest, 22, 6, 1977].
- CHAO, YUEN REN: "Models in linguistics and models in general", *Logic, Methodology and Psychology of Science, Proceedings of the 1960 Intern. Congress*, Stanford, California, 1962, pp. 558-66.
- CHAVASSE, P.: "L'application des moyennes d'analyse de la qualité des transmissions télégraphiques", *La Cybernétique*, Paris, 1951, pp. 161-95.
- CHEBANOV, S. G.: "O podchinenii recevnykh ukladov indo-evropeiskoi gruppy zakoma Puassona", [Sobre la conformidad del tipo de estructura de las lenguas indoeuropeas a la ley de Poisson], *Akademia Nauk SSSR, Doklady*, [Reseñas], 55, 2, 1947, pp. 103-9.
- CHEITIN, G. S.: "K voprosu o postroenii matematicheskikh modelei iazyka", [Sobre la cuestión de la construcción del modelo matemático de la lengua], *Doklady na Komferentsii po Obrabotke Informatsii Mashinnomu Perevodu i Avtomaticheskomu Chiteniu Teksta*, Moscú, 1961.
- CHERRY, E. C.; HALLE, M. y JAKOBSON, R.: "Towards the logical description of languages in their phonemic aspects", *Language*, 29, 1, 1953, pp. 34-46.
- CHOMSKY, N.: "Three models of the description of language", *IRE Transaction on Information Theory*, IT-2, 3, 1956, pp. 113-24.
- CHOMSKY, N.: "On certain formal properties of grammars", *Information and Control*, 2, 1959, pp. 137-67.
- CHOMSKY, N.: "Explanatory models in linguistics", *Logic, Methodology and Psychology of Science, Proceedings of the 1960 Intern. Cong.*, Stanford, California, 1962.
- CHOMSKY, N.: "On the notion rule of grammar", *Proceedings of the Symposium on the Structure of Language and its Mathematical Aspects*, 1961, pp. 6-24.
- CHOMSKY, N.: "Some methodological remarks on generative grammar", *Word*, 17, 1961, pp. 219-39.

- CHOMSKY, N.: "Formal properties of grammars", *Handbook of Mathematical Psychology*, 1963, pp. 323-428.
- CHOMSKY, N.: *Syntactic Structures*, La Haya, 1957. Versión española: *Estructuras sintácticas*, México, 1974.
- CHOMSKY, N.: *Aspects of the Theory of Syntax*, Cambridge, Mass., 1965. Versión española: *Aspectos de la teoría de la sintaxis*, Madrid, 1970.
- CHOMSKY, N. y MILLER, G. A.: "Finite state languages", *Information and Control*, 1, 1958, pp. 91-112.
- CHOMSKY, N. y MILLER, G. A.: "Finitary models of language users", *Handbook of Mathematical Psychology*, 2, 1963, pp. 429-91.
- [CHOMSKY, N. y HALLE, M., *The Sound Pattern of English*, Nueva York, 1968. Versión francesa refundida: *Principes de phonologie générative*, Paris, 1973].
- CHURCH, A.: *The Calculi of Lambda-conversion*, Princeton, 1951.
- CIOCULESCU, Ș.: "Valori muzicale in opera lui Sadoveanu", [Valores musicales en la obra de Sadoveanu], *Varietăți Critice*, Bucarest, 1966, pp. 350-61.
- CONSTANTINESCU, I.; CONDREA, S. y NICOLAU, E.: *Teoria informației*, [Teoría de la información], Bucarest, 1958.
- COOPER, W. S.: *Set Theory and Syntactic Description*, Londres-La Haya-Paris, 1964.
- CORSTIUS, H. B.: *Exercises in Computational Linguistics*, Amsterdam, 1970.
- COSERIU, E.: *Sincronía, diacronía e historia*, Montevideo, 1958. Reedición: Madrid, 1974.
- [COSERIU, E.: "Critique de la glottochronologie appliquée aux langues romanes", *Actes du Xe Congr. Intern. de Linguistique et de Philologie Romanes*, Estrasburgo, 1962, pp. 87-96].
- COSERIU, E.: *Einführung in die transformationelle Grammatik*, Tubinga, 1968.
- COTTERET, J. M. y MOREAU, R.: *Le vocabulaire du Général De Gaulle*, Paris, 1969.
- CUCIUREANU, S.: "Versul enigmă din Floarea Albastră", [El verso enigma de *La flor azul*], *Limba Română*, 9, 3, 1960, pp. 54-61.
- ČULIK, K.: "Some notes on finite states languages and events represented by finite automata using labelled graphs", *Časopis pro Pěstovani Mat.*, 86, 1961, pp. 43-55.
- DAVIS, M.: *Computability and Unsolvability*, Nueva York, 1958.
- DELAVENAY, E.: *La machine à traduire*, Paris, 1959.
- DELAVENAY, E. y K.: *Bibliography of Mechanical Translation*, La Haya, 1960.
- DIACONESCU, P.: "Pe marginea unor lucrări despre morfem", [Al margen de unos trabajos sobre el morfema], *Studii și Cercetări Lingvistice*, 4, 1962, pp. 519-41.
- DOBRUSHIN, R. L.: "Elementarnaia grammaticheskaia kategorija", [La categoría gramatical elemental], *Biuletén Obedinenia po Problema Mashinnogo Perevoda*, Moscú, 5, 1957, pp. 19-57.
- DOBRUSHIN, R. L.: "Matematicheskie metody v lingvistike", [Métodos matemáticos en lingüística], *Matematicheskie Prosveshchenie*, 6, 1961, pp. 37-40.
- DOLEŽEL, L.: "Význam teorie informace pro marxistickou jazykovědu", [La importancia de la

- teoria de la información para la lingüística marxista], *Problémy Marxistické Jazykovědy*, Praha, 1962, pp. 61-5.
- DOMONKOŞ, E.: *Asupra algoritmului de traducere automată din limba engleză în limba română*, [Sobre el algoritmo de traducción automática de la lengua inglesa a la lengua rumana], Bucarest, 1962.
- [DONZÉ, R.: *La grammaire générale et raisonnée de Port-Royal*, Berna, 1967. Traducción española: *La gramática general y razonada de Port-Royal*, Buenos Aires, 1970].
- DOPUHANOV, M. P.: *Vvedenie v teoriu peredachi informatsii*, [Introducción a la teoría de la transmisión de la información], Moscú, 1955.
- DUNN, H. K.: "Methods of measuring vowel formant bandwidths", *JASA* 33, 1961, pp. 1737-46.
- EBELING, C.: *Linguistic Units*, La Haya, 1960.
- EMINESCU, M.: *Poezii*, [Poesías], Bucarest, 1953.
- ERMILOVA, E.: "Poezia și matematica", [La poesía y la matemática], *Viața Literară Sovietică*, 11, 1962, pp. 45-54.
- EVEY, R. JAMES: *The Theory and Applications of Pushdown Store Machines*, Cambridge, Mass., 1963.
- FADEEV, D. K.: "K poniatiiu entropii konechnoi veroiatnostnoi sjemy", [Sobre el concepto de entropía y de la probabilidad del esquema], *Uspeji Matematicheskij Nauk*. Traducción rumana en *Probleme de Cibernetică*, Bucarest, 1959, pp. 514-20.
- [FERNÁNDEZ GONZÁLEZ, R.; HERVÁS, S. y BÁEZ, V.: *Introducción a la semántica*, Madrid, 1977].
- FISCHER, W.: "Topologische Stylcharakteristiken von Texten", *Grundlagenstudien aus Kybernetik und Geisteswissenschaft*, 10, 4, 1969, pp. 111-9.
- FISCHER, W.: "Beispiele für topologische Stylcharakteristiken von Texten", *Grundlagenstudien aus Kybernetik und Geisteswissenschaft*, 11, 1, 1970, pp. 1-11.
- FITALOV, S. I.: "Formal'no-matematicheskie modeli iazykov i struktura algoritmov perevoda", [Modelos matemáticos de las lenguas y la estructura de los algoritmos de traducción], *Tezisy Soveshchenia po Matematicheskoi Lingvistike*, Leningrado, 1959. Traducción rumana en *Probleme de Lingvistică Matematică*, Bucarest, 1960, pp. 33-4.
- FITALOV, S. I.: "O postroenii formalnoi morfologii v sviazi s mashinom perevodom", [Sobre la construcción de la morfología formal a la luz de la traducción mecánica], *Doklady na Konfertsii po Obrabotke Informatsii Mashinnomu Perevodu i Avtomaticheskomu Chiteniu Texta*, Moscú, 1961.
- FITALOV, S. I.: "O modelirovania sintaksisa v strukturnoi lingvistike", [Sobre la modelación de la sintaxis en lingüística estructural], *Problemy Strukturnoi Lingvistiki*, 1962, pp. 100-14.
- FOTINO, S. y MARCUS, S.: "Gramatica basmului", I, [La gramática del cuento], *Revista de Etnografie și Folclor*, 18, 4, 1973.
- FOTINO, S. y MARCUS, S.: "Gramatica basmului", II, *Revista de Etnografie și Folclor*, 19, 5, 1973, pp. 349-63.
- FOWLER, M.: "Herdan's statistical parameter and the frequency of English phonemes", en Pulgram (ed.), *Studies Presented to Ioshua Whatmough on his 60th Birthday*, pp. 47-52.
- FREUDENTHAL, H.: "Toward a cosmic language", *Yale Scient. Mag.*, 35, 3, 1960, pp. 6-11.

- FREUNDENTHAL, H.: *Lincos, Design for a Language for Cosmic Intercourse*, Amsterdam, 1960.
- FRISCH, H. L. y WASSERMAN, E.: "Chemical topology", *Journal of the American Chemical Society*, 83, 1961, pp. 3789-95.
- FROGER, D.: *La critique des textes et son automatisaton*, Paris, 1968.
- FUCKS, W.: "Mathematical theory of word-formation", *Information Theory, Third London Symposium* (Colin, ed.), Londres, 1956, pp. 154-70.
- GAIFMAN, C.: *Dependency Systems and Phrase-structure Systems*, Santa Mónica, California, 1961.
- GAMOW, G.: *Unu, doi, trei, ... infinit*, [Uno, dos, tres, ... infinito], Bucarest, 1958.
- [GARCÍA HOZ, V.: *Vocabulario usual común y fundamental*, Madrid, 1953].
- GARVIN, P.: *On Linguistic Method*, La Haya, 1964.
- GILBERT, E. N.: "Random plane net works", *Journal of the Society for Industrial and Applied Mathematics*, 9, 1961, pp. 533-43.
- GINESTIER, P.: *Le théâtre contemporain dans le monde*, Paris, 1961.
- GINSBURG, S.: *The Mathematical Theory of Context Free Languages*, Nueva York, 1966.
- GLADKI, A. V.: *Formalnye grammatiki i iazyki*, [Gramáticas y lenguas formales], Moscú, 1973.
- GLADKI, A. V. y MELCHUK, I. A.: *Elementy matematicheskoi lingvistiki*, [Elementos de lingüística matemática], Moscú, 1969. Traducción española: *Introducción a la lingüística matemática*, Barcelona, 1972.
- GLEASON, A. H.: *An Introduction to Descriptive Linguistics*, Nueva York, 1955. Traducción española: *Introducción a la lingüística descriptiva*, Madrid, 1970.
- GOLOPENȚA, S. y TOMA, P.: "Statistica și stilurile limbii", [La estadística y los estilos de la lengua], *Limba Română*, 4, 1960, pp. 58-65.
- GOODMAN, N.: "Graphs for linguistics", *Structure of Language and its Mathematical Aspects*, *PSAM*, vol. 12, 1962, pp. 51-5.
- GORUN, I.: "Une application de l'algèbre homologique à l'étude des textes poétiques", *Cahiers de Linguistique Théorique et Appliquée*, 10, 2, 1973, pp. 175-83.
- [GOUGENHEIM, G. et al.: *Problèmes de la traduction automatique*, Paris, 1968].
- GRAUR, A.: *Încercare asupra fondului principal lexical al limbii române*, [Ensayo sobre el fondo principal léxico de la lengua rumana], Bucarest, 1954.
- GRAUR, A. y WALD, L.: *Scurtă istorie a lingvisticii*, [Breve historia de la lingüística], Bucarest, 1965.
- GREENBERG, J. H.: *An Axiomatization of the Phonologic Aspect of Language*, Nueva York, 1959.
- GREGORCZYK, A.: *Fonctions récursives*, Paris, 1961.
- GREIBACH, S.: *Inverse of Phrase Structure Generators*, Cambridge, Mass., 1963.
- GREIMAS, A. J.: *Sémantique structurale. Recherche et méthode*, Paris, 1966. Traducción española: *Semántica estructural*, Madrid, 1971.
- GRENIEWSKI, H.: *Elementy Cybernetyki*, [Elementos de cibernética], Varsovia, 1959.

- GROSS, M.: *Mathematical Models in Linguistics*, Prentice-Hall, 1972.
- GROSS, M. y LENTIN, A.: *Notions sur les grammaires formelles*, Paris, 1968.
- GRUMĂZESCU, M.: "Studii asupra foneticii statistice a limbii române", [Estudios sobre la fonética estadística de la lengua rumana], *Comunicările Academiei R. P. R.*, 10, 1955, pp. 1429-37.
- GRUMĂZESCU, M.: "Frecvența sunetelor în limba română", [La frecuencia de los sonidos en la lengua rumana], *Comunicările Academiei R. P. R.*, 6, 1955.
- GUIRAUD, P.: *Les caractères statistiques du vocabulaire*, Paris, 1953.
- GUIRAUD, P.: *Problèmes et méthodes de la statistique linguistique*, Paris, 1960.
- GUTENMACHER, L. I.: "Modelirovanie matematicheskoe", [La modelación matemática], *Bolshaja Sovetskaia Entsiklopedia*, 28, 1954.
- HADLICH, R. L.: *A Transformational Grammar of Spanish*, Prentice-Hall, 1971. Traducción española: *Gramática transformativa del español*, Madrid, 1973.
- HALLE, M.: "On the role of simplicity in linguistic description", *Structure of Language and its Mathematical Aspects, Proceedings of Symposia in Applied Mathematics*, 12, 1961, pp. 89-94.
- HAMMING, R. W.: "Error-detecting and error-correcting codes", *Bell Syst. Techn.*, 29, 1950, pp. 147-60.
- HARARY, F. y PAPER, H. H.: "Toward a general calculus of phonemic distribution", *Language*, 33, 1957, pp. 143-69.
- HARKEVICH, A. A.: *Ocherki obshchei teorii svyazi*, [El concepto de entropía sobre la teoría de probabilidades], Moscú, 1955.
- HARPER, K. E. y HAYS, D. G.: "The use of machines in the construction of a grammar and computer program for structural analysis", *Proceedings of the Intern. Cong. of Information Processing*, Unesco, Paris, 1959.
- HARRIS, J. W.: *Spanish Phonology*, Cambridge, Mass., Londres, 1969.
- HARRIS, Z. S.: "Co-occurrence and transformation in structural linguistics", *Language*, 1957, pp. 283-340.
- HARRIS, Z. S.: "The transformational model of language structure", *Anthropological Linguistics*, 1, 1959, pp. 27-9.
- HARRIS, Z. S.: *Methods in Structural Linguistics*, Chicago, 1951. Refundido en *Structural Linguistics*, Chicago, 1961.
- HARRIS, Z. S.: *String Analysis of Sentence Structure*, La Haya, 1962.
- HARTUNG, W.: "Die Passivtransformationen in Deutschen", *Studia Grammatika*, 1, 1962, pp. 90-114.
- HARWOOD, F. W.: "Axiomatic syntax. The construction and evaluation of a syntax calculus", *Language*, 31, 1955, pp. 409-13.
- HAUGEN, E.: "Directions in modern linguistics", *Language*, 2, 1951.
- HAYS, D. G.: *Grouping and Dependency Theories*, Santa Mónica, California, 1960.
- HAYS, D. G.: *Basic Principles and Technical Variations in Sentence Structure Determination*, Santa Mónica, California, 1960.

- HAYS, D. G.: *Readings in Automatic Language Processing*, Nueva York, 1966.
- HERDAN, G.: *Language as Choice and Chance*, Groninga, 1956.
- HERDAN, G.: *Type-token Mathematics*, La Haya, 1960.
- HERDAN, G.: *The Calculus of Linguistic Observations*, La Haya, 1962.
- HIGHLEYMAN, W. H.: "An analog method for character recognition", *Trans. IRE-EC-10*, 1961, pp. 510-2.
- HIRSCHBERG, L. y LYNCH, I.: *Discussions sur l'hypothèse de projectivité*, Rapport CETIS, Euratom, 35, 1961.
- HIZ, H. I.: *The Intuitions of Grammatical Categories*, Univ. de Pennsylvania, 29.
- HJELMSLEV, L.: "La catégorie des cas", *Acta Jutlandica*, 7, 1935, p. 81.
- HJELMSLEV, L.: *Omkring sprogteoriens grundlaeggelse*, Copenhagen, 1943. Traducción española: *Prolegómenos a una teoría del lenguaje*, Madrid, 1971.
- HOCKETT, CH. F.: *A Course in Modern Linguistics*, Nueva York, 1958. Traducción española: *Curso de lingüística moderna*, Buenos Aires, 1971.
- HOCKETT, CH. F.: "Two models of grammatical description", *Word*, 10, 1954, pp. 210-34.
[HOCKETT, CH. F.: *Language, Mathematics and Linguistics*, La Haya, 1967].
- HORECKÝ, J.; KALMÁR, L. y MARCUS, S. (eds.): *Recueil linguistique de Bratislava*, vol. 4, Bratislava, 1973.
- HUPCROFT, J. E. y ULLMAN, J. D.: *Formal Languages and their Relation to Automata*, Reading, Mass., 1969.
- IAGLOM, I. M.; DOBRUSHIN, R. L. e IAGLOM, A. M.: "Teoria informației și lingvistica", [La teoría de la información y la lingüística], *Analele Româno-sovietice, Seria Matematică-Fizică*, Bucarest, 1963.
- IAGLOM, A. M. e IAGLOM, I. M.: *Veroiatnost' i pontiata*, [Probabilidad e información], Moscú, 1960. Traducción rumana: *Probabilitate si informație*, Bucarest, 1963.
- IANOV, I.: "O logicheski skemaj algoritmov", [Sobre los esquemas lógicos de los algoritmos], *Problemy Kibernetiki*, 1, 1958, pp. 75-127.
- Inzhenernata lingvistika*, [Lingüística para ingenieros], I, II, Gos. Ped. Institut, Leningrado, 1971.
- IUSHCHENKO, E. L.: *Adresnoe programirovanie*, [Programación a direcciones], Kiev, 1963.
- IVANOV, V. V.: "Pontiatie neutralizachia v morfologii i leksike", [La noción de neutralización en la morfología y en el léxico], *Biuletén Obediunnenia po Problemam Mashinnogo Perevoda*, 5, 1957, p. 56.
- IVANOV, V. V.: *Lingvisticheskie voprosy sozdania mashinnogo dlia informatsionnoi mashinu*, [Cuestiones lingüísticas de la creación automática para máquinas informacionales], Leningrado, 1958.
- JAKOBSON, R. (ed.): *Structure of Language and its Mathematical Aspects*, en *Proceedings of the Symposia in Applied Mathematics*, 12, Providence, Rhode Island, 1961.
- JAKOBSON, R.: "Linguistics and Communication Theory", *Structure of Language and its Mathematical Aspects, PSAM*, 12, 1961, pp. 245-52.
- JAKOBSON, R.; FANT, G. y HALLE, M.: *Preliminaries to Speech Analysis*, Cambridge, Mass., 1969.

- JAKOBSON, R. y HALLE, M.: *Fundamentals of Language*, La Haya, 1956. Traducción española: *Fundamentos del lenguaje*, Madrid, 1967.
- JINCHIN, A. I.: "Poniatie entropii v teorii veroiatnosti", [El concepto de entropía en la teoría de la probabilidad], *Uspeji Matematicheskij Nauk*, 29, 1949, pp. 35-86. Traducción rumana en *Analele Româno-sovietice, Seria Matematică-Fizică*, 2, 1954.
- JUILLAND, A.: *Outline of a General Theory of Structural Relations*, La Haya, 1958.
- [JUILLAND, A. y CHANG RODRÍGUEZ, E.: *Frequency Dictionary of Spanish Words*, La Haya, 1964].
- JUILLAND, A.; EDWARDS, P. M. H. y JUILLAND, I.: *Frequency Dictionary of Rumanian Words*, La Haya, 1965.
- KALUZHNNIN, L. A.: "Ob informatsionnom iazyke dlia matematiki", [Sobre la lengua informacional por la matemática], *Prikladnaia Lingvistika i Mashinnyi Perevod*, Kiev, 1963, pp. 21-9.
- KARPOV, V. I.: *Poluprovodnikovye stabilizatory napriazhenia*, [Semiconductores estabilizadores, Tesis], Moscú-Leningrado, 1963.
- KATZ, J. J.: *Semantic Theory*, Nueva York, 1972.
- KATZ, J. J. y POSTAL, P. M.: *An Integrated Theory of Linguistic Descriptions*, Cambridge, Mass., 1964.
- KAUFMAN, B.: *Iterative Computation of String Nesting*, Universidad de Pennsylvania, nr. 20.
- KIEFER, F.: *Mathematical Linguistics in Eastern Europe*, Nueva York, 1968.
- KIMBALL, J. P.: *The Formal Theory of Grammar (Foundations on Modern Linguistics)*, Prentice-Hall, 1973.
- KING, R. D.: *Historical Linguistics and Generative Grammar*, Prentice-Hall, 1969.
- КІТОВ, А. І. y KRINIŤKI, N. A.: *Maşini electronice cifrice şi programare*, [Máquinas electrónicas cíficas y programación], Bucarest, 1963.
- KLEENE, S. G.: *Introduction to Mathematics*, Groninga, 1962.
- [KOCK, J. DE: *Introducción a la lingüística automática en las lenguas románicas*, Madrid, 1974].
- KOLMOGOROV, A. N.: "O poniatii algoritma", [Sobre el concepto de algoritmo], *Uspeji Matematicheskij Nauk*, 8, 4, 1953 (1956), pp. 175-6.
- KOLMOGOROV, A. N.: "Modelirovanie", [Modelación], *Bolshaia Sovetskaia Entsiklopedia*, 28, 1954.
- KORTLAND, F. H. H.: *Modelling the Phoneme*, La Haya, 1972.
- KRÁMSKÝ, J.: *The Word as a Linguistic Unit*, La Haya, 1969.
- KRENN, H. y MÜLLNER, K.: *Bibliographie zur Transformations-Grammatik*, Heidelberg, 1968.
- KULAGUINA, O. S.: "Ob odnom sposobe opredelnia grammaticeskij poniatii, na baze teorii mnozhestv", [Un método para la definición de los conceptos gramaticales a partir de la teoría de los quanta], *Problemy Kibernetiki*, 1, 1958, pp. 203-14.
- KUNO, S. y OETTINGER, A. G.: "Multiple-path syntactic analyzer", *Proceedings of the IFIP Congress*, 1962.
- KÜPFEMÜLLER, K.: "Die Entropie der deutschen Sprache", *Fernmelde technische Zeitschrift*, 7, 1954, pp. 265-72.

- KURYŁOWICZ, J.: "La notion de l'isomorphisme", *Travaux du Cercle Linguistique de Copenhague*, 1949, pp. 48-60.
- KUZNETSOV, A. F.; PADUCHEVA, E. V. y ERMOLEEVA, N. M.: "Ob informatsionnom iazyke dlia geometrii i algoritme perevoda s russkogo iazyka na informatsionnyi", [Sobre la lengua informacional por la geometría y los algoritmos de la traducción con la lengua rusa como informacional], *Mashiny Perevod i Prikladnaia Lingvistika*, 6, 1961, pp. 9-18.
- LAFON, C. J.: "La reconnaissance phonétique et sa mesure", *Annales de Télécommunication*, 15, 1960, pp. 27-37.
- LAMBEK, J.: "The mathematics of sentence structure", *American Mathematical Monthly*, 65, 1958, pp. 154-70.
- LAMBEK, J.: "Contributions to a mathematical analysis of the English verb-phrase", *Journal of the Canadian Linguistic Association*, 5, 1959, pp. 83-9.
- LAMBEK, J.: "On the calculus of syntactic types structure of language and its mathematical aspects", *XII Symposium in Applied Mathematics* (R. JAKOBSON, ed.), American Mathematical Society, 1961, pp. 166-78.
- LANDA, DIEGO DE: *Relation des choses de Yucatan*, Paris, pp. 128-9.
- LECERF, Y. e IHM, P.: *Éléments pour une grammaire générale des langues projectives*, Rapport GRISA, Euratom, 1, 1960.
- LEE, C. Y.: "An algorithm for path connections and its applications", *IRE Trans EC-10*, 1961, pp. 346-65.
- [LEES, R. B.: "The basis of glottochronology", *Language*, 29, 1953, pp. 113-27].
- LEES, R. B.: "The grammar of English nominalizations", *International Journal of American Linguistics*, 26, 3, 1960.
- LEKOMTSEV, I. K.: "K voprosu ob analoguiã v stroenii shem sloga i prostogo predlozhenia", [Sobre la cuestión de la analogía estructural entre la sílaba y la frase], *Problemy Strukturnoi Lingvistiki*, 1962, pp. 31-42.
- LEPSCHY, G. C.: *La linguistica strutturale*, Torino, 1966. Traducción española: *La lingüística estructural*, Barcelona, 1969.
- LIAPUNOV, A. A.: "O nekotoryj obtschij voprosaj kibernetiki", [Sobre algunas cuestiones generales de cibernética], *Problemy Kibernetiki*, 1, 1958. Traducción rumana en *Analele Româno-sovietice, Seria Matematică-Fizică*, 3, 1959, pp. 10-28.
- LIAPUNOV, A. A. y SHESTOPAL, G. A.: "Ob algoritmicheskom opisaniï protsessov upravleniã", [Sobre el concepto algorítmico de los procesos de dirección], *Matematicheskoe Proveshchenie*, 2, 1957. Traducción rumana en *Analele Româno-sovietice, Seria Matematică-Fizică*, 3, 1959, pp. 29-42.
- LIEB, H. H.: *Sprachstadium und Sprachsystem*, Stuttgart, 1970.
- LLORENTE MALDONADO DE GUEVARA, A.: *Morfología y sintaxis. El problema de la división de la gramática*, Granada, 1955. Refundido en *Teoría de la lengua e historia de la lingüística*, Madrid, 1967, pp. 215-370.
- LOCKE, W. y BOOTH, A. D.: *Machine Translation of Language*, Londres, 1955.
- LUTZEIER, P. R.: *Der "Aspekt" Welt als Einstieg zu einem nützlichen Kontextbegriff für eine natürliche Sprache*, Stuttgart, 1974.

- LYONS, J.: *Structural Semantics: An Analysis of Part of the Vocabulary of Plato*, Londres, 1963.
- [LYONS, J.: *Semantics* (2 vols.), Cambridge, 1977. Traducción española: *La semántica*, Barcelona, en prensa].
- MACREA, D.: "Frecvența fonemelor în limba română", [La frecuencia de los fonemas en la lengua rumana], *Dacoromania*, 10, 1941, pp. 40.
- MALMBERG, B.: *Nya vägar inom språkforskningen*, Estocolmo, 1962. Traducción española: *Los nuevos caminos de la lingüística*, México, 1969.
- MANDELBROT, B.: *An Informational Theory of the Statistical Structure of Language*, Londres, 1953.
- MANDELBROT, B.: *Contribution à la théorie mathématique des communications*, Paris, 1953.
- MANDELBROT, B.: "Structure formelle des textes et communication. Deux études", *Word*, 10, 1954, pp. 1-27.
- MANDELBROT, B.: "Word frequencies and Markovian models of discourse", *Structure of Language and its Mathematical Aspects, Proceedings of Symposia in Applied Mathematics*, 12, American Mathematical Society, Providence, Rhode Island, 1961, pp. 190-219.
- MANFRINO, R.: "Entropia della lingua italiana", *Alta Frequenza*, 29, 1960, pp. 4-29.
- MANOLIU, M.: "Morfeemul în lingvistica modernă", [El morfema en la lingüística moderna], *Limba Română*, 1, 3, 1963, pp. 3-15.
- MARCUS, S.: "Legi cantitative în limbă", [Leyes cuantitativas en la lengua], *Gazeta Matematică și Fizică*, 11, 4, 1960, pp. 199-208.
- MARCUS, S.: "Description, à l'aide de la théorie des ensembles, de certains phénomènes morphologiques", *Revue des Mathématiques Pures et Appliquées*, 6, 1961, pp. 735-44.
- MARCUS, S.: "Traducerea dintr-o limbă în alta, cu mașina electronică de calcul", [La traducción de una lengua a otra con la máquina electrónica de cálculo], *Gazeta Matematică și Fizică*, seria B, 3, 1961.
- MARCUS, S.: "Automates finis, progressions arithmétiques et grammaire à un nombre fini d'états", *Comptes Rendus de l'Académie des Sciences de Paris*, 256, 17, 1963, pp. 3571-4.
- MARCUS, S.: *Lingvistică matematică. Modele matematice în lingvistica*, [Lingüística matemática. Modelos matemáticos en lingüística], Bucarest, 1963. Traducción francesa ampliada: *Introduction mathématique à la linguistique structurale*, Paris, 1967.
- MARCUS, S.: "O analiză sincronică a genului gramatical", [Un análisis sincrónico del género gramatical], *Studii și Cercetări Lingvistice*, 13, 1962, pp. 337-48. Traducción inglesa: "A synchronic analysis of the grammatical gender", *Revue Roumaine de Linguistique*, 8, 1963, pp. 99-111.
- MARCUS, S.: "Modèles mathématiques pour le catégorie grammaticale du cas", *Revue des Mathématiques Pures et Appliquées*, 8, 1963, pp. 585-610.
- MARCUS, S.: "Modelarea matematică a limbii", [La modelación matemática de la lengua], *Limba Română*, 5, 1963.
- MARCUS, S.: *Gramatici și automate finite*, [Gramáticas y autómatas finitos], Bucarest, 1964.
- MARCUS, S.: "Structures linguistiques et structures topologiques", *Revue des Mathématiques Pures et Appliquées*, 6, 1964, pp. 501-6.

- MARCUS, S.: *Algebraic Linguistics. Analytical Models*, Nueva York-Londres, 1967.
- MARCUS, S.: "Trois types d'écarts syntaxiques et trois types de figures dans le langage poétique", *Cahiers de Linguistique Théorique et Appliquée*, 7, 1970, pp. 181-8.
- MARCUS, S.: *Poetica matematică*, [La poética matemática], Bucarest, 1970.
- MARCUS, S.: "On types of meters of a poem and their informational energy", *Semiotica*, 4, 1, 1971, pp. 31-6.
- [MARCUS, S.: *Din gândirea matematică românească*, [Del pensamiento matemático rumano], Bucarest, 1975].
- MARCUS, S. y VASILIU, E.: "Matematică și fonologie. Teoria grafelor și consonantismul limbii române", [Matemática y fonología. La teoría de los grafos y el consonantismo de la lengua rumana], *Fonetică și Dialectologie*, 3, 1961, pp. 15-55.
- MARKOV, A. A.: "Teoria algoritmov", [La teoría de los algoritmos], *Trudy Matematicheskii Instituta im. Steklova*, 38, 1951, pp. 176-89 y 42, 1954.
- MARTINET, A.: *Économie des changements phonétiques*, Berna, 1955. Traducción española: *Economía de los cambios fonéticos (Tratado de fonología diacrónica)*, Madrid, 1974.
- MARTINET, A.: *Éléments de linguistique générale*, Paris, 1960. Traducción española: *Elementos de lingüística general*, Madrid, 1965.
- MARTINET, A.: *A Functional View of Language*, Londres, 1962. Traducción española: *El lenguaje desde el punto de vista funcional*, Madrid, 1971.
- MASTERMAN, M.: *A Theory of the Semantic Basis of Human Communication. Applied to the Phonetics of Intonational Form*, 21, 1960.
- MEYER-EPPLER, W.: "Zur Anwendung informationstheoretischer Methoden auf sprachliche Probleme", *Lexis*, 4, 1954, p. 30.
- MILLER, G. A.: *Language and Communication*, Nueva York, 1957. Traducción española: *Lenguaje y comunicación*, Buenos Aires, 1974.
- MITCHELL, R. P.: "Properties of a class of categorial grammars", *Reprints of Papers for the 9th Intern. Cong. of Linguists*, Cambridge, Mass., 1962, p. 68.
- MIZUTANI, S.: "On algebra of the junction of linguistic elements", *Reprints of Papers for the 9th Intern. Cong. of Linguists*, Cambridge, Mass., 1962.
- MOISIL, G. C.: "Traducerea automată", [La traducción automática], *Limba Română*, 1, 1960.
- MOISIL, G. C.: "Probleme puse de traducerea automata. Conjugarea verbelor în limba română scrisă", [Problemas planteados por la traducción automática. La conjugación de los verbos en la lengua rumana escrita], *Studii și Cercetări Lingvistice*, 11, 1, 1960.
- MOISIL, G. C.: "Preliminariile traducerilor automate", [Los preliminares de las traducciones automáticas], *Limba Română*, 9, 1, 1960, pp. 3-27.
- MOISIL, G. C.: "La linguistique mathématique dans la République Populaire Roumaine", *La Traduction Automatique*, 1, 5, 1960, pp. 37-9.
- MOISIL, G. C.: "Problèmes posées par la traduction automatique. La déclinaison dans la langue roumaine écrite", *Cahiers de Linguistique*, 1, 1962.
- MOLOSHNAIA, T. N.: "Voprosy razlichenia omonymov pri mashinnom perevode angliiskogo iazyka na russkii", [El problema de la distinción de los homónimos en la traducción a máquina

- del inglés al ruso], *Problemy Kibernetiki*, 1, 1958. Traducción rumana en *Analele Româno-sovietice, Seria Matematică-Fizică*, 3, 1959, pp. 56-62.
- MOLOSHNAIA, T. N.: "O poniatii grammaticheskoi konfiguratsii", [Sobre el concepto de configuración gramatical], *Strukturno-tipologicheskoi Issledovania*, Moscú, 1963, pp. 46-59.
- MOSKOVICH, V. A.: *Statistika i semantika*, [Estadística y semántica], Moscú, 1969.
- MOTSCH, W.: *Zur Stellung der Wortbildung in einem formalen Sprachmodell*, Thesen über die theoretische Grundlagen einer wissenschaftliche Grammatik, Deutsche Akademie der Wissenschaften, Berlín, 1961.
- [MOUNIN, G.: *Les problèmes théoriques de la traduction*, Paris, 1963. Traducción española: *Los problemas teóricos de la traducción*, Madrid, 1971].
- [MOUNIN, G.: *Histoire de la linguistique des origines au XXe siècle*, Paris, 1967. Traducción española: *Historia de la lingüística desde los orígenes al siglo XX*, Madrid, 1968].
- [MOUNIN, G.: *La linguistique du XXe siècle*, Paris, 1972].
- MUJIN, A. M.: *Funktsionalnyi analiz sintaksicheskij elementov*, [Análisis funcional de los elementos sintácticos], Moscú-Leningrado, 1964.
- [MULJAČIĆ, Ž.: *Fonologia generale e fonologia della lingua italiana*, Bolonia, 1969. Traducción española: *Fonología general. Revisión crítica de las nuevas corrientes fonológicas*, Barcelona, 1974].
- MÜLLER, Ch.: *Initiation à la statistique linguistique*, Paris, 1968. Traducción española: *Estadística lingüística*, Madrid, 1973.
- NAIR, K. K. y RAMACHANDRAN, V.: "A note on the statistical studies in Kannada speech sounds", *J. Inst. Telecomm. Eng.*, 4, 4, 1957, pp. 197-8.
- [NAVARRO, T.: *Estudios de fonología española*, Syracuse, N. Y., 1946].
- NEBESKÝ, L.: *Algebraic Properties of Trees*, Universita Karlova, Praga, 1969.
- NICOLAU, E.: "Cibernetica și lingvistica", [La cibernética y la lingüística], *Studii și Cercetări Lingvistice*, 10, 4, 1958, pp. 471-82.
- NICOLAU, E.: "Utilizarea calculatoarelor electronice de lingvistică, cu privire specială asupra traducerii mecanice", [El empleo de los ordenadores electrónicos por la lingüística, con especial atención a la traducción automática], en *Probleme de Automatizare*, 3, Bucarest, 1960.
- NICOLAU, E.: "Einige Untersuchungen zur Nachrichtenübertragung dur Vermittlung der Sprache", *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 14, 2, 1961, pp. 173-87.
- NICOLAU, E.: "Asupra calculării entropiei", [Sobre el cálculo de la entropía], *Studii și Cercetări de Matematică*, Cluj, 13, 1962, pp. 181-90.
- NICOLAU, E.: *Întroducere în cibernetică*, [Introducción en la cibernética], Bucarest, 1964.
- NICOLAU, E.: "Langage et stratégie", *Cahiers de Linguistique Théorique et Appliquée*, 1, 1966, pp. 153-77.
- NICOLAU, E.: "Observații asupra limbajului științific", [Observaciones sobre el lenguaje científico], *Omagiu lui Alexandru Rosetti*, Bucarest, 1966, pp. 623-6.
- NICOLAU, E.; SALA; C. y ROCERIC, A.: "Observații asupra entropiei limbii române", [Observa-

- ciones sobre la entropía de la lengua rumana], *Studii și Cercetări Lingvistice*, 10, 1, 1959, pp. 35-53.
- NICOLAU, E. y BĂLĂCEANU, C.: *Cibernetica*, [La cibernética], Bucarest, 1961.
- NICOLAU, E.; WEBER, I. y GAVĂȚ, Ș.: "Aparate pentru recunoașterea automată a vocalelor", [Aparatos para el reconocimiento automático de las vocales], *Automatica și Electronica*, 7, 6, 1963, pp. 255-61.
- NICOLAU, E.: *Omul informațional* [El hombre informacional], Yasi, 1971.
- NICOLAU, E. y BALACEANU, C.: "Semnificat - semn - semnificație; interpretări sistematice" [Significado - signo - significación; interpretaciones sistémicas], *Progresele Științei*, 11, 9, 1973, pp. 557-62.
- NICOLESCU, G. C.: "Contribuții la lămurirea "enigmatică" *Floare albastră*", [Contribuciones al esclarecimiento de la enigmática 'Floare albastră'], *Limba Română*, 12, 3, 1963, pp. 264-73.
- NIDA, E. A.: *Morphology. The Descriptive Analysis of Words*, Michigan, University Press, 1957.
- NIDA, E. A.: *A Synopsis of English Syntax*, University of Oklahoma, 1960.
- NIKONOV, V. A.: "Statistika padezhei russkogo iazyka", [La estadística de los casos en la lengua rusa], *Mashinnyi Perevod i Prikladnaia Lingvistika*, 10, 3, 1959, pp. 45-65.
- [NIVETTE, J.: *Principes de grammaire générative*, Bruselas, 1971. Traducción española: *Principios de gramática generativa*, Madrid, 1973].
- NOVAK, L. y PIOTROVSKI, R.: "Experimentul de predicție și entropia limbii române", [El experimento de predicción y la entropía de la lengua rumana], *Studii și Cercetări Lingvistice*, 19, 1968, pp. 209-36.
- NOWAKOWSKÁ, M.: *Language of Motivation and Language of Action*, La Haya, 1974.
- OETTINGER, A. G.: "Linguistics and Mathematics", *Studies Presented to Ioshua Whatmough on his 60th Birthday* (E. PULGRAM, ed.), La Haya, 1957, pp. 179-86.
- OETTINGER, A. G. et al.: "Linguistic and machine methods for compiling and updating the Harvard Automatic Dictionary", *Proceedings of the Int. Conference on Scientific Information*, Washington, 1958, pp. 137-59.
- OLSON, H. y BELAR, H.: "Phonetic typewriter", *Trans. IRE-AUS*, 1957, pp. 90-5.
- ONICESCU, O.: *Strategia jocurilor*, [La estrategia de los juegos], Bucarest, 1961.
- ONICESCU, O.: "Énergie informationnelle", *Comptes Rendus de l'Académie des Sciences de Paris*, 63, nr. 22, 1966, pp. 841-2.
- ONICESCU, O.: "Procedee de estimare comparativă a unor obiecte purtatoare de mai multe caracteristici", [Procedimientos de estimación comparativa de objetos portadores de varias características], *Revista de Statistică*, 19, 4, 1970, pp. 3-11.
- ONICESCU, O. y MIHOC, G.: *Lección de statistică matematică*, [Lecciones de estadística matemática], Bucarest, 1958.
- ORTIZ, A. y ZIERER, E.: *Set Theory and Linguistics*, La Haya, 1968.
- PADUCHEVA, E. V.: "Ob opisanií padezhnoi sistemy russkogo sushchestvitelnogo", [Sobre la descripción del sistema casual de los sustantivos rusos], *Voprosy Iazykoznaníia*, 5, 1960, pp. 104-11.
- PANOV, D. I.: "Avtomaticeskii perevod", [La traducción automática], *Izd. Akad. Nauk SSSR*,

- Moscú, 1956. Traducción rumana en *Analele Româno-sovietice, Seria Matematică-Fizică*, 4, 1956, pp. 17-46.
- PARKER-RHODES, A. F.: *The Lattice Properties of Syntactic Relation in an Open Language*, Cambridge, Inglaterra, 1961.
- PARKER-RHODES, A. F. et al. (miembros de la Cambridge Language Research Unit): *A Lattice Model of Syntactic Description*, Cambridge, Inglaterra, 1961.
- PARKER-RHODES, A. F.; WOOD, R.; MC KINNON; KAY, M. y BRATLEY, P.: *The Cambridge Language Research Unit Computers Program for Syntactic Analysis*, Cambridge, Inglaterra, M. L. 142.
- PARKER-RHODES, A. F.; WOOD, R., MC KINNON; KAY, M. y BRATLEY, P.: *The Derivation of Syntactic Relations from a Lattice Model*, Cambridge, Inglaterra, 1961.
- PELEVINA, N. F.: "Ustanovlenie etimologicheskogo tozhdestva s pomoshchiu umnozhenia veroiatinostei", [Establecimiento de la identidad etimológica con ayuda del cálculo de probabilidades; ensayo de lingüística aplicada], *Pitania Prikladnoi Lingvistiki*, Chernovsky, 1960, pp. 38-40.
- PEREBEINOS, V. I.: "K voprosu ob ispol'zovani strukturnyj metodov v leksikologii", [Sobre la cuestión de los métodos estructurales en la lexicología], *Problemy Strukturnoi Lingvistiki*, Moscú, 1962, pp. 163-74.
- PÉTER, R.: *Rekursive Funktionen*, Budapest, 1951.
- PIKE, K. L.: "Taxemes and immediate constituents", *Language*, 10, 2, 1943, pp. 65-82.
- PIOTROVSKAIA, A. A.; PIOTROVSKI, R. G. y RAZZHIVI, K. A.: "Entropia russkogo iazyka", [La entropía de la lengua rusa], *Voprosy Iazykoznania*, 6, 1962, pp. 115-30.
- PIOTROVSKI, R. G.: "Problema loculii în cuvînt al elementelor purtătoare de informații", [El problema del lugar en la palabra de los elementos portadores de informaciones], *Probleme de Lingvistică Matematică*, Biblioteca Analelor Româno-sovietice, Seria Tehnică, 79, 1960.
- PIOTROVSKI, R. G.: "O teoretiko-informatsionnyj parametraj ustnoi pismennoi form iazyka", [Parámetros teórico-informativos sobre la forma de la lengua escrita y oral], *Problemy Strukturnoi Lingvistiki*, Moscú, 1962, pp. 43-59.
- POLETAEV, I. A.: *Signal*, Sovetskoe Radio, Moscú, 1958.
- POST, E. L.: "Recursively enumerable sets of positive integers and their decision problems", *Bulletin of the American Mathematical Society*, 50, 1944, pp. 284-316.
- POSTAL, P.: *Constituent Structure. A Study of Contemporary Models of Syntactic Description*, La Haya, 1964.
- POTTIER, B.: *Recherches sur l'analyse sémantique en linguistique et en traduction automatique*, Publications Linguistiques de la Faculté des Lettres et Sciences Humaines de l'Université de Nancy, 1963.
- POYEN, J. y POYENNE, J.: *Le langage électronique*, Paris, 1960.
- PRATT, A. W.; ROBERTS, A. H. y LEWIS, K. (eds.): *Seminar on Computational Linguistics*, U. S. Department of Health, Education and Welfare, octubre, 1966.
- PULGRAM, E. (ed.): *Studies Presented to Ioshua Whatmough on his 60th Birthday*, La Haya, 1957.
- [QUESADA, D.: *La lingüística generativo-transformacional: supuestos e implicaciones*, Madrid, 1974].

- QUINE, W. V.: "Logic as a source of syntactical insights", *Structure of Language and its Mathematical Aspects, PSAM, 12, American Mathematical Society, Providence, Rhode Island, 1961*, pp. 1-5.
- RABIN, M. O. y SCOTT, D.: "Finite automata and their decision problems", *I. B. M. J. Res. Dev. III, 2, 1959*, pp. 114-25.
- RĂDOL, I.: "Studiu matematic comparativ al unor variante românești ale baladei *Meșterul Manole*", [Estudio matemático comparativo sobre unas variantes rumanas de la balada 'El maestro Manole'], *Revista de Etnografie și Folclor, 19, 3, 1974*.
- RAMAKRISHNA, B. S. y SUBRAMANIAN, R.: "Relative efficiency of English and German languages for communication of semantic content", *Trans IRE IT-4, 3, 1958*, pp. 127-8.
- RAMAKRISHNA, B. S. et al.: "Statistical studies in some Indian languages", *Journal of the Institute of Telecommunication Engineering, 4, 1, 1957*, pp. 25-35.
- REVZIN, I. I.: *Modeli iazyka*, [Modelos de la lengua], Moscú, 1962. Traducción francesa: *Les modèles linguistiques*, París, 1968.
- REVZIN, I. I.: "O nekotoryj voprosaj distributivnogo analiza i ego dal'neishei formalizatsii, (primenitel'no k sintaksicheskomu analizu)", [Sobre algunas cuestiones del análisis distribucional y su ulterior formalización (principalmente en el análisis sintáctico)], *Problemy Strukturnoi Lingvistiki*, Moscú, 1962, pp. 13-21.
- REVZIN, I. I.: "Oj odnom podjode k modeliam distributsionogo fonologicheskogo analiza", [Sobre un método de modelos para el análisis fonológico distributivo], *Problemy Strukturnoi Lingvistiki*, Moscú, 1962, pp. 80-5.
- REVZIN, I. I.: *Metod modelirovania i tipologua slavianskyj iazykov*, [El método de la modelación y la tipología de las lenguas eslavas], Moscú, 1967.
- RICHARD, P.; LÉVY, F. y VIRVILLE, M.: "Essai de description des contes merveilleux", *Ethnologie Française, 1, 3-4, 1971*, pp. 95-120.
- ROBINS, R. H.: "In defence of "word-and-paradigm"", *Transactions of the Philological Society, 1959 (1960)*, pp. 116-44.
- ROBINS, R. H.: *A Short History of Linguistics*, Londres, 1969. Traducción española: *Breve historia de la lingüística*, Madrid, 1974.
- ROCERIC-ALEXANDRESCU, A.: "Cu privire la aplicarea metodei statistice în studiul vocabularului", [En torno a la aplicación del método estadístico en el estudio del vocabulario], *Limba Româna, 3, 1960*, pp. 14-22.
- ROCERIC-ALEXANDRESCU, A.: "Recherches statistiques sur l'initiale des mots en roumain", *Cahiers de Linguistique Théorique et Appliquée, 1, 1962*, pp. 209-16.
- ROSENBLUETH, A. y WIENER, N.: "The role of models in science", *Philosophy of Science, 12, 1945*, pp. 317-20.
- RUWET, N.: *Introduction à la grammaire générative*, París, 1968. Traducción española: *Introducción a la gramática generativa*, Madrid, 1973.
- SADOVEANU, M.: *Hanu-Ancuței*, [La venta de Ancutsa], Bucarest, 1958.
- SĂHLEANU, V.: "Valoarea gnoseologică a modelelor și a metodei modelări", [El valor gnoseológico de los modelos y del método de la modelación], *Cercetări Filozofice, 9, 1962*, pp. 1221-37.
- SALOMAA, A.: *Formal Languages*, Nueva York-Londres, 1973.

- [SÁNCHEZ DE ZAVALA (ed.): *Semántica y sintaxis*, Madrid, 1974].
- SAPIR, E.: *Language*, Nueva York, 1921. Traducción española: *El lenguaje*, México, 1954 (eds. sucesivas).
- SAPOZHKOVA, M. A.: *Rechevoi signal v kibernetike i svyazi*, [La señal vocálica en la cibernética y en la comunicación], Sveazhisdats, Moscú, 1963.
- SAUSSURE, F. de: *Cours de linguistique générale*, París, 1916. Traducción española: *Curso de lingüística general*, Buenos Aires, 1945 (eds. sucesivas).
- SCHUTZENBERGER, M. P.: "A problem in the theory of context free language", *Preprints of Papers for the 9th Intern. Cong. of Linguists*, Cambridge, Mass., 1962, p. 96.
- SCHUTZENBERGER, M. P.: "On context-free languages and push-down automata", *Information and Control*, 7, 1963, pp. 246-64.
- [SERRANO, S.: "Resolución de una homonimia léxico-gramatical (catalán *pas*) por medio de un algoritmo", *Revista Española de Lingüística*, 2, 2, 1972, pp. 443-4].
- [SERRANO, S.: *Elementos de lingüística matemática*, Barcelona, 1975].
- [SERRANO, S.: *Lógica, lingüística y matemáticas*, Barcelona, 1977].
- SERVIEN, P.: *Le langage des sciences*, París, 1931.
- SERVIEN, P.: *Esthétique*, París, 1953.
- SESHU, S.: "Mathematical methods for sequential machines", *IRE Nat. Convent.*, 7, 2, 1959, pp. 4-16.
- SGALL, P.; NEBESKÝ, L.; GORALČIKOVÁ, E. y HAJIČOVÁ, E.: *A Functional Approach to Syntax in Generative Description of Language*, Nueva York, 1969.
- SGALL, P. et al.: *Cesty moderní jazykovedy*, [Las vías de la lingüística moderna], Praga, 1964.
- SHAMIR, E.: "On sequential languages", *Applied Logic Branch, Technical Report*, 7, Hebrew University, Jerusalén, 1961.
- SHAMIR, E.: "A remark on discovery algorithms for grammars", *Information and Control*, 5, 1962, pp. 246-51.
- SHANNON, C. E.: "A mathematical theory of communication", *Bell System Technical Journal*, 27, 1948, pp. 379-423 y 623-56.
- SHANNON, C. E.: "Prediction and entropy in printed English", *Bell System Technical Journal*, 30, 1951, pp. 50-64.
- SHAUMIAN, S. K.: "Lingvisticheskie voprosy kibernetiki i strukturalnaia lingvistika", [Cuestiones lingüísticas de la cibernética y lingüística estructural], *Voprosy Filozofii*, 9, 1960.
- SHAUMIAN, S. K.: *Problems of Theoretical Phonology*, La Haya, 1968.
- SHAUMIAN, S. K.: *Strukturalnaia lingvistika*, [Lingüística estructural], Moscú, 1965. Traducción inglesa: *Principles of Structural Linguistics*, La Haya, 1971.
- SHAUMIAN, S. K. (ed.): *Matematicheskaia lingvistika*, [Lingüística matemática], Moscú, 1973.
- SHAUMIAN, S. K. y SOBOLEVA, P. A.: *Applikativnaia porozhdaiushchaia model i ischislenie transformatsii v russkom iazyke*, [Modelo aplicativo-generativo y cálculo transformacional aplicado a la lengua rusa], Moscú, 1963.

- [SINGH, T.: *Great Ideas in Information Theory, Language and Cybernetics*. Traducțiune spaniolă: *Teoría de la información, del lenguaje y de la cibernética*, Madrid, 1972].
- SLAMA-CAZACU, T. y ROCERIC, A.: "Statistica fonemelor și valoarea experimentului de predicție", [La estadística de los fonemas y el valor del experimento de predicción], *Fonetică și Dialectologie*, 2, 1961, pp. 63-9.
- SLUTSKER, G. S.: "Poluchenie vsejdopustimyj variantov sintaksicheskogo analiza teksta pri pomoshchi mashiny", [Obtención de todas las variantes posibles del análisis sintáctico de un texto con ayuda de la máquina], *Problemy Kibernetiki*, 10, 1963, pp. 215-25.
- SOLOMONOFF, R.: "A new method for discovering the grammar of phrase structure languages", *Information Processing*, Paris-Munich-Londres, 1960, pp. 285-90.
- SOURIAU, É.: *Les deux cent mille situations dramatiques*, Paris, 1950.
- STATI, S.: "Categoriile de posibilitate și realitate în lingvistică", [Las categorías de posibilidad y realidad en lingüística], *Probleme de Lingvistică Generală*, vol. 3, Bucarest, 1961.
- STATI, S.: "Constituenții imediați", [Los constituyentes inmediatos], *Limba Română*, 2, 1964, pp. 206-23.
- STATI, S.: *Teoria e modo nella sintassi*, Bolonia, 1967.
- STEINHAUS, H.: "The collaboration of various sciences as illustrated by mathematics, and its role in Wrocław scientific circles", *Review of the Polish Academy of Sciences*, 1, 4, 1956, pp. 1-20.
- STEVENS, N. N.: "Toward a model for speech recognition", *Journal of the Acoustical Society of America*, 32, 1960, pp. 47-55.
- STOFF, V. A.: "Despre rolul modelelor în mecanica cuantică", [Sobre la función de los modelos en la mecánica cuántica], *Probleme de Filozofie*, 12, 1958.
- STOFF, V. A.: "Critica concepției neopozitiviste asupra rolului modelelor în cunoaștere", [La crítica de la concepción neopositivista sobre la función de los modelos en el conocimiento], *Analele Româno-sovietice, Seria Filozofie*, 3, 1961.
- SUTEU, V.: "Observații asupra frecvenței cuvintelor în operele unor scriitori români", [Observaciones sobre la frecuencia de las palabras en las obras de varios escritores rumanos], *Studii și Cercetări Lingvistice*, 10, 3, 1959, pp. 419-43.
- SUTEU, V.: "Observații asupra structurii acustice asupra vocalelor românești i, e, a, o și u", [Observaciones sobre la estructura acústica de las vocales rumanas i, e, a, o, u], *Studii și Cercetări Lingvistice*, 14, 2, 1963, pp. 179-98.
- [SWADESH, M.: "Lexico-statistic dating of prehistoric ethnic contacts", *Proceedings of the American Philosophical Society*, 96, 1952, pp. 453-63].
- TAGLIAVINI, C.: *Panorama di storia della linguistica*, Bolonia, 1970.
- TELLEGEN, B. D. H.: "Geometrical configuration and duality of electrical networks", *Philips Technical Review*, 11, 1940, pp. 324-30.
- TESNIÈRE, L.: *Éléments de syntaxe structurale*, Paris, 1959.
- [*The Principles of the International Phonetic Association*, Londres, 1949 (reimpresión, 1974)].
- THOM, R.: *Stabilité structurelle et morphogénèse. Essai d'une théorie générale des modèles*, Reading, Mass., 1972.
- TOPOROV, V. N.: "Introducerea probabilității în lingvistică și consecințele ei", [La introducción

- de las probabilidades en lingüística y sus consecuencias], *Probleme de Lingvistică Matematică*, Bucarest, 1960, pp. 11-5.
- TRAJTENBROT, B. A.: *Algoritmy i mashinnoe reshenie zadachi*, [Los algoritmos y los problemas de decisiones automáticas], Moscú, 1957.
- TRIER, J.: *Der deutsche Wortschatz im Sinnbezirk des Verstandes*, Heidelberg, 1931.
- [TRNKA, B., et al.: *El círculo de Praga*, Barcelona, 1971].
- TRUBETSKOI, N. S.: *Principes de phonologie*, Paris, 1957. Traducción española: *Principios de fonología*, Madrid, 1973.
- TURING, A. M.: "On computable numbers, with an application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, 42, 2, 1936-7, pp. 230-65; correction, *ibid.*, 43, 1937, pp. 544-6.
- UEMOV, A. I.: "Analogie și model", [Analogía y modelo], *Probleme de Filozofie*, 3, 1962, pp. 163-72.
- UHLENBECK, E. M.: *Critical Comments on Transformational-Generative Grammar*, La Haya, sin fecha.
- USPENSKI, V. A.: "K opredleniu padezha po A. N. Kolmogorova", [Sobre la definición del caso según A. N. Kolmogorov], *Biuleten Obedinenia po Problemam Mashinnogo Perevoda*, 5, 1957, pp. 11-8.
- USPENSKI, V. A.: "K opredleniu chasti rechi v teoretiko-mnozhestvennoi sisteme iazyka", [Sobre la definición de las partes del discurso en el sistema teórico-cuantitativo de la lengua], *Biuleten Obedinenia po Problemam Mashinnogo Perevoda*, 5, 1957, pp. 22-6.
- VACHEK, J.: *The Linguistic School of Prague*, Bloomington-Londres, 1966.
- VAIDA, D.: *Avtomatizatsia programirovania*, [Automatización de la programación], *Sbornik Perevodov pod Redaktsiei A. P. Ershova*, Moscú, 1961.
- VAIDA, D.: *Utilizări ale calculatoarelor electronice de la I. F. A.*, [Usos de los ordenadores electrónicos de la I. F. A.], Bucarest, 1961.
- VAIDA, D.: *Stroje na zpracování informací*, [Máquinas para la elaboración de la información], Praga, 1962.
- VAKAR, N. P.: *A Word Count of Spoken Russian: The Soviet Usage*, Ohio State University Press, 1966.
- VALT, L. O.: "Despre funcția cognitivă a reprezentărilor-model în fizica modernă", [Sobre la función cognitiva de las representaciones-modelo en la física moderna], *Analele Româno-sovietice, Seria Filozofie*, 3, 1961, p. 5.
- VARGA, D.: "Morphological analysis by help of the method of successive delimitation", *Computational Linguistics, I (Computing Center of the Ungarian Academy of Sciences)*, Budapest, 1963, pp. 223-55.
- VASILIU, E.: "Fonologicheskie opisaniie rumynskogo vokalizma", [Descripción fonológica del vocalismo rumano], *Problemy Strukturnoi Lingvistiki*, Moscú, 1962, pp. 86-91.
- VASILIU, E.: "Niveluri lingvistice și structuri transformabile", [Niveles lingüísticos y estructuras transformativas], *Studii și Cercetări Lingvistice*, 13, 1962, pp. 15-22.
- VASILIU, E.: "Gramaticile generative", [Las gramáticas generativas], *Limba Română*, 12, 1963, pp. 219-31.

- VASILIU, E.: *Outline of a Semantic Theory of Kernel Sentences*, La Haya, 1972.
- VASILIU, E. y GOLOPENȚIA-ERETESCU, S.: *The Transformational Syntax of Romanian*, La Haya-Paris, 1972.
- VIANU, T.: *Problemele metaforei și alte studii de stilistică*, [Los problemas de la metáfora y otros estudios de estilística], Bucarest, 1957.
- VIANU, T.: "Statistica lexicală și o problemă a vocabularului eminescian", [La estadística léxica y un problema del vocabulario de Eminescu], *Limba Română*, 8, 1959, pp. 25-33.
- VILLAMS, G. D.: *Sovershennyi strateg*, [El perfecto estratega], Sovetskie Radio, Moscú, 1960.
- VOEGELIN, C. F.: "Model-directed structuralization", *Anthropological Linguistics*, 1, 1959, pp. 9-25.
- WALL, R.: *Introduction to Mathematical Linguistics*, Prentice-Hall, 1972.
- WELLS, R. S.: "Immediate constituents", *Language*, 23, 1947, pp. 81-117.
- WHITFIELD, F. J.: "Criteria for a model of language", *Logic, Methodology and Philosophy of Science, Proceedings of the 1960 Intern. Congress*, Standford, California, 1962, pp. 577-83.
- YNGVE, V. H.: "A programming language for mechanical translation", *Mechanical Translation*, 5, 1958, pp. 25-41.
- YNGVE, V. H.: "A model and an hypothesis for language structure", *Proceedings of the American Philosophical Society*, 104, 1960, 5.
- YULE, G. U.: *The Statistical Study of Literary Vocabulary*, Cambridge University Press, 1944.
- ZARECHNAK, M.: "Three levels of linguistic analysis in machine translation", *Journal of Association for Computing Machinery*, 1, 1959, pp. 24-35.
- ZASORINA, L. N.: "Model imennogo sklonenia dlia russkogo pismennogo teksta", [El modelo de la declinación nominal en el ruso de los textos escritos], *Problemy Strukturnoi Lingvistiki*, Moscú, 1962, pp. 130-40.
- ZIERER, E.: *La teoría de los gráficos en la lingüística*, Lima, 1967.
- ZIFF, P.: *Semantic Analysis*, Ithaca, N. Y., 1960.
- ZINDER, L. R.: "Despre probabilitatea lingvistică", [Sobre la probabilidad lingüística], *Probleme de Lingvistica Matematică*, Bucarest, 1960, p. 189.
- ZINOVIEV, A. A.: "Despre lingvistica matematică", [Sobre la lingüística matemática], *Probleme de Filozofie*, 9, 1959.
- ZINOVIEV, A. A. y REVZIN, I. I.: "Modelul logic, mijloc al investigației științifice", [El modelo lógico, recurso de la investigación científica], *Probleme de Filozofie*, 1, 1960, pp. 101-14.
- ZIFF, G. K.: *Human Behavior and the Principle of Least Effort*, Cambridge, Mass., 1949.
- ZVEGHINȚEV, V. A.: "Ce se înțelege prin lingvistică structurală și aplicată", [Qué se entiende por lingüística estructural y aplicada], *Buletin de Informare Științifică*, 19, 1963, pp. 13-4.

Publicaciones dedicadas total o parcialmente a la lingüística matemática:

- American Journal of Computational Linguistics* (EE-UU).
- Analele Universității "Babeș-Bolyai", Seria Filologie* (Rumania).
- Aplikace Matematiki* (Checoslovaquia).
- Beiträge zur Linguistik und Informationsverarbeitung* (R. D. Alemania).
- Cahiers de Linguistique Théorique et Appliquée* (Rumania).
- Cercetări Lingvistice* (Rumania).
- Communications of the Association of Computing Machinery* (EE-UU).
- Computational Linguistics* (Hungria).
- Computers and the Humanities* (EE-UU).
- Computing Reviews* (EE-UU).
- Discrete Mathematics* (Holanda).
- Foundations of Language* (Holanda).
- Gazeta Matematică, Seriile A și B* (Rumania).
- Information and Control* (EE-UU).
- Information Processing Journal* (EE-UU).
- International Journal of Computer Mathematics* (EE-UU).
- IRE Transactions on Information Theory* (EE-UU).
- Journal of Linguistics* (Inglaterra).
- Journal of the Association of Computing Machinery* (EE-UU).
- Keiryō Kokugo Gakkai* (Japón).
- Kiberneticheski Sbornik* (U. R. S. S.).
- Langages* (Francia).
- Language* (EE-UU).
- Lenguaje y Ciencias* (Perú).
- Limba Română* (Rumania).
- Lingua e Stile* (Italia).
- Linguistic Inquiry* (Holanda).
- Linguistics* (Holanda).
- Mashinnyi Perevod i Prikladnaia Lingvistika* (U. R. S. S.).
- Mathematical Reviews* (EE-UU).
- Mathematical System Theory* (EE-UU).
- Mechanical Translation* (EE-UU).
- Nauchno-tehnicheskaia Informatsia* (U. R. S. S.).

- Poetica* (Holanda).
- Prague Studies of Mathematical Linguistics* (Checoslovaquia).
- Problemy Kibernetiki* (U. R. S. S.).
- Problemy Strukturnoi Lingvistiki* (U. R. S. S.).
- Recueil Linguistique de Bratislava* (Checoslovaquia).
- Referativnyi Zhurnal, Sección Matemáticas Problemy Semiotiki* (U. R. S. S.).
- Revue Française du Traitement de l'Information* (Francia).
- Revue Roumaine de Linguistique* (Rumanía).
- Revue Roumaine de Mathématiques Pures et Appliquées* (Rumanía).
- Semiotica* (Holanda).
- Slovo a Slovesnost* (Checoslovaquia).
- Statistical Methods in Linguistics* (Suecia).
- Strukturno-tipologicheskie Issledovanie* (U. R. S. S.).
- Studia Grammatica* (R. D. Alemania).
- Studii și Cercetări Lingvistice* (Rumanía).
- Studii și Cercetări Matematice* (Rumanía).
- T. A. Informations* (Francia).
- The Finite String* (EE-UU).
- Theoretical Linguistics* (R. F. Alemania).
- The Prague Bulletin of Mathematical Linguistics* (Checoslovaquia).
- Traduction Automatique* (Francia).
- Voprosy Iazykoznanía* (U. R. S. S.).
- Word* (EE-UU).
- Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* (R. D. Alemania).
- Zentralblatt für Mathematik und ihre Grenzgebiete* (R. F. de Alemania - R. D. Alemana).

Índice alfabético

- α -equivalencia, 86.
ACHENWALL, G., 278
adjetivo, 2, 86, 133, 154, 163, 209, 213, 214, 215, 217, 218, 220, 221, 222, 237, 255, 325, 326.
adverbio, 3, 133, 152, 153, 255.
AJMANOVA, O. S., 25, 83, 260.
ALARCOS, E., 256.
ALEKSEIEV, P. M., 226.
alemán, 4, 13, 16, 86, 258, 259, 260, 261, 272, 294, 295.
alfabeto Morse, 194, 310, 315.
álgebra, 42, 51, 53, 59, 69, 83, 123, 196.
álgebra homológica, 135.
álgebra lógica, 42.
algoritmo, 52, 54, 58, 66, 71, 79, 80, 122, 143, 144, 146, 147, 151, 152, 153, 154, 157, 158, 160, 162, 163, 164, 165, 166, 167, 170, 189, 195, 207, 210, 221, 223, 224, 322.
algoritmo de Euclides, 145, 146.
algoritmo general, 153.
algoritmo individual, 153.
algoritmo normal, 94, 158, 159, 161, 322.
algoritmos de resolución, 185-86.
algoritmos de traducción, 104.
algoritmos individuales, 167.
algoritmos numéricos, 147.
alomorfo, 21.
ALVAR, M., 113.
ambigüedad, 163.
amplitud acústica, 205, 331, 333, 337, 340.
análisis algorítmico, 150.
análisis "clusterial", 343.
análisis componencial, 39.
análisis de Fourier, 336, 337.
análisis de probabilidades, 52.
análisis distribucional, 21, 30, 35, 87, 256.
análisis en constituyentes inmediatos, 24-29, 31, 33, 35, 105.
análisis formal, 35.
análisis lógico-semántico, 11.
análisis matemático, 59.
análisis morfemático, 82.
análisis proposicional, 99.
análisis sintagmático, 84.
análisis transformacional, 89.
analogía, 2, 61, 62, 75, 78.
ANDREEV, N. D., 79, 81, 83, 84, 167.
antiguo eslavo, v. eslavo.
APOSTEL, L., 322.
Appendix Probi, 2.
APRESIAN, J. D., 8. 11.
árabe, 271.
arbitrariedad del signo lingüístico, 5.
archifonema, 59, 325.
archilexemas, 40.
ARGHEZI, T., 131, 231, 233, 275, 287, 288, 290.
ARISTÓTELES, 1.
aritmética, 43, 52, 143, 144, 171, 176, 186, 194, 198, 199, 207.
armenio clásico, 6.
arqueología, 122.
artes plásticas, 226-31.
artículo, 2, 3, 51, 124, 128, 149, 150, 163, 213, 219, 221, 222, 255, 273.
astronomía, 54, 75.
autómata con un número finito de estados, 91, 94, 170, 226.
automatización, 48, 186.
autoprogramación, 186.
AVOGADRO, A., 278.
AVRAM, A., 113, 337.
AVRAMESCU, A., 226.
BABITSKI, K. I., 83.
BÁEZ, V., 94.
BĂLĂCEANU, C., 231, 245.
BĂLĂNESCU, T., 135.
BAR-HILLEL, Y., 78, 79, 83, 94, 114, 246.
BATOG, T., 69, 83, 84.
BAUDELAIRE, CH., 129.
BECKMANN, P., 71.
BELEVITCH, V., 61, 83.
BELSKAIA, I. K., 47.
BENVENISTE, É., 6.
BERG, A. I., 55.
BERNOULLI, D., 278.
BERNOULLI, J., 278.
Biblia de Uifilas, 287.
binarismo, 56, 312.
biofísica, 55.
biología, 54, 55, 68, 79, 80, 109, 144.
bioquímica, 55.
bit, 170, 250, 251, 252, 253, 267, 291, 303, 307.
BLAGA, L., 231, 233.
BLOCH, B., 10, 22, 34.
BLOOMFIELD, L., 9, 10, 22, 25.
BOGZA, G., 233.

- BOIANGIU, D., 169.
 BOLOCAN, GH., 245.
 BOLTZMANN, L., 279.
 BOOLE, G., 42.
 BOOTH, A. D., 208.
 BOPP, F., 6.
 BOREL, É., 53.
 BOSE, J. CH., 279.
 BOYLE, R., 278.
 BRAINERD, B., 70, 71.
 BRANDT CORSTIUS, H., 71.
 BRANWOOD, L., 208.
 BRATCHIKOV, I. L., 79, 81, 82, 84.
 BRÉAL, M., 38.
 BREDICEANU, M., 226.
 BRUGMANN, K., 6.
 búlgaro, 274, 294.
 BUNIAKOVSKI, V. I., 52, 53.
 BURTON, N. G., 257, 263.
 BUSTOS, E. DE, 235.
- cadena de Markov, 62, 79, 129, 324.
 cadenas con enlaces completos, 62.
 calculador, v. ordenador.
 cálculo de probabilidades, 45, 53, 59, 83, 84, 129, 157, 164, 278, 322, 330.
 cálculo diferencial, 300.
 cálculo integral, 300.
 cálculo lógico, 167.
 cálculo sintáctico asociativo, 106.
 cálculo sintáctico no-asociativo, 107, 108.
 CĂLINESCU, G., 233.
 cambios fonéticos, 5.
 CAMION, P., 81, 83, 89.
 campo semántico, 39.
 campos léxicos, 38.
 canal de transmisión, 309, 310, 311.
 canal informacional, 266.
 CANTINEAU, J., 61.
 cantidad de información, 250, 253, 254, 255, 268.
 CARNAP, R., 246.
 caso gramatical, 149, 151, 210, 215, 218, 236, 255, 274, 313, 324, 325, 326.
 casos, 6.
 castellano, v. español.
 catalán, 5, 6.
 categoría de objetos, 331.
 categoría gramatical, 78, 86, 140, 213.
- CERDA, R., 111.
 CHAMPOLLION, J. F., 119.
 CHANG, E., 235.
 CHAO, YUEN REN, 110.
 CHERANOV, S. G., 286, 287, 288, 289.
 CHEITIN, G. S., 79, 84, 90.
 CHEJOV, A., 236.
 CHERRY, E. C., 260, 312.
 CHOMSKY, N., 10, 37, 38, 57, 58, 70, 78, 82, 83, 89, 90, 93, 94, 108, 135, 136, 258.
 CHURCH, A., 143.
 cibernética, 10, 43, 55, 79, 80, 91, 114, 117, 122, 138, 144, 155, 165, 166, 169, 207, 209, 245, 246, 277, 280, 309, 329, 342, 343.
 ciencias naturales, 236.
 CIHAC, A. DE, 282.
 CIOCULESCU, S., 131.
 circuitos en serie, 304.
 círculo lingüístico de Copenhague, v. escuela de Copenhague.
 círculo lingüístico de Praga, v. escuela de Praga.
 clase de contextos, 32.
 clase de equivalencia, 51, 92.
 clase de focos, 30.
 clase de secuencias, 29, 30, 31, 32.
 clase distribucional, 30, 31, 32, 91, 92, 114.
 clasificación tipológica, 134.
 CLEAVE, J. P., 208.
 codificación, 56, 309, 310, 311, 315, 323.
 código, 56, 81, 267, 268, 269, 270, 275, 309, 310, 311, 312, 313, 315, 316, 317, 319, 320, 322, 323, 325.
 código binario, 311, 313, 315, 321, 324, 325.
 código corrector de errores, 315-23, 324-7.
 código detector de errores, 315, 324-7.
 código electrónico, 167-79, 186.
 código grupal, 322.
 código Morse, v. alfabeto Morse.
 código operacional, 175, 176.
 coeficiente de contracción, 323.
 COEURDOUX, A.
 complementariedad distribucional, 20.
 componente armónico, 332.
- componente fundamental, 332.
 comportamiento, 213-5.
 computador, v. ordenador.
 comunicación, 267-78.
 concordancia, 214, 233, 258, 326.
 condición de cuasiproyectividad, 104.
 condición de proyectividad, 103, 104.
 condición lógica, 145, 154, 188.
 condicionamiento lógico, 16, 17.
 condiciones de Dirichlet, 332.
 condiciones lógicas, 146, 147, 150, 151, 152, 156, 157, 161, 162, 163, 165, 180, 183, 185.
 CONDON, 281.
 CONDREA, S., 314.
 configuración de rango, 101-105.
 configuración progresiva, 105.
 configuración regresiva, 105.
 configuración sintáctica, 66.
 conjugación, 2, 214.
 conjunción, 2, 3, 123, 154, 217, 220, 233.
 conjunción lógica, 197.
 conjunto inicial, 86.
 conjunto ordenado, 200.
 conmutación fonológica, 17.
 connotación, 130.
 conocimiento artístico, 109.
 conocimiento científico, 109.
 constante de Rydberg, 206.
 constante, 125, 126, 128.
 CONSTANTINESCU, I., 314.
 CONSTANTINESCU, P., 343.
 constelación lingüística, 16, 17.
 constituyente inmediato, 9, 10, 25, 26, 28, 29, 31, 32, 33, 34, 35.
 constituyente último, 25.
 construcción, 25.
 contenido, 114, 249, 311.
 contexto, 30, 31, 32, 34, 35, 83, 87, 95, 122, 127, 165, 248, 267, 268, 270, 271, 272, 273, 274.
 contigüidad, 98.
 convergencia, 90.
 coordenadas cartesianas, 238, 239, 290.
 coordenadas logarítmicas, 238, 239, 240, 241, 242, 243.
 coordinación, 24, 33-34.
 copto, 119.
 CORSTIUS, B. H., 71.
 COSERIU, E., 11, 13, 21, 39, 113.

- cosmología, 206.
 COTTERET, J. M., 285.
 COURTENAY, BAUDOIN DE, 8, 53, 255.
 CREANGĂ, I., 288, 290.
 crítica literaria, 63, 129.
 cuanta de información, 165.
 cuantificación, 124, 126, 128.
 CUCIUREANU, S., 265.
 ČULIK, K., 78, 83.
 CURRY, H. B., 67.
- danés, 9.
 datos iniciales, 171-192.
 décimo problema de Hilbert, 148.
 declinación, 2, 6, 214.
 DEDEKIND, R., 200.
 DE GAULLE, CH., 285-6.
 DELAVENAY, E., 152.
 DEMÓCRITO, I.
 denotación, 130.
 dependencia monótona, 102, 103.
 derivación, 37.
 desciframiento textual, 118, 119, 120, 195, 313.
 descodificación, 56.
 descripción algorítmica, 151, 155, 156, 157, 163, 164, 165, 166, 167.
 descripción no-modelada, 109.
 descriptivismo americano, 8, 9.
 descriptivismo lingüístico, 18, 24, 25, 28.
 descubrimiento de errores, 267.
 desigualdad, 197.
 determinación heterosintagmática, 258.
 determinación lingüística, 16.
 diacronía, 13, 14.
 diacrónica lingüística histórica, 12.
 dialectología estructural, 113.
 diccionario, 4, 208, 209, 210, 214, 215, 216, 217, 221, 235, 243, 244.
 diccionario de frecuencia, 63, 66, 235.
 diccionario electrónico, 66.
 diccionario Motul, 119, 120, 121.
 DICKENS, CH., 48.
 dicotomía lingüística, 13, 15, 26.
 DIEZ, FR., 6.
 DIONISIO TRACIO, 2, 3.
 DIRAC, P. A., 279.
 direcciones electrónicas, 172-79, 216-224.
- discurso directo, 125.
 discurso indirecto, 123-28.
 distancia, 56.
 distancia entre palabras, 299, 315, 316, 326.
 distribución lingüística, 18-21, 31, 45, 101, 245.
 disyunción, 123.
 disyunción lógica, 197.
 DOBRUSHIN, R. L., 83, 84, 86, 256, 257, 260.
 DOLEŽEL, L., 275.
 DOMONKOŠ, E., 207, 210, 223.
 DONZÉ, R., 4.
- economía informacional, 249-50.
 economía lingüística, 144, 273.
 economía ortográfica, 272.
 ecuación, 147, 148, 183.
 ecuaciones diofánticas, 147, 148.
 EDWARDS, P. M. H., 66, 244, 253.
 egipcio, 119.
 EINSTEIN, A., 279.
 electrónica, 301.
 ELIOT, T. S., 230-1.
 EMINESCU, M., 132, 245, 265, 275, 287, 288, 290.
 energía cinética, 132.
 energía informacional, 132.
 enseñanza de lenguas, 42.
 entropía, 44, 62, 63, 67, 227, 233-275, 279, 282, 290, 291, 295, 298, 299, 300, 301, 303, 307, 322, 323, 324.
 entropía absoluta, 265.
 entropía de las letras, 263-69.
 entropía de las palabras, 263-69.
 entropía de orden n , 255, 256, 257, 259, 261, 262, 263, 264, 265, 266, 322, 323.
 entropía fonológica, 255.
 entropía informacional, 279.
 entropía lingüística, 255, 261.
 entropía morfológica, 258.
 entropía relativa, 265, 266, 324.
 enunciado, 22, 23, 25, 27, 28, 29, 30, 31, 33, 34, 35, 38.
 enunciado endocéntrico, 34.
 enunciado exocéntrico, 34.
 epopeya homérica, 2.
 ergodicidad, 62.
 ERMILOVA, E., 265.
 ERMOLEVA, N. M., 83.
 escuela de Copenhague, 9.
 escuela de Praga, 8, 9, 22.
- escuela descriptivista, v. descriptivismo lingüístico.
 escuela estructuralista, v. estructuralismo lingüístico.
 escuela generativista, v. gramática generativa.
 eslavo, 6, 234, 244, 282.
 espacio, 205.
 espacio euclidiano, 205.
 espacio métrico, 321.
 espacio n -dimensional, 321, 330, 343.
 espacio vectorial, 205.
 español, 4, 5, 16, 27, 36, 40, 48, 92, 102, 115, 116, 118, 119, 120, 121, 133, 139, 152, 153, 163, 164, 236, 244, 256, 258, 259, 260, 261, 262, 270, 272, 284, 286, 291, 295.
 espectrografía, 331, 332, 333, 334, 335, 336, 337, 339.
 esquema lógico, 186.
 esquema lógico de cálculo, 184, 179-84.
 estabilidad frecuencial, 62.
 estadística matemática, 322.
 estenografía, 42, 80, 248, 271.
 estética, 129.
 estilística, 129, 280.
 estilo lingüístico, 38, 39, 233, 235, 236, 241, 245.
 estímulo, 138, 156, 162, 182, 183.
 ESTOUP, J. B., 64, 237, 241, 281.
 estrategia de los juegos, 281.
 estratificación lingüística, 22.
 estructura, 224.
 estructura estadística, 53, 59, 61, 63, 79, 234, 246, 253, 255, 262, 274, 277-305, 322, 323.
 estructura estadística de la lengua, 49, 50, 62.
 estructura estadística de las leyes, 45.
 estructura etimológica del léxico, 244.
 estructura fonológica, 56, 263, 275.
 estructura gramatical, 47, 125, 275.
 estructura léxica, 244.
 estructura lingüística, 51, 58, 59, 64, 67, 77, 108, 114, 117, 141, 165.
 estructura profunda, 38, 138.
 estructura superficial, 38.

- estructuralismo lingüístico, 8, 9, 10.
 estructuras progresivas, 106.
 estructuras regresivas, 81, 83.
 estructuras repetitivas, 135.
 etimología, 2, 3, 52, 111, 112, 141, 244.
 etnografía, 122.
 etrusco, 3.
 EULER, L., 149.
 eventos estadísticos, 277.
 expansión, 30.
 expresión/contenido, 13, 273.
 extensión de la frase, 95, 101.
 extensión de las palabras, 56, 97, 102, 264, 267, 286-9, 315, 323, 325.
 extensión de un texto, 237.
 extensión media de las palabras, 286-9, 290, 291.
 extensión textual, 238, 240.

 familias de palabras, 87, 89.
 fase acústica, 205, 331.
 FERMAT, P., 278.
 FERMI, E., 279.
 FERNÁNDEZ, A. R., 39.
 filología, 71.
 filología comparada, 52.
 filosofía, 113.
 FISCHER, W., 135.
 física, 11, 47, 54, 55, 62, 144, 247, 278.
 física nuclear, 138.
 física ondulatoria, 279.
 fisiología, 11, 54.
 FRIALOV, S. I., 76, 78, 79, 81, 82, 83, 84.
 FITZGERALD, E., 230.
 flexibilidad del texto, 265.
 flexión gramatical, 49, 50, 51, 58, 208, 218, 222, 223.
 foco, 30, 31, 32.
 folklore, 134, 135, 138.
 fondo léxico básico, 289.
 fonema, 2, 8, 10, 13, 15, 17, 18, 23, 27, 37, 39, 56, 59, 60, 61, 83, 115, 117, 246, 255, 256, 257, 259, 260, 261, 262, 263, 264, 265, 270, 271, 272, 273, 274, 280, 289, 291, 295, 325.
 fonética, 2, 11, 12, 122.
 fonética acústica, 7.
 fonética articuladora, 7.
 fonética instrumental, 7.
 fonético, 11.

 fonología, 7, 8, 9, 10, 13, 17, 21, 27, 44, 55, 67, 69, 114, 115, 117, 214, 256, 271, 275, 312, 326.
 forma lingüística, 12, 13, 14, 15.
 forma mínima libre, 22.
 formalización, 12, 54, 55, 98.
 formalización lingüística, 206.
 formante acústico, 334, 335, 336, 337, 339, 340.
 FOTINO, S., 135.
 Fourier, análisis, v. análisis de Fourier.
 FOURIER, J. B. J., 332.
 francés, 4, 5, 6, 40, 88, 133, 152, 226, 234, 245, 259, 260, 285, 287, 291, 294, 295.
 francés antiguo, 6.
 frase, 81, 84, 85, 86, 87, 89, 91, 92, 93, 95, 96, 97, 101, 102, 103, 108, 120, 122, 125, 126, 136, 233, 270, 271, 342.
 frase correctamente construida, 51.
 frase cuasiproyectiva, 103, 104.
 frase de rango n , 101, 102.
 frase marcada, 84, 89; 101, 102.
 frase nula, 95.
 frase proyectiva, 103, 104.
 frase terminal, 95, 97.
 frase unilacunar, 85.
 frase unilacunar marcada, 85.
 frases atómicas, 108.
 frases equivalentes, 85.
 frecuencia, 62, 239, 240, 241, 242, 243, 245, 246, 253.
 frecuencia acústica, 202, 203, 205, 331, 333, 336, 339, 340.
 frecuencia de aparición, 49, 225, 280, 281, 282, 287, 290, 291, 294, 298, 299, 301, 313, 314.
 frecuencia relativa, 65, 66, 67, 119.
 FREUDENTHAL, H., 123, 124, 125, 128, 194, 206.
 FROGER, Dom J., 71.
 FRUMKINA, R. M., 83, 260.
 FRUCKS, W., 290.
 función Bessel, 190.
 función booleana, 61, 83.
 función de Estoup, 240.
 función de Estoup, Zipf y Mandelbrot, 242, 243.
 función de Mandelbrot, 240.
 función de subordinación, 117.
 función de Zipf, 240.
 función lineal, 242.

 función periódica, 331.
 función recursiva, 81.
 función de Riemann, 240.

 GAIFMAN, C., 83.
 GAMOW, G., 255.
 GARCÍA HOZ, V., 235.
 GAY-LUSSAC, J. L., 278.
 generalización, 128.
 género gramatical, 140, 214, 221, 272.
 género literario, 129.
 geografía lingüística, 7.
 geometría, 43, 51.
 geometría analítica, 329.
 GIBBS, J. W., 279.
 GILLIÉRON, J., 7.
 GINSBURG, S., 70.
 GLADKI, A. V., 69, 71, 84, 101.
 GLEASON, H. A., 22, 34, 87.
 glosemática, 8, 9, 16, 89, 113.
 glotocronología, 113.
 GOETHE, J. W., 68.
 GOLOPENTIA, S., 11, 38, 39, 245.
 GOODMAN, N., 83.
 GORALČIKOVÁ, A., 70.
 GORKI, M., 236.
 GORUN, I., 135.
 GOUGHENHEIM, G., 208.
 grado de dependencia, 102.
 GRAESER, E., 282.
 grafos, teoría de los, v. teoría de los grafos.
 gramática, 2, 4, 10, 11, 44, 47, 50, 52, 67, 69, 82, 87, 91, 95, 99, 100, 103, 114, 115, 117, 118, 208, 209, 212, 213, 224, 233, 255, 256, 273, 275, 280, 312, 313, 326.
 gramática analítica, 274.
 gramática categorial, 83.
 gramática con un número finito de estados, 10, 81, 83, 91, 93, 94, 95, 113, 114.
 gramática cuasiproyectiva, 104.
 gramática de constituyentes inmediatos, 35, 82, 83, 93, 94, 108.
 gramática de Port Royal, 4.
 gramática dependiente de contexto, 96.
 gramática formal, 69.
 gramática generativa, 8, 10, 25, 35, 38, 39, 91, 93, 231.
 gramática independiente de contexto, 95, 96, 97.

- gramática matricial, 96, 97.
 gramática natural, 94.
 gramática normativa, 2.
 gramática proyectiva, 83, 103, 104.
 gramática secuencial, 83.
 gramática sintética, 274.
 gramática tradicional, 24, 35, 37, 50.
 gramática transformacional, 10, 11, 37, 82, 83, 91, 93.
- HADAMARD, J., 53.
 HADLICH, R. L., 11.
 HAJIČOVÁ, E., 70.
 HALLE, M., 260, 312.
 HAMMING, R. W., 56.
 HARRIS, J. W., 18.
 HARRIS, Z. S., 9, 18, 24, 78, 87, 89.
 HARTLEY, R. V., 250.
 HARWOOD, F. W., 78.
 HASDEU, B. P., 244, 282, 283, 284, 286.
 HAUGEN, E., 11.
 HAYS, D. G., 71, 72.
 hebreo, 4, 294.
 HEMINGWAY, E., 288.
 HERÁCLITO, 1.
 HERDAN, G., 65, 66, 70, 79, 83.
 HEYTING, A., 144.
 HIGHLEYMANN, W. H., 342.
 HILBERT, D., 148.
 himnos védicos, 1.
 hindí, 294, 295.
 hipótesis de Yngve, 104.
 HIRSCHBERG, L., 83, 89.
 historia, 123.
 historicismo lingüístico, 12.
 HJELMSLEV, L., 9, 10, 12, 13, 16, 21, 56, 78, 114, 115, 117, 258, 313.
 HOCKETT, CH. E., 9.
 homonimia, 38, 51, 65, 86, 111, 150, 151, 152, 153, 154, 164, 166, 209, 213, 217, 221, 311, 327.
 homonimia casual, 164.
 homonimia gramatical, 221.
 homonimia léxica, 87.
 homonimia léxico-gramatical, 152, 153, 154, 155, 162, 163.
 homonimia morfológica, 86, 149, 152, 153, 162, 163, 164, 165, 166, 214, 325, 326.
 homonimia preposicional, 221.
- HORECKY, J., 72.
 HUMBOLDT, W. von, 6.
 HUPCROFT, J. E., 69.
 HUWĀRIZM, AL, 143.
- (i, j)-dependencia, 258.
 IAGLOM, A. M., 256, 257, 260.
 IAGLOM, I. M., 256, 257, 260.
 IANOV, I., 162, 163.
 idealismo lingüístico, 141.
 IHM, P., 81, 83, 89, 103.
 implicación, 123, 197.
 implicación bilateral, 197.
 indeterminación informacional, 249.
 indoeuropeo, 286, 287.
 influencia a distancia, 256-61.
 información, 57, 172, 179, 233-275, 319.
 ingeniería, 44, 55, 80, 144, 226.
 inglés, 4, 45, 48, 105, 106, 107, 109, 119, 133, 139, 150, 153, 154, 155, 162, 163, 164, 186, 187, 207, 210, 212, 213, 215, 216, 219, 221, 222, 226, 230, 250, 257, 258, 259, 260, 261, 263, 264, 272, 274, 284, 286, 294, 295, 312, 325, 342.
 instrucciones lógicas, 167-79.
 interdependencia lingüística, 16.
 interjección, 3.
 interrogación, 124, 128.
 invariables, 127.
 IOFE, 236.
 isomorfismo, 12, 16, 114, 115, 116, 117.
 isomorfismo lingüístico, 114, 115.
 isomorfismo matemático, 116.
 italiano, 4, 5, 40, 261, 274, 291, 295.
 IVANOV, V. V., 79, 84, 90, 167.
- JAKOBSON, R., 9, 53, 55, 56, 260, 312.
 jeroglífico, 118, 119, 120, 121, 122, 123.
 JINCHIN, A. I., 62, 324.
 JOSSELSOHN, H. H., 285.
 juegos de azar, 278.
 JUILLAND, A., 66, 244, 253.
 JUILLAND, I., 66, 244, 253.
- KALMAR, L., 72.
 kannada, 294, 295.
- KARAMZIN, N. M., 236.
 KATZ, J. J., 10, 38, 39.
 KHAYYAM, O., 230.
 KIEFER, F., 72.
 KIMBALL, J. P., 69.
 KING, R. D., 11, 13.
 KLEE, P., 228.
 KLEENE, S., 143.
 KOCK, J., 235, 253.
 KOLMOGOROV, A. N., 76, 79, 84, 85, 143, 245, 256, 264, 265.
 KORTLANDT, F. H. H., 69.
 KRAMSKY, J., 21.
 KRENN, H., 11.
 KRÖNIG, 278.
 KULAGUINA, O. S., 58, 78, 79, 81, 82, 83, 84, 87, 101, 109.
 KALUZHNIN, L. A., 83.
 KURYLOWICZ, J., 6, 78, 114, 115, 116.
 KUZNETSOV, A. F., 60, 83.
- LAGRANGE, J. L., 282.
 LAMBEK, J., 78, 83, 106, 108.
 latin, 3, 4, 5, 6, 12, 16, 21, 234, 282.
 latin tardío, 274.
 latin vulgar, 2.
 LECERF, Y., 81, 83, 89, 103.
 LEES, R. B., 113.
 lengua adecuada, 88.
 lengua conforme, 88.
 lengua con número finito de estados, 91, 92, 94.
 lengua cósmica, 123.
 lengua cuasiproyectiva, 104.
 lengua dependiente de contexto, 97.
 lengua-imagen, 208, 209, 214.
 lengua independiente de contexto, 97.
 lengua lógica, 79, 80, 82, 90.
 lengua natural, 43, 44, 45, 69, 82, 85, 90, 91, 92, 93, 94, 104, 105, 125, 128, 131, 133, 209, 269.
 lengua-objeto, 209, 210.
 lengua-original, 208, 209.
 lengua proyectiva, 103, 104.
 lengua románica, 213, 274.
 lengua semítica, 271.
 lengua terminal, 92.
 lenguaje Algol, 186, 189, 190, 191.
 lenguaje artístico, 138, 139.
 lenguaje científico, 287-9, 300-8.

- lenguaje con un número finito de estados, 78.
 lenguaje cósmico, 193-206.
 lenguaje literario, 287-9, 300.
 lenguaje matricial, 97.
 lenguaje poético, 129, 130, 138, 287-9.
 lenguas artificiales, 43, 44, 128.
 lenguas de información, 58, 93.
 lenguas eslavas, 4.
 lenguas formales, 42, 226.
 lenguas formalizadas, 43, 44.
 lenguas informacionales, 169, 191.
 LENIN, V. I., 57, 236.
 LENTIN, A., 69.
 LEPSCHY, G. C., 8, 11, 13, 38.
 LEROY, M., 1.
 LEŚNIEWSKI, S., 69.
 LESSING, G. E., 296.
 LÉVY, F., 134.
 LEWIS, K., 71.
 léxico, 11, 118, 119, 120, 122, 209.
 lexicología, 2.
 ley natural, 206.
 leyes cuantitativas, 233-34.
 ley de Avogadro, 278.
 ley de Boyle-Mariotte, 278.
 ley de Estoup-Zipf, 237, 241.
 ley de Fucks, 291.
 ley de Gay-Lussac, 278.
 ley de los eventos raros, 286, 287.
 ley de Mandelbrot, 281, 282, 283.
 ley de Maxwell, 279.
 ley de número-frecuencia, 238, 239, 240.
 ley de Poisson, 287.
 ley de rango-frecuencia, 238, 239, 240, 241, 243.
 ley de Zipf, 263, 264, 280, 281, 282, 283, 286, 291.
 leyes estadísticas, 56, 277.
 LIAPUNOV, A. A., 84, 156, 163, 165.
 LICKLIDER, J. C., 257, 263.
 LIEB, H. H., 72.
 Lincos, 193-206.
 lingüística aplicada, 41, 44, 53, 77, 79, 80, 81, 82, 94, 103, 104, 111.
 lingüística comparativa, 3.
 lingüística computacional, 71.
 lingüística descriptiva, 3, 7, 8, 12, 87, 89.
 lingüística estadística, 63, 64, 65, 66, 67, 70.
 lingüística estructural, 1, 7, 10, 11, 16, 17, 18, 21, 22, 57, 59, 60, 81, 89, 104, 110, 117, 164, 312.
 lingüística genética, 141.
 lingüística histórica, 3, 7, 12.
 lingüística histórico-comparativa, 4, 6.
 lingüística matemática, *passim*.
 lingüística tradicional, 11, 20, 60, 104, 122.
 literatura, 47, 129, 141, 227, 229.
 LIVIO, TITO, 287.
 LORENTE, A., 12.
 logaritmo, 148, 250, 252, 268, 282.
 lógica, 4, 11, 42, 44, 55, 67, 68, 123, 201, 207, 255, 273.
 lógica aristotélica, 59.
 lógica formal, 69.
 lógica matemática, 43, 52, 53, 59, 78, 79, 80, 82, 83, 93, 143, 164, 166, 196, 300, 322.
 lógica simbólica, 106, 123.
 LOMONOSOV, M. V., 4.
 longitud, 106, 206, 246.
 longitud de la palabra, v. extensión de las palabras.
 LUTZEER, P. R., 70.
 LYNCH, I., 83, 89.
 LYONS, J., 13, 39.
 MACREA, D., 244.
 macroestructura semántica, 38, 39.
 macro-lingüística, 65.
 Mahābharata, 1.
 MALAKOVSKI, V., 79.
 malayo, 294, 295.
 MALMBERG, B., 4, 8, 11, 21.
 MANDELBROT, B., 61, 64, 79, 83, 240, 241, 245, 281, 282, 322.
 máquinas analógicas, 170.
 máquinas de cálculo, 48, 165, 166.
 máquina de Turing, 81, 93, 94.
 máquina digital, v. ordenador digital.
 máquina electrónica de cálculo, 80, 104, 118, 120, 224.
 máquinas generativas, 137.
 máquina informacional, 170, 275.
 marathi, 294, 295.
 MARCUS, S., 70, 72, 78, 84, 94, 135, 164.
 MARIOTTE, E., 278.
 MARKOV, A. A. (hijo), 62, 94, 122, 143, 144, 158, 163, 322.
 MARKOV, A. A. (padre), 158.
 MARTINET, A., 9, 10, 13, 18, 24, 273.
 masa, 106, 205, 206, 278.
 matemática, 12.
 matematización, 52, 54.
 materialismo dialéctico, 55, 58, 77, 81, 110.
 MATHESUS, V., 9.
 MATORE, G., 64.
 matriz de orden m , 96, 97.
 MAXWELL, J. C. 278, 279.
 maya, 117, 118, 119, 120, 121, 122.
 mecánica, 54, 123, 301.
 mecánica estadística, 279.
 media ponderada, 253, 254.
 medicina, 68, 75.
 MEILLET, A., 6.
 MELCHUK, I. A., 57, 71, 83, 167, 260.
 memoria de un ordenador, 48, 81, 170-9, 209-224.
 MENÉZDEZ PIDAL, R., 7.
 mereología, 69.
 metáfora, 109.
 metalengua, v. metalenguaje.
 metalenguaje, 52, 129, 130, 194, 195, 199, 201, 204, 209.
 metatexto, 203, 204.
 método de correlación, 298-300.
 método de las funciones arbitrarias, 62.
 método estadístico, 122, 244, 245.
 método "rebusista", 122.
 métrica, 131, 132.
 MEYER-EPPLER, W., 271.
 MEYER-LÜBKE, W., 6.
 microestructura semántica, 39.
 microgramática, 209.
 MIHOC, G., 62.
 MIKAELAN, G., 25.
 MILLER, G. A., 78, 83, 94, 274.
 MITCHELL, R. P., 83.
 mitología, 121.
 modelación, 45, 75, 76.
 modelación algorítmica, 93, 117, 143.
 modelación analógica, 76.
 modelación de enseñanza, 76.
 modelación gramatical, 28.
 modelación ideal, 76.
 modelación lógica, 108, 123.
 modelación matemática, 76, 77,

- 78, 79, 80, 81, 82, 84, 89, 90, 91, 92, 93, 98, 101, 104, 108, 109, 110, 113, 114, 116, 117, 118, 135, 139-40, 141, 164, 166, 256.
- modelación material, 76.
- modelo, 26, 27, 30, 35.
- modelo algorítmico, 84.
- modelo analítico, 82, 84.
- modelo configuracional, 83.
- modelo constructivo, 84.
- modelo deductivo, 77.
- modelo descriptivo, 84.
- modelo de las partes del discurso, 109.
- modelo de los constituyentes inmediatos, 89, 92.
- modelo de Ingve, 81.
- modelo generativo, 38, 83, 90, 91, 93.
- modelo lógico, 57, 58.
- modelo matemático, 69, 82.
- modelo no-algorítmico, 84.
- modelo paradigmático, 83, 84.
- modelo probabilista, 81, 111-13.
- modelo proyectivo, 81.
- modelo sintagmático, 79, 83.
- modelo sintético, 83.
- modelo transformacional, 89, 90, 93.
- modo gramatical, 214, 218, 237.
- MOISIL, GR. C., 151, 207, 209, 223.
- MOLOSHNAIA, T. N., 57, 79, 153.
- MOREAU, R., 285.
- MORF, A., 322.
- morfema, 8, 10, 15, 17, 18, 19, 20, 21, 23, 25, 29, 30, 32, 35, 59, 84, 246, 255, 256, 258, 259, 261, 262, 265, 272, 273, 274, 275.
- morfología, 2, 9, 11, 12, 18, 26, 35, 50, 51, 67, 81, 209, 233, 255, 274, 275.
- MOSCOVICH, V. A., 70.
- MOUNIN, G., 1, 39.
- movimiento, 205.
- movimiento kepleriano, 206.
- MOZART, W. A., 226.
- MÜLLER, CH., 70, 235.
- MÜLLNER, K., 11.
- MUJIN, A. M., 10, 24.
- música, 226-8, 231.
- neogramática, 11.
- neopositivismo, 58, 110.
- neutralización, 59.
- NICOLAU, E., 169, 208, 229-31, 245, 256, 260, 275, 298, 314.
- NICOLESCU, G. C., 265.
- NIDA, E. A., 9, 10, 24, 28, 34.
- NIKOLAEVA, T., 57.
- NIKONOV, V. A., 235, 255.
- ΝΙΤΧ, C., 169.
- NIVETTE, J., 94.
- nombre, 2, 6, 51, 89, 106, 125, 126, 127, 133, 149, 150, 151, 152, 163, 164, 209, 210, 213, 214, 215, 216, 217, 218, 220, 221, 236, 237, 253, 255, 325, 326.
- nombre común, 2.
- nombre propio, 2.
- NOVAC, L., 256.
- NOWAKOWSKA, M., 73.
- núcleo del sintagma, 100.
- numeral, 3.
- número cardinal, 200.
- número complejo, 200.
- número entero, 198, 199.
- número gramatical, 214, 221.
- número natural, 97, 101, 102, 159, 160, 161, 199, 237, 238, 239, 321.
- número negativo, 198.
- número positivo, 199.
- número primo, 199.
- número racional, 200.
- OBLIGADO, R., 296.
- ONICESCU, O., 62, 132.
- operador, 145, 146, 147, 150, 151, 152, 154, 156, 157, 161, 162, 163, 183, 185, 186.
- operador de ciclización, 189.
- operadores lógicos, 187.
- operador O, 154-157.
- oposición, 56, 61.
- oposición fonológica, 214.
- oposición lingüística, 18.
- oración, 15, 22, 23, 24, 26, 27, 35, 37, 38, 47, 93, 114, 273, 275.
- oraciones nucleares, 37, 38.
- orden, 101, 104.
- orden de las palabras, 209, 210, 213, 216, 219, 221, 224, 233.
- orden estructural, 99.
- orden linear, 98, 99.
- ordenador digital, 72, 169-172, 179, 208, 209, 342.
- ordenador electrónico, 41, 42, 71, 117, 134, 169, 170, 207, 208, 210, 224, 226-31.
- ORTIZ, A., 72.
- ortografía, 333.
- OSTHOFF, H., 6.
- PADUCHEVA, E. V., 58, 83, 260, 313.
- PÄNINI, I.
- PANOV, D. I., 49, 154, 155, 208.
- par predicativo, 114.
- paradigma, 86.
- paradigmática, 15, 16, 18, 57, 131.
- PARKER-RHODES, A. F., 82, 83.
- partes de la oración, v. partes del discurso.
- partes del discurso, 2, 15, 22, 23, 24, 35, 38, 39, 53, 58, 78, 88, 109, 114, 139, 140, 255.
- partición derivada, 87.
- partición en vecindades, 87.
- participio, 3.
- PASCAL, B., 278.
- PASK, G., 229.
- PAUL, H., 6.
- PAVLOV, J. P., 156, 236.
- PELEVINA, N. F., 111.
- periodos, 15.
- PERLES, M., 83.
- persona gramatical, 218, 219, 222.
- PETRESCU, C., 133, 134.
- PIKE, K. L., 9, 24, 89.
- PIOTROVSKI, R. G., 71, 226, 256, 271.
- plano del contenido, v. contenido.
- plano de la expresión, v. expresión.
- PLATÓN, I.
- plerema, 258.
- "plotter", 228.
- POE, E. A., 62, 119.
- poesía, 129, 130, 229-30, 264-65.
- poética matemática, 129, 130, 131, 135.
- POISSON, S. D., 286.
- polisemia, 39, 49, 225.
- Port Royal, v. gramática de Port Royal.
- POST, E., 143.
- POSTAL, P., 10, 38, 39.
- POTTIER, B., 39, 40.
- POYEN, J., 169.
- POYENNE, J., 169.

- Praga, escuela de, v. escuela de Praga.
- PRATT, A. W., 71.
- preposición, 3, 155, 163, 233, 324.
- probabilidad, 233-275.
- probabilidad de aparición, 63, 132, 253, 254, 256, 257, 258, 259, 261, 263, 264, 266, 270, 275, 282, 295, 313, 323.
- proceso de formalización, 164.
- problema del "leader", 135.
- proceso de Markov, 158.
- proceso de modelación, 75.
- procesos de ordenación, 144, 155.
- procesos recursivos, 138.
- profundidad de una estructura regresiva, 105.
- programación, 47, 166, 170, 171, 172-79, 179-84, 185, 186, 188, 189, 190.
- programación gramatical, 47.
- pronombre, 3, 125, 153, 154, 163, 221, 222, 255, 325.
- propiedad de proyectividad, 134.
- propiedad de Tesnière, 133, 134.
- propiedades algorítmicas, 162.
- propiedades de asociatividad, 108.
- proposición, 29, 31, 36, 98, 99, 100, 104, 105, 107, 109, 115, 116, 117, 123, 124, 133, 134, 210, 216, 223, 270, 271, 272.
- proposiciones no proyectivas, 133.
- provenzal, 6.
- proyectividad sintáctica, 132, 133.
- psicología, 11, 54, 55, 67, 144.
- PUSHKIN, A. S., 62, 79, 129, 236, 296.
- QUESADA, D., 94.
- química, 80, 296.
- química física, 55.
- QUINE, W. V. O., 83, 127, 200.
- RABIN, M. O., 78, 94.
- radio, 80.
- radioquímica, 55.
- RADOI, I., 135.
- RĀMĀYANA, 1.
- rango, 237, 238, 239, 240, 241, 242, 243, 263, 281, 291, 299.
- rasgo acústico, 312.
- rasgos fonológicos, 61.
- rección, 258.
- reconocimiento automático, 329-43.
- recursividad, 92, 93, 108.
- redacción formalizada, 120.
- redundancia, 262, 265-274, 275.
- redundancia de orden n , 266.
- redundancia morfológica, 272.
- redundancia relativa, 300.
- reflejo condicionado, 156, 157, 182, 183, 186.
- regla con autoinclusión, 95.
- regla combinatoria, 269.
- regla de generación, 36.
- regla dependiente de contexto, 95.
- regla de transformación, 37.
- regla generativa, 35, 37.
- regla gramatical, 93, 206, 208, 216, 271.
- regla independiente de contexto, 95.
- regla morfofonémica, 37.
- regla terminal, 95.
- regla transformativa, 38.
- regulador de Watt, 165.
- relación de dependencia, 102.
- relación de equivalencia, 87.
- relación de orden, 86.
- relación de subordinación, 93, 98, 99, 100, 101, 102, 103, 114, 116.
- relación gramatical, 33, 38.
- REMBRANDT, H. VAN R., 229.
- resultante de la configuración, 101.
- retículos, 88.
- retórica, 1.
- REVZIN, I. I., 58, 60, 71, 76, 78, 82, 83, 84, 86, 88, 89, 101, 109.
- RICHARD, P., 134.
- RIEMANN, G. F. B., 240.
- RIMBAUD, A., 296.
- ROBERTS, A. H., 71.
- ROBINS, R. H., 1, 38.
- ROGERIC-ALEXANDRESCU, A., 245, 256, 260, 263, 275, 299.
- RODORÉDA, M., 288.
- ROSETTI, A., 244.
- ROZENTVEIG, V. I., 82.
- ruido informacional, 248, 266-68, 270, 274, 295, 316.
- rumano, 5, 13, 16, 18, 20, 40, 44, 47, 48, 51, 63, 66, 88, 89, 133, 134-37, 139, 149, 150, 163, 164, 165, 166, 207, 208, 209, 210, 213, 214, 215, 216, 219, 222, 230, 233, 235, 244, 253, 255, 256, 258, 259, 260, 261, 269, 270, 271, 272, 273, 274, 275, 282, 283, 284, 286, 287, 290, 291, 295, 296, 298, 313, 314, 325, 326, 336, 340.
- ruso, 16, 45, 48, 88, 89, 111, 112, 120, 133, 150, 207, 226, 255, 258, 259, 260, 285-6, 294, 295, 324, 325.
- RUWET, N., 10, 38.
- SADOVEANU, M., 233, 287, 288, 291, 296.
- SĂHLEANU, V., 76.
- SALA, C., 256, 260, 275, 298.
- SALOMAA, A., 69.
- SÁNCHEZ DE ZAVALA, V., 12.
- sánscrito, 1, 4, 6.
- SAPIR, E., 21.
- satélite del sintagma, 100.
- SAUSSURE, F. DE, 5, 6, 7, 8, 9, 13, 39, 53, 55, 56, 310, 312.
- SCOTT, D., 78, 94.
- SCHLEICHER, A., 6.
- SCHNELLE, H., 72.
- SCHÜTZENBERGER, M. P., 83.
- SEBEOK, TH. A., 73.
- secuencia terminal, 37.
- segmentos narrativos, 136.
- sema, 39.
- semántica, 10, 12, 18, 19, 20, 21, 28, 38, 58, 70, 90, 111, 125, 126, 131, 132, 228, 231, 246, 255, 262, 270, 310.
- semántica estructural, 38, 39.
- semántica lógica, 43, 164.
- semas, 40.
- serie armónica, 238, 240.
- series, 180.
- SERRANO, S., 152.
- SERVIEN, P., 129.
- SGALL, P., 70.
- SHAKESPEARE, W., 296.
- SHAMIR, E., 83, 94, 114.
- SHANNON, C. E., 63, 250, 261, 262, 263, 264, 277, 279.
- SHAUMIAN, S. K., 60, 68, 69, 72, 79, 313.
- SHESTOPAL, G. A., 156, 163, 165.
- significado, 13, 15, 18, 19, 20, 22, 28, 45, 49, 52, 81, 87, 98, 115, 119, 120, 122, 127, 128, 130, 151, 167, 193, 195, 210, 212, 234, 243, 266, 268, 271.
- significado gramatical, 212.
- significante, 13.

- signo lingüístico, arbitrariedad del, v. arbitrariedad del signo lingüístico.
- signo lingüístico, 9, 10, 13.
- silaba, 2, 114, 115, 116, 117, 121, 289.
- simbolo no-terminal, 95.
- simbolos auxiliares, 95.
- simbolos terminales, 95.
- sincretismo gramatical, 325.
- sincronía, 12, 13, 140.
- SINGH, T., 257.
- sinonimia, 38, 111, 112, 233, 311.
- sintagma, 10, 22, 35, 59, 100, 102, 131, 256, 272.
- sintagma mínimo, 21.
- sintagmática, 15, 16, 18, 57, 221.
- sintaxis, 2, 9, 10, 11, 12, 22, 24, 35, 49, 50, 67, 82, 132, 209, 213, 219, 223, 233, 237, 271, 275.
- sintaxis estructural, 24.
- sintaxis poética, 132.
- sintaxis tradicional, 23.
- sisntesis acústica, 339-42.
- sistema Baudot, 313.
- sistema binario, 48.
- sistema de puntuación, 196.
- sistema formal, 77.
- sistema gramatical, 272.
- sistema lingüístico, 8, 113.
- sistema Morse, v. alfabeto Morse.
- SLAMA-CAZACU, T., 263.
- SLUTSKER, G. S., 215.
- SMIRNOV-TROIANSKI, P., 45, 207.
- SOBOLEV, S. L., 117.
- SOURIAU, É., 129.
- STATI, S., 22, 25, 272.
- STEINFELDT, 285.
- STEINHAUS, H., 68.
- STROE, A., 226.
- subordinación, 22, 24, 33-34, 102.
- subordinación de rango n , 100.
- subprogramación, 180, 181, 191, 212-13.
- sucesión terminal, 92.
- sueco, 294.
- sustancia lingüística, 14, 15, 57, 58.
- sustantivo, v. nombre.
- sustitución, 28.
- ŞUTEU, P., 282.
- ŞUTEU, V., 66, 245, 253, 283, 337.
- SWADESH, M., 113.
- tablas de verdad, 187.
- TAGLIAVINI, C., 1, 4, 261.
- tamil, 294, 295.
- tautología, 197.
- taxonomía, 343.
- telefonía, 80.
- telego, 294, 295.
- telegrafía, 46, 80, 313.
- televisión, 227-8.
- temperatura de información, 241.
- teorema de Fourier, 332.
- teorema de Pitágoras, 148.
- teoría algebraica, 91.
- teoría cinética, 278.
- teoría cinética de los gases, 279.
- teoría convencionalista, 3.
- teoría de Fucks, 290.
- teoría de códigos, 309.
- teoría de la automoción, 53.
- teoría de la comunicación, 53, 277.
- teoría de la información, 43, 63, 71, 81, 246, 247, 248, 249, 250, 256, 257, 258, 309, 323, 324.
- teoría de la relatividad, 206.
- teoría de la representación de los grupos, 322.
- teoría de las catástrofes, 72.
- teoría de las funciones recursivas, 53.
- teoría de las gramáticas generativas, 135.
- teoría de las oscilaciones, 205.
- teoría de las probabilidades, 62, 158.
- teoría de los algoritmos, 81, 322.
- teoría de los árboles, 131.
- teoría de los autómatas, 82, 91, 114.
- teoría de los autómatas finitos, 78, 81, 114.
- teoría de los códigos, 43, 56, 71, 322, 325.
- teoría de los conjuntos, 42, 43, 59, 61, 67, 69, 78, 79, 81, 83, 84, 114, 164.
- teoría de los grafos, 69, 71, 78, 83, 131.
- teoría de los grupos finitos, 322.
- teoría de los modelos, 10.
- teoría de los números, 82.
- teoría de los retículos, 83.
- teoría de los semigrupos, 82, 322.
- teoría generativa, 70, 136.
- teoría leninista de la gnoseología, 57, 117.
- término dependiente, 100.
- término regente, 100.
- termodinámica, 114, 255, 279.
- TESNIÈRE, L., 10, 18, 24, 89.
- THOM, R., 72.
- tiempo, 106, 201, 202, 203, 205.
- tiempo gramatical, 214, 218, 237.
- TIMIRIAZEV, K. A., 236.
- tipología generativa, 137.
- tipo primitivo, 108.
- tipos, 88.
- tipos primarios, 108.
- tipos sintácticos, 106.
- TOMA, P., 245.
- topología, 53, 59, 114, 306.
- TOPOROV, V. N., 68.
- traducción, 164, 275.
- traducción automática, 9, 44, 45, 46, 47, 49, 52, 57, 58, 66, 79, 80, 81, 93, 122, 151, 167, 225, 277, 313.
- traducción electrónica, 44, 65.
- traducción mecánica, v. traducción automática.
- traducción "palabra por palabra", 99, 123, 223.
- TRAGER, G. L., 10, 22, 34.
- TRAJTENBROT, B. A., 147.
- transferencia condicionada, 176, 178.
- transformación de Fourier, 334.
- TRIER, J., 38, 39.
- trigonometría, 50.
- TRNKA, B., 9.
- TRUBETSKOI, N. S., 9, 55, 56, 61.
- TURING, A., 143.
- UHLENBECK, E. M., 11, 39.
- ULDALL, H., 9.
- ULLMAN, J. D., 69.
- USPENSKI, V. A., 83, 84, 88, 109.
- VACHEK, J., 9, 11.
- VAIDA, D., 169, 176.
- VAKAR, N. P., 285-6.
- valores booleanos, 183.
- VARGA, D., 324.
- variable, 124, 125, 126, 127, 128, 196.
- variable booleana, 336.
- variable libre, 124.
- variación fonológica, 17.
- variante fonética, 17.
- VARRÓN, 3.
- VASILIU, E., 11, 38, 39, 70, 83.
- vecindad lingüística, 87, 88, 89.

- velocidad, 206.
- verbo, 2, 3, 48, 133, 152, 153, 163, 209, 210, 213, 214, 215, 216, 217, 220 221, 222, 237, 253, 255.
- verdades estadísticas, 235.
- verificaciones de paridad, 318-20.
- VIANU, T., 109, 245.
- VIRVILLE, M., 134.
- vocabulario, 38, 39, 49, 50, 85, 91, 95, 97, 101, 102, 206, 216, 227-30, 237, 238, 240, 241, 245, 263, 265, 284, 296.
- "vocoder", 339.
- WALD, L., 1, 4, 11.
- WALL, R., 69.
- WELLS, R. S., 25, 30, 89.
- WHITFIELD, F. J., 141.
- YNGVE, V. H., 83, 105, 106, 208.
- YULE, G. U., 65, 66.
- ZARECHNAK, 208.
- ZIERER, E., 72, 78, 83.
- ZINDER, L. R., 79, 273.
- ZINOVIEV, A. A., 76, 78.
- ZIFF, G. K., 49, 64, 237, 241, 243, 281, 286.
- ZVEGUNTSEV, V. A., 89, 90.

Índice general

PREFACIO	V
NOTA A LA EDICIÓN ESPAÑOLA	VI
PRÓLOGO A LA VERSIÓN ESPAÑOLA	VII
TABLA DE SÍMBOLOS	IX
1. <i>La lingüística estructural</i>	1
1.1. Orígenes de la lingüística	1
1.2. Constitución de la lingüística científica	4
1.3. Breve historia de la lingüística estructural	7
1.4. La crítica a la lingüística tradicional	11
1.5. Algunas dicotomías	13
1.6. Tipos de relaciones	16
1.7. La fonética estructural (fonología)	17
1.8. La morfología estructural	18
1.9. La palabra y las unidades superiores a la palabra	21
1.10. La sintaxis tradicional	23
1.11. La sintaxis estructural	24
1.12. Técnica de segmentación en constituyentes inmediatos	28
1.13. Las relaciones gramaticales	33
1.14. Gramáticas generativas	35
1.15. Semántica estructural	38
2. <i>Sinopsis general de la lingüística matemática</i>	41
2.1. Objeto y método	41
2.2. Cómo y cuándo ha nacido la nueva disciplina	41
2.3. El origen matemático de la lingüística matemática	42

2.4.	El origen técnico de la lingüística matemática	43
2.5.	La traducción de una lengua a otra con máquinas electrónicas de cálculo	45
2.6.	El principio regulador de las máquinas electrónicas	45
2.7.	Cómo se buscan las palabras en el diccionario electrónico	46
2.8.	La máquina también "aprende" gramática	47
2.9.	Otras operaciones al alcance de la máquina	48
2.10.	La importancia del conocimiento de las leyes estadísticas del vocabulario	49
2.11.	Dificultades relativas a la introducción de la gramática en la máquina	50
2.12.	¿Qué ayuda puede proporcionar la matemática?	50
2.13.	Antecedentes en el proceso de la matematización de la lingüística	52
2.14.	El sentido del proceso de matematización	54
2.15.	Premisas lingüísticas de la lingüística matemática	55
2.16.	La importancia del análisis lingüístico estructural y sus relaciones con el análisis de la sustancia lingüística	57
2.17.	La lingüística matemática como etapa superior en el desarrollo de la lingüística estructural	59
2.18.	La lingüística estructural no matemática como premisa necesaria para el desarrollo de la lingüística matemática	60
2.19.	La estructura probabilística de la lengua	61
2.20.	La lingüística estadística	63
2.21.	Anexo bibliográfico	69
3.	<i>Modelos matemáticos en la lingüística</i>	75
3.1.	Concepto de modelo y modelación	75
3.2.	Necesidades de la modelación matemática de la lengua	77
3.3.	Modelaciones concomitantes	77
3.4.	Analogía, modelo, teoría	78
3.5.	El modelo omite aspectos esenciales	79
3.6.	La importancia de la modelación matemática para la lingüística aplicada	79
3.7.	Clasificación de los modelos matemáticos de la lengua	82
3.8.	Ejemplos de modelos analíticos en gramática	84
3.9.	Modelación y lingüística estructural	89
3.10.	La dialéctica de la modelación a través del modelo	90
3.11.	Los modelos generativos y su función explicativa	91
3.12.	La jerarquía de Chomsky	94
3.13.	La modelación de las relaciones subordinadas	98

3.14.	Configuraciones	101
3.15.	Relaciones de dependencia	102
3.16.	La relación de proyectividad	103
3.17.	La hipótesis de Yngve	104
3.18.	El cálculo de los tipos sintácticos	106
3.19.	Dificultades que aparecen en el proceso de modelación	108
3.20.	Modelo y metáfora	109
3.21.	La función gnoseológica de la modelación matemática de la lengua	110
3.22.	Función explicativa de los modelos matemáticos	111
3.23.	El carácter específico de las abstracciones matemáticas	113
3.24.	Isomorfismo y modelo	114
3.25.	La modelación algorítmica	117
3.26.	La modelación matemática y el desciframiento de la lengua maya	118
3.27.	El papel de la modelación lógica en la formalización del discurso indirecto	123
3.28.	Los modelos matemáticos en el estudio de la literatura	129
3.29.	Algunos problemas especiales de la modelación matemática en lingüística	139
3.30.	Perspectivas y significación	140
4.	<i>La modelación algorítmica de la lengua</i>	143
4.1.	Introducción	143
4.2.	Algoritmo de la sustracción	144
4.3.	El algoritmo de Euclides	145
4.4.	Otros algoritmos numéricos	147
4.5.	La resolución de la homonimia morfológica	149
4.6.	La resolución de la homonimia léxico-gramatical	152
4.7.	El hallazgo de una correspondencia castellana a una palabra inglesa	154
4.8.	Algunos problemas no lingüísticos	155
4.9.	El concepto de "algoritmo normal"	158
4.10.	Representación de un algoritmo normal con ayuda de operadores y de condiciones lógicas	161
4.11.	Conclusiones sobre el concepto general de algoritmo. Las propiedades algorítmicas. Implicaciones lingüísticas	162
5.	<i>Lenguas informacionales</i>	169
5.1.	Introducción	169

5.2.	Los ordenadores digitales	169
5.3.	El programador del ordenador	172
5.4.	Esquemas lógicos de cálculo	179
5.5.	Algoritmos de resolución	185
5.6.	Automatización de las programaciones	186
6.	<i>El lenguaje cósmico</i>	193
6.1.	Introducción	193
6.2.	Nociones temporales	201
6.3.	Comportamiento	203
6.4.	Espacio, movimiento, masa	205
7.	<i>Traducción y composición automáticas</i>	207
7.1.	Nociones introductorias	207
7.2.	El tratamiento de los datos	207
7.3.	El algoritmo de la traducción automática del inglés al rumano	210
7.4.	Puntos de vista sobre la traducción automática	225
8.	<i>Probabilidad, información y entropía en la lengua</i>	233
8.1.	Las leyes cuantitativas en la lengua	233
8.2.	Precisiones numéricas en algunas cuestiones de vocabulario	234
8.3.	Fenómenos lingüísticos que obedecen a determinadas fórmulas	237
8.4.	Preocupaciones y perspectivas en el estudio estadístico de la lengua rumana	244
8.5.	La necesidad de una teoría científica de la información	245
8.6.	Economía y derroche	247
8.7.	Cómo nos ayuda la “teoría de la información”	248
8.8.	Cómo se mide la información	249
8.9.	Concepto de entropía	253
8.10.	Influencia a distancia	256
8.11.	Determinación experimental de la entropía lingüística	261
8.12.	La entropía de las palabras y su relación con la energía de las letras	263
8.13.	La entropía de los textos poéticos	264
8.14.	La redundancia	265
8.15.	La redundancia en las lenguas naturales	269
8.16.	Ejemplos y aplicaciones	271
8.17.	Observaciones finales	274

9. <i>Otros métodos estadísticos en lingüística</i>	277
9.1. Introducción	277
9.2. Generalidades históricas	277
9.3. Los métodos estadísticos en lingüística	279
9.4. La ley de Zipf	280
9.5. La distribución de las palabras en la lengua rumana y en otras lenguas	282
9.6. La longitud media de las palabras	286
9.7. Teoría de Fucks	290
9.8. Leyes estadísticas relativas a letras y fonemas	291
9.9. Relación entre número de palabras nuevas y número total de palabras	296
9.10. Métodos de correlación	298
9.11. El lenguaje científico	300
10. <i>Lengua y código</i>	309
10.1. Introducción	309
10.2. El proceso de codificación	309
10.3. La noción matemática de código	310
10.4. Los códigos binarios y su empleo en lingüística	311
10.5. Desciframiento de un texto codificado	313
10.6. Distancia entre dos palabras	315
10.7. Códigos que descubren errores	315
10.8. Códigos correctores	316
10.9. El método de las verificaciones de paridad	318
10.10. Verificaciones de paridad para la corrección de errores sim- ples y el descubrimiento de los dobles	318
10.11. Una interpretación geométrica	320
10.12. Códigos, grupos, algoritmos	322
10.13. Código y entropía	322
10.14. Algunas ilustraciones de analogías lingüísticas	324
11. <i>El reconocimiento automático de las formas</i>	329
11.1. Generalidades	329
11.2. Reconocimiento de las vocales. Espectros	331
11.3. Aparatos que reconocen las señales vocálicas	335
11.4. La síntesis de las señales vocálicas	339
11.5. El reconocimiento de los caracteres escritos	342

BIBLIOGRAFÍA E ÍNDICES	345
<i>Bibliografía</i>	347
<i>Índice alfabético</i>	371
<i>Índice general</i>	381

En esta colección:

Elementos y técnica del trabajo científico. D. Romano.
Classificació decimal. Adaptació per a les biblioteques catalanes.
J. Rubió.

El lenguaje. J. Roca Pons
Introducción a la gramática. J. Roca Pons
Introducción en la lingüística teórica. J. Lyons
Introducción en la lingüística matemática.
S. Marcus, E. Nicolau, S. Stati.

Introducción a la historia. J. Reglá
Historia y crítica. P. Salmon
Lecciones de prehistoria. A. Arribas
Manual de arqueología medieval. De la prospección a la historia.
M. de Boüard y M. Riu
Lecciones de historia medieval. M. Riu
Textos comentados de época medieval (siglos V al XII).
M. Riu y colaboradores

Textos, mapas y cronología de historia moderna y contemporánea.
E. Giralt, R. Ortega, J. Rolg
Historia de España. Ubleto, Reglá, Jover, Seco
La Península en la Edad Media. J. L. Martín
Historia de las doctrinas económicas. L. Beltrán

Teoría de la probabilidad. K. Krickeberg
Espais de probabilitat finits. E. Bonet, G. Ferrater
Introducción al espacio de Hilbert. S. K. Berberian
Espacios vectoriales topológicos. H. H. Schaefer

El pH en medicina. W. R. Frisell

obras básicas de formación cultural

En esta colección:

Elementos y técnica del trabajo científico. D. Romano.
Classificació decimal. Adaptació per a les biblioteques catalanes.
J. Rubió.

El lenguaje. J. Roca Pons
Introducción a la gramática. J. Roca Pons
Introducción en la lingüística teórica. J. Lyons
Introducción en la lingüística matemática.
S. Marcus, E. Nicolau, S. Stati.

Introducción a la historia. J. Reglá
Historia y crítica. P. Salmon
Lecciones de prehistoria. A. Arribas
Manual de arqueología medieval. De la prospección a la historia.
M. de Boüard y M. Riu

Lecciones de historia medieval. M. Riu
Textos comentados de época medieval (siglos V al XII).
M. Riu y colaboradores

Textos, mapas y cronología de historia moderna y contemporánea.
E. Giralt, R. Ortega, J. Roig

Historia de España. Ubieto, Reglá, Jover, Seco
La Península en la Edad Media. J. L. Martín
Historia de las doctrinas económicas. L. Beltrán

Teoría de la probabilidad. K. Krickeberg
Espais de probabilitat finits. E. Bonet, G. Ferrater
Introducción al espacio de Hilbert. S. K. Berberian
Espacios vectoriales topológicos. H. H. Schaefer

El pH en medicina. W. R. Frisell

obras básicas de formación cultural