



MPGI UC
MAGISTER EN PROCESAMIENTO Y
GESTIÓN DE LA INFORMACIÓN

Internet y sociedad: comunicación y cultura en la era digital

César Antonio Aguilar
Facultad de Lenguas y Letras
27/08/2020

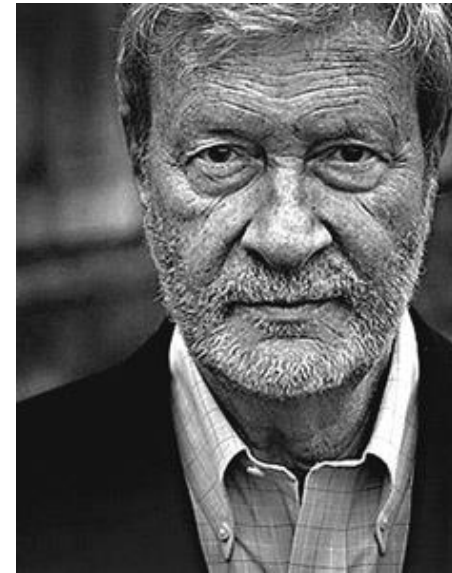
Cesar.Aguilar72@gmail.com

Síntesis de la clase anterior (1)

En la sesión pasada hicimos una breve revisión sobre dos de las corrientes de pensamiento que delinearon tanto el marco teórico como las metodologías que caracterizan a las Humanidades: el **estructuralismo** y el **postmodernismo**.



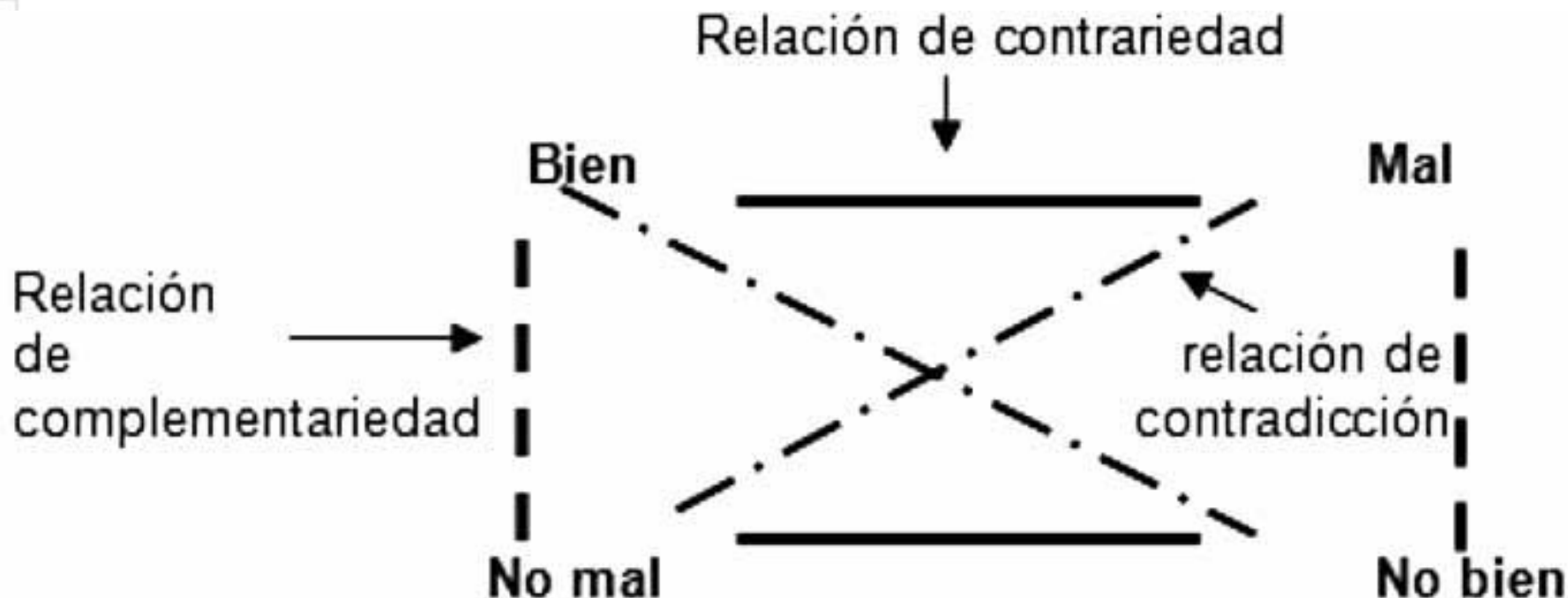
**Claude Lévi-
Strauss**
(1908-2009)



**Gianni
Vattimo**
(1936)

Síntesis de la clase anterior (2)

Estas dos corrientes de pensamiento han logrado un gran impacto en varios campos de conocimiento:



Cuadrado semiótico desarrollado por A. J. Greimas

Síntesis de la clase anterior (3)

Y por supuesto, la literatura y las artes:

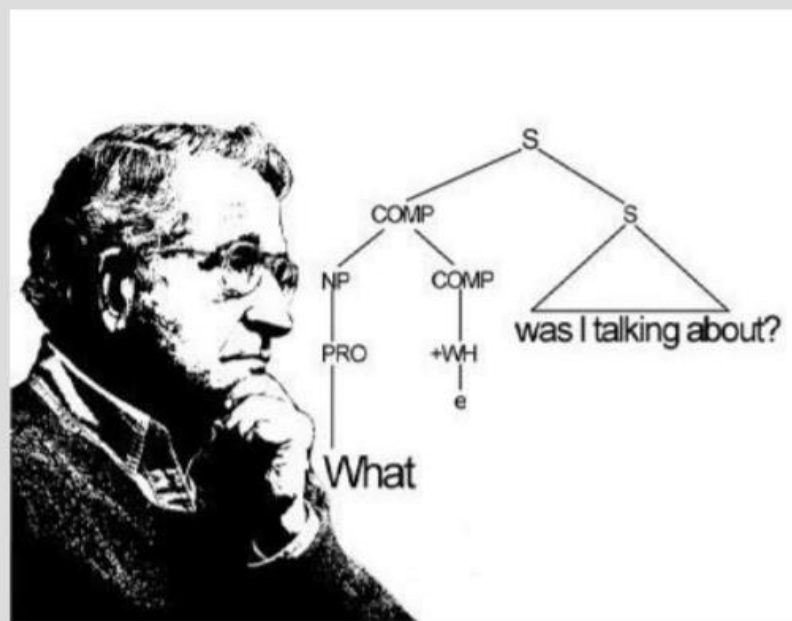
El nombre
Sus sombras
El hombre La hembra
El mazo El gong
La i La o
La torre El ajibe
El índice La hora
El hueso La rosa
El rocío La huesa
El venero La llama
El tizón La noche
El río La ciudad
La quilla El ancla
El hembro La hombra
El hombre
Su cuerpo de nombres
Tu nombre en mi nombre. En tu nombre mi nombre
Uno frente al otro uno contra el otro uno en torno al otro
El uno en el otro
Sin nombres

“Custodia”,
un poema de
Octavio Paz

Del estructuralismo al generativismo (1)

En paralelo a estos movimientos, en el ámbito anglosajón se iban desarrollando y consolidando propuestas alternas que trataban de mantener el estatus de “ciencia dura” para las Humanidades. De hecho, se dieron al menos 3 tendencias:

1. Una cuyo soporte sigue siendo el método hipotético-deductivo, pero poniendo mayor énfasis en **desarrollar hipótesis y pruebas de índole formal, cuyo poder descriptivo y explicativo aspire a ser universal**. El mejor ejemplo de aplicación de este método son las llamadas **gramáticas generativas**, derivadas de los planteamientos de **Chomsky**.



Del estructuralismo al generativismo (2)

Lo que Chomsky y sus colegas se plantearon fue desarrollar una teoría que fuese capaz de explicar —con precisión matemática, de ser posible— cómo funciona el lenguaje humano si lo consideramos como un mecanismo de generación y comprensión de secuencias de sonidos y caracteres, con el cual venimos *programados* de manera natural.

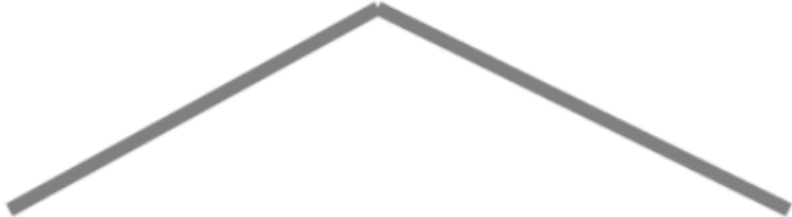


Los modelos generativistas nos invitan a estudiar al lenguaje y la mente humana como si fueran **sistemas computacionales**, con la pequeña gran diferencia que nosotros no sabemos cómo es el *software* y el *hardware* que ocupa nuestro cerebro.

Así, lo que hay que hacer es analizar estos procesos de generación y comprensión lingüística de forma hipotética, proponiendo teorías que, conforme las vamos corrigiendo según aparecen nuevas evidencias que nos ayudan a falsear y corregir tales hipótesis.

Del estructuralismo al generativismo (3)

Para comprender mejor esto, Chomsky consideró que las lenguas humanas funcionan con dos módulos básicos: la **competencia** y la **actuación**.



Competencia: es un módulo que concentra el conjunto de reglas necesarias para producir y/o comprender cadenas sintácticas.

Actuación: es un módulo que proyecta todas las manifestaciones reales de una cadena sintáctica en una situación dada.

Del estructuralismo al generativismo (4)

Veamos un ejemplo para ver cómo funcionan estos módulos. ¿Cuál es la diferencia entre estas dos oraciones? ¿Cuál les parece que es más “fácil” de comprender?:

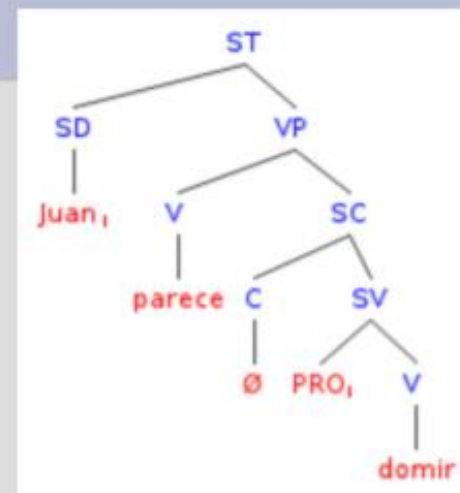
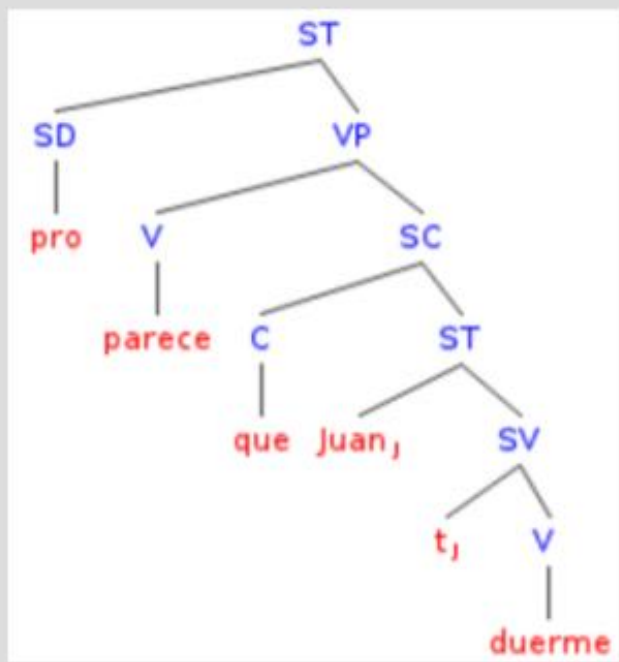
1. Juan parece dormir
2. Parece que Juan duerme

Explicación: si toman la primera opción, lo que están suponiendo es que una oración que siga un patrón *nombre + verbo finito + verbo infinitivo* es más *básico* (o por lo menos más esperable) que *verbo finito + pronombre + nombre + verbo finito*.

De acuerdo con la gramática generativa, la oración (1) sería un patrón de oración que pertenece al **módulo de la competencia**; en cambio la oración (2) es una **transformación del primero**, y por tanto lo situamos en el **módulo de la actuación**.

Del estructuralismo al generativismo (5)

Si lo anterior es así, digamos que la **competencia** es una enorme base de datos virtual que nos ofrece **estructuras sintácticas** como la siguiente:



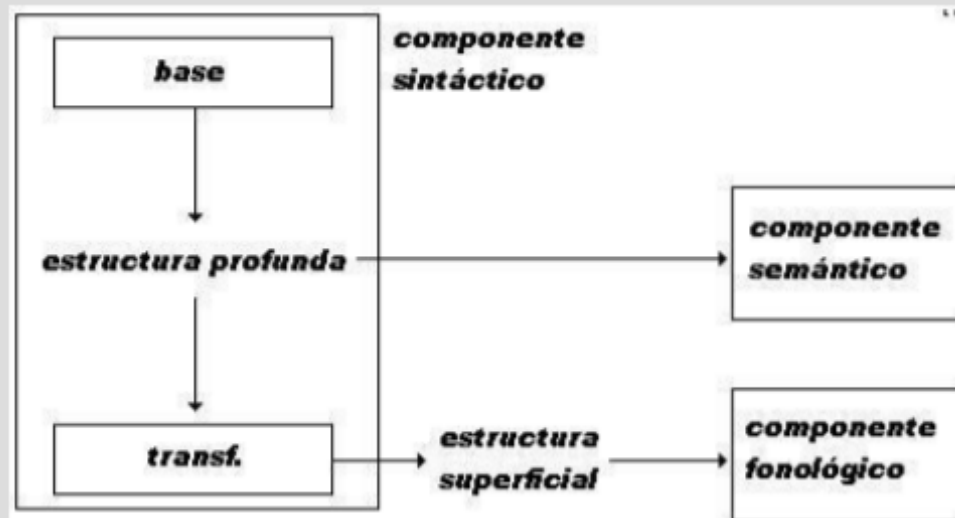
Por su parte, la **actuación** procesa patrones como el anterior, y nos brinda secuencias alternas, las cuales son generadas aplicando **procesos de transformación**, que se van ajustando a lo que queremos expresar.

Del estructuralismo al generativismo (6)

Tomando en cuenta esta diferencia entre competencia y actuación, el generativismo se plantea **desarrollar modelos teóricos que expliquen cómo opera un sistema lingüístico humano, partiendo del análisis de los datos observables que arroja una lengua (el módulo de la actuación), y haciendo inferencias para construir un modelo abstracto y universal de tal sistema (el módulo de la competencia).**

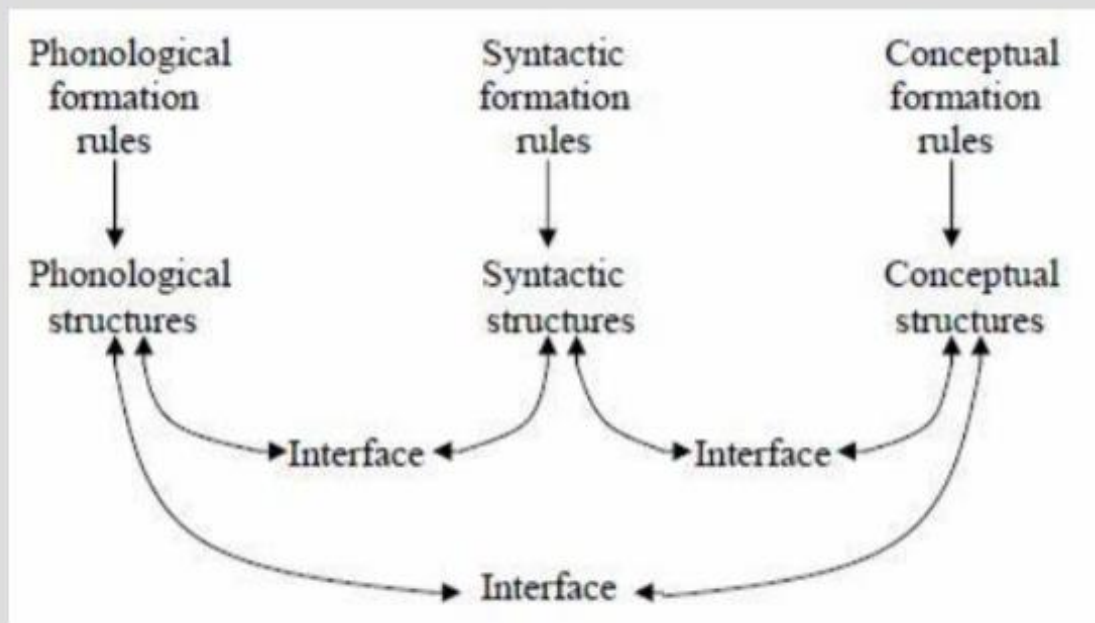
Algunas preguntas que se hacen tomando en cuenta estos presupuestos son:

1. ¿Todas las lenguas humanas comparten los mismos patrones y las mismas reglas de construcción?
2. ¿Podemos diseñar un modelo que represente cómo operan estos elementos universales?



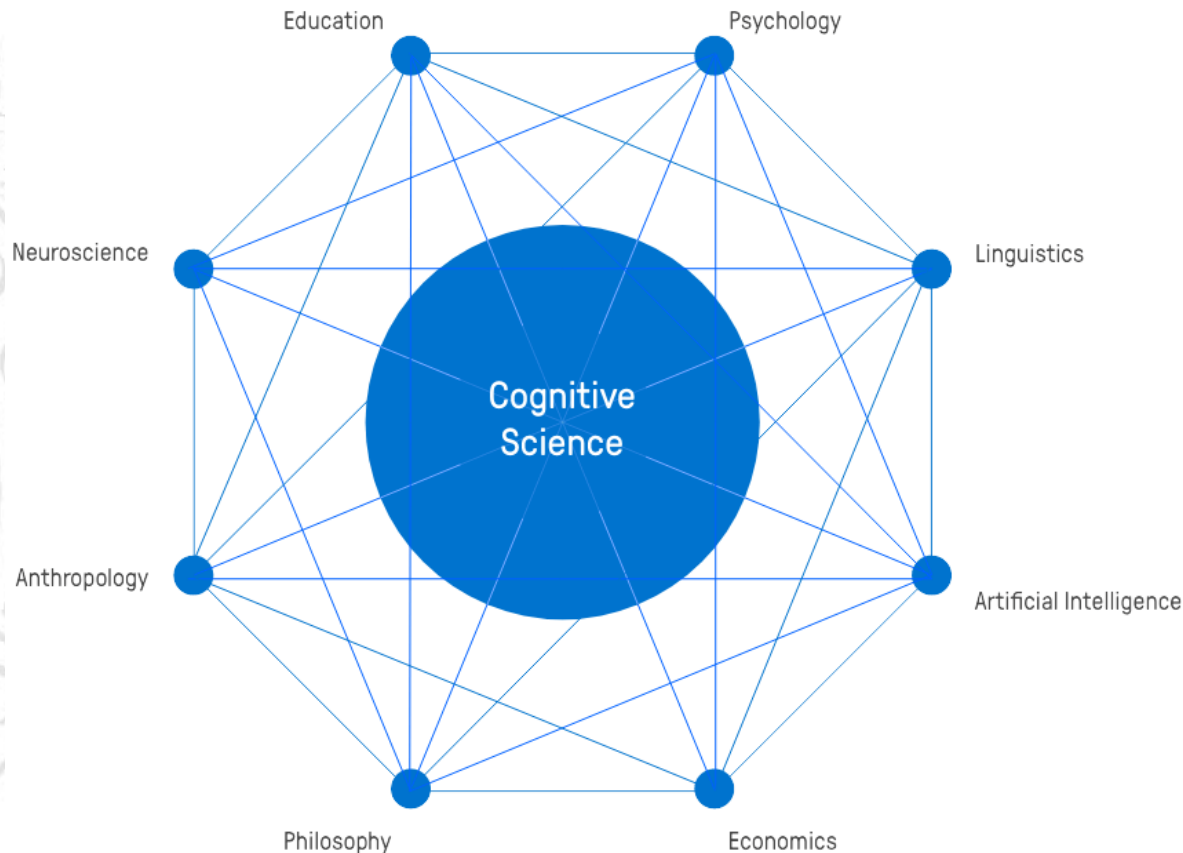
Del estructuralismo al generativismo (7)

Para el generativismo, el mejor objeto de estudio que hay para responder a preguntas como la anterior es el **nivel de la sintaxis o de la gramática**, en concreto porque este nivel **funciona como una interfaz que nos permite estudiar cuáles son los componentes formales de una lengua** (fonemas y palabras), **y cuáles son sus componentes conceptuales** (significados). Cuando juntamos y combinamos estos componentes, **creamos lo que se conoce como estructuras sintácticas** (frases y oraciones), las cuales son objetos de estudio perceptibles para nosotros.



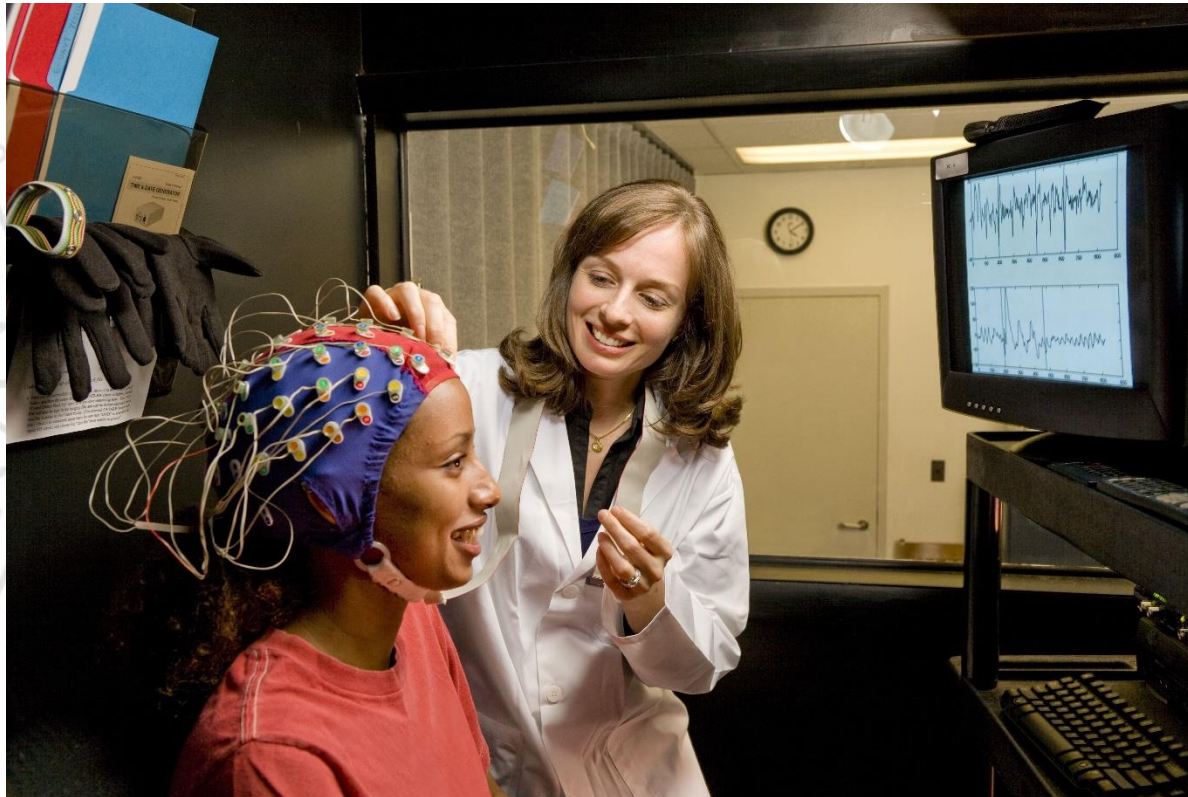
Spreading the news around the world (1)

La irrupción del generativismo propuesto y promovido por Noam Chomsky y sus colegas derivó en todo un cambio paradigmático en la lingüística y en otras áreas relacionadas. Para los años 70 y 80, p.e., se había desarrollado una nueva área de investigación, las llamadas **ciencias cognitivas**:



Spreading the news around the world (2)

Las ciencias cognitivas básicamente se enfoca en analizar y describir cómo opera la mente humana. Para ello, echan mano de diferentes perspectivas que les permiten concebir y determinar qué clase de procesamiento se da en nuestro cerebro para resolver problemas como la percepción del espacio y el tiempo, el reconocimiento de patrones, la comprensión y generación de lenguaje, etc.

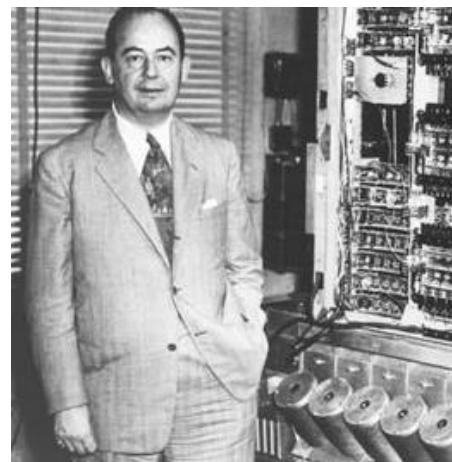


Spreading the news around the world (3)

Y por supuesto, desde las ciencias cognitivas la gramática generativa saltó a la palestra de un área emergente, la cual tenía como padres fundadores a un matemático británico que se había dedicado a decodificar mensajes para el Ejército Británico en la 2ª Guerra Mundial; y otro húngaro radicado en Estados Unidos, quien entre otras cosas había colaborado en la implementación de ENIAC, uno de los primeros grandes computadores creados por el Ejército de Estados Unidos: **Alan Turing** y **John von Neumann**.



Alan Turing
(1912-1954)



John von Neumann
(1903-1957)

Spreading the news around the world (4)

Los trabajos de Turing y von Neumann marcaron buena parte de líneas de investigación emergentes como la informática, la cibernética y la robótica (apenas recién formadas).

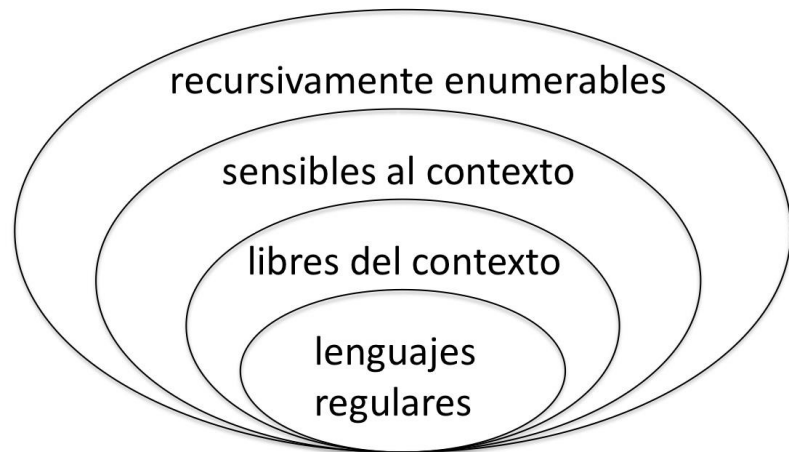
Para 1956, un recién doctorado en matemáticas llamado **John McCarthy** (1927-2011) convocó a varios colegas a participar en una conferencia en el **Dartmouth College** (Estados Unidos) la cual propuso que se llamara **Dartmouth Summer Research Project on Artificial Intelligence**. Por supuesto, en ese año nadie tenía una idea clara sobre lo que quería decir eso de “Artificial Intelligence”.



Participantes de la convención de dartmouth, incluidos Marvin Minsky, Claude Shannon y Ray Solomonoff

Spreading the news around the world (4)

Y para poner las cosas más interesantes, un joven **Noam Chomsky** (1928) publicó en el mismo año un artículo titulado: “Three models for the description of language”, en donde planteaba una jerarquía de lenguajes formales, los cuales eran vistos como claros ejemplos respecto a las limitaciones que hay para crear un lenguaje artificial que pueda “imitar” el humano. Por supuesto, este artículo era un desafío directo a las ideas que McCarthy y sus colegas estaban discutiendo.



All Together Now! (1)

McCarthy y un colega que lo acompañó en la conferencia de Dartmouth, **Marvin Minsky** (1927-2016) se pusieron manos a la obra y decidieron fundar un laboratorio de inteligencia artificial en el **Massachusetts Institute of Technology (MIT)** en 1959.



John
McCarthy

Marvin
Minsky



MIT Computer Science &
Artificial Intelligence Lab

www.csail.mit.edu

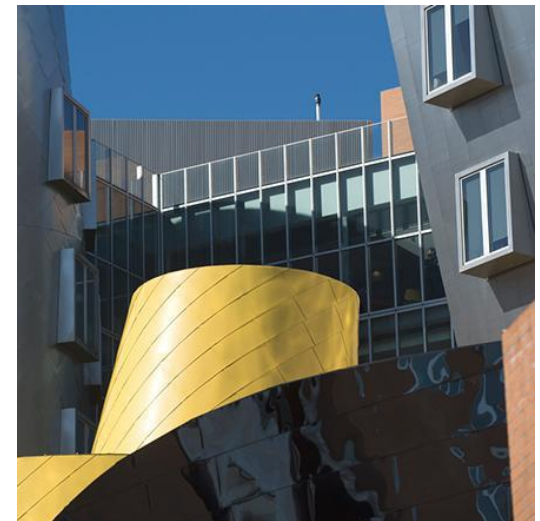
All Together Now! (2)

Y por azares del destino, en 1961 el MIT fundó su **Departamento de Lingüística y Filosofía**, el cual contaba con algunos profesores e investigadores de amplia trayectoria como **Roman Jakobson**, y otros recién doctorados que empezaban metidos de lleno en la recién fundada gramática generativa, tales como **Morris Halle** (1923-2018), o el mismo Chomsky.



Morris
Halle

Noam
Chomsky



MIT Linguistic &
Philosophy Department

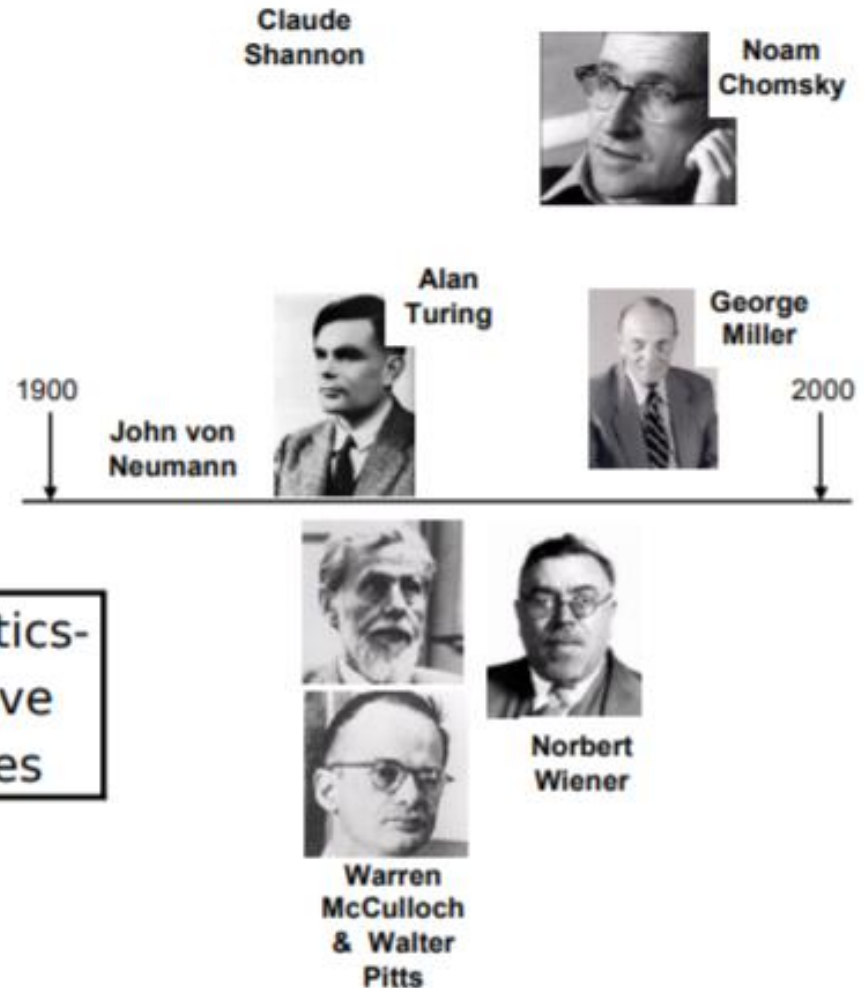
<http://linguistics.mit.edu>

All Together Now! (3)

Una forma resumida de presentar esta coincidencia entre la lingüística y la inteligencia artificial es la que propone **John Goldsmith** (University of Chicago):

Para saber más, véase:

<http://people.cs.uchicago.edu/~jagoldsm/slides/2010-historyoflinguistics.pdf>

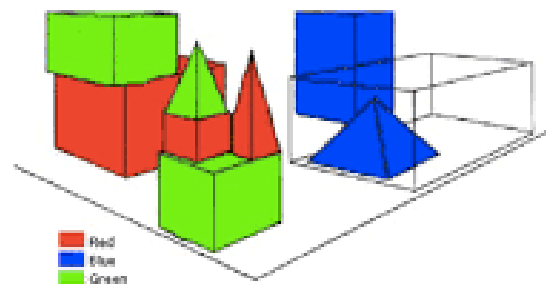


No todo lo que brilla es oro... (1)

Durante el periodo de 1956 hasta 1974 se dio la *edad dorada* de la inteligencia artificial, en la cual ésta —junto con la gramática generativa— se expandió por el mundo, desarrollando proyectos de investigación y programas de formación universitaria que plantearon un nuevo horizonte. Un ejemplo es el sistema **SHRDLU**, desarrollado por **Terry Winograd** (1946), de 1968 a 1970 en el MIT.



Terry Winograd



Person: Pick up a big red block.

Computer: OK.

Person: Grasp the pyramid.

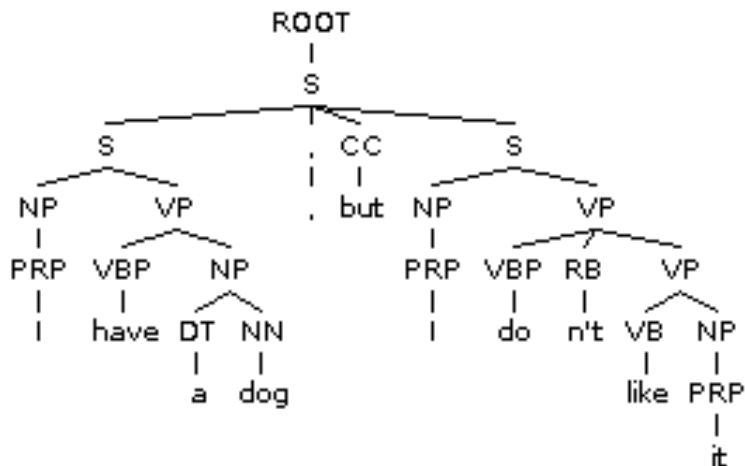
Computer: I don't understand which pyramid you mean.

No todo lo que brilla es oro... (2)

Empero, de 1980 a 1990 la inteligencia artificial fue pausando su desarrollo, sobre todo en lo que respecta a sus proyectos relacionados con la generación y comprensión de lenguaje humano.

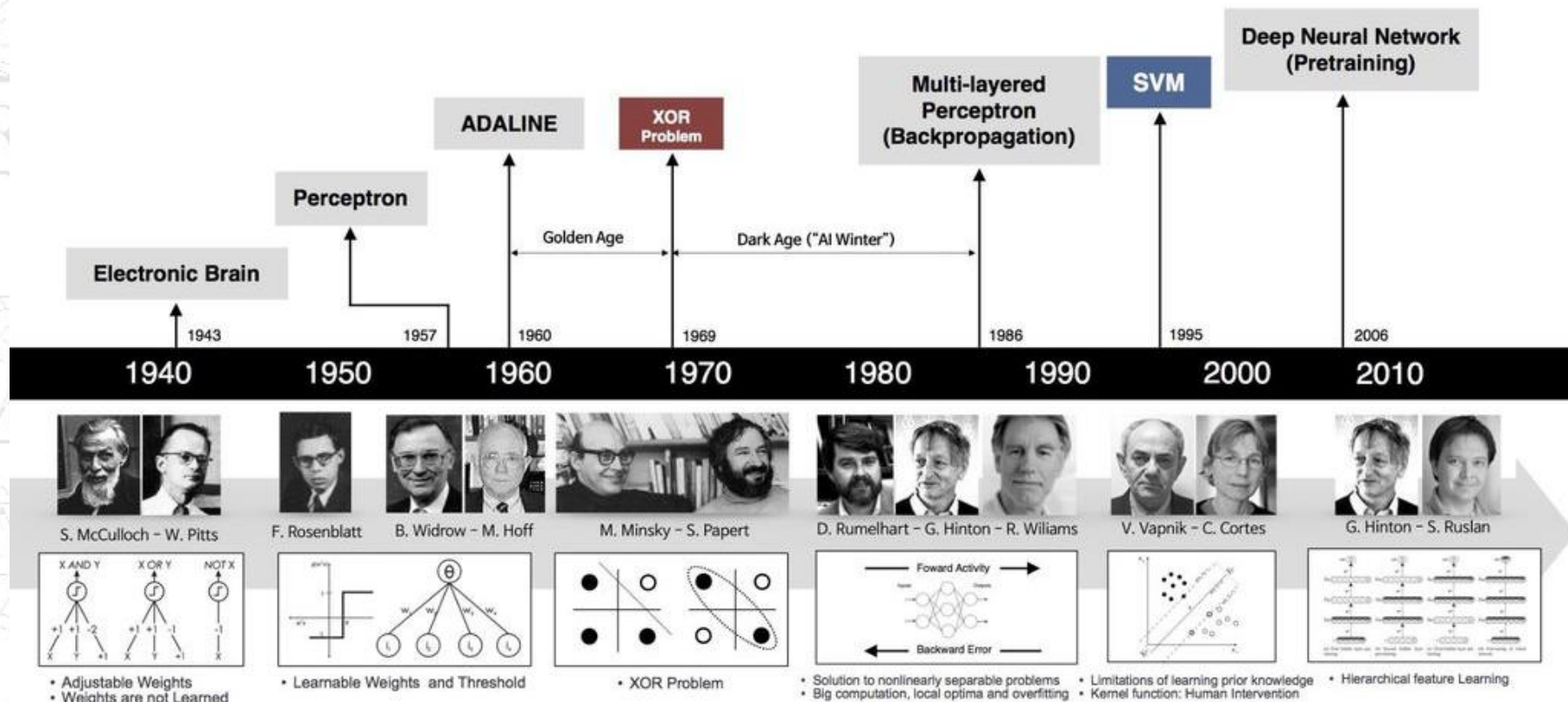
Uno de los problemas era que los modelos propuestos por la gramática generativa eran demasiado formales y complejos de programar.

```
(ROOT
  (S
    (S
      (NP (PRP I))
      (VP (VBP have)
        (NP (DT a) (NN dog))))
    (, ,)
    (CC but)
    (S
      (NP (PRP I))
      (VP (VBP do) (RB n't)
        (VP (VB like)
          (NP (PRP it))))))))
```



No todo lo que brilla es oro... (3)

Por otro lado, propuestas recién desarrolladas como las **redes neuronales** no lograban resolver algunos problemas para su implementación, p. e., un modelado eficaz que ayudara a lograr un mejor entrenamiento a partir de un mínimo de reglas.



Las goteras en el techo (1)

De parte del generativismo, tampoco las cosas iban bien. Un primer problema que tuvieron fue su decisión de dejar de lado la relación que mantiene el lenguaje humano con la cultura. Los lingüistas norteamericanos estaban interesados por esta relación desde principios del XX, y un excelente ejemplo de tal interés son los estudios de **Edward Sapir**.

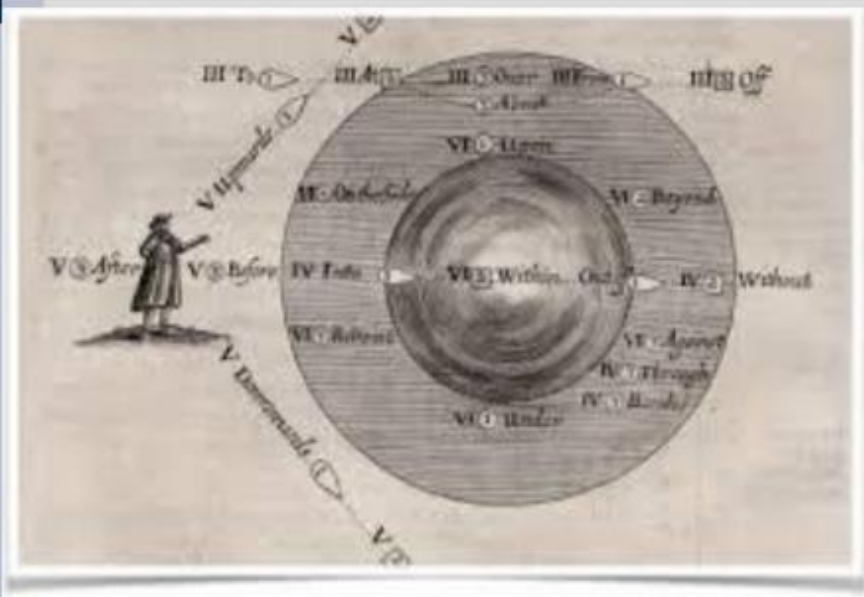


Edward Sapir
(1884-1939)

Sapir estudió muchos idiomas indígenas de Estados Unidos, particularmente el **hopi**, una lengua que pertenece a una familia llamada **yuto-azteca**, por lo que el hopi comparte rasgos con el **náhuatl** (la lengua que hablaban los aztecas de México). Fue precisamente estudiando esta lengua que descubrió un fenómeno sumamente curioso: cuando le preguntaba a algún hopi qué palabras ocupaban para referirse a días pasados (*ayer, anteayer*) o futuros (*mañana, pasado mañana*)..., la respuesta era que no entendían la pregunta, porque **no sabían qué era eso de dividir los días en pasados y futuros.**

Las goteras en el techo (2)

Los estudios que realizó Sapir sobre el hopi dieron origen a otra perspectiva teórica alterna al generativismo que se conoce como **teoría relativista**, la cual sostiene que **la lengua que domina un hablante influye en su visión de la realidad**, o por lo menos, **le ayuda a clasificar tal realidad en unidades, estructuras y categorías lingüísticas que le son propias.**



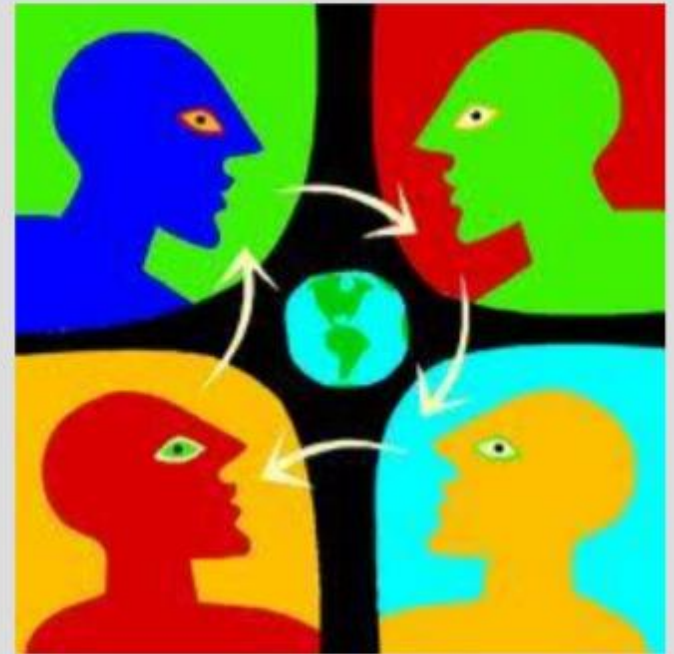
El relativismo se ha ido modificando con el paso de los años, pero nos ha dejado muchas cuestiones a considerar:

- i. Hay una relación estrecha entre la percepción de la realidad y su expresión por medios lingüísticos
- ii. Los sistemas lingüísticos **no solamente generan o procesan oraciones**, también sirven para comunicar esquemas cognitivos, culturales, sociales, etc.
- iii. Estos fenómenos deben ser estudiados por la lingüística también.

Las goteras en el techo (3)

La teoría relativista, al igual que la generativista, también toma como base el método hipotético-deductivo, pero sus objetivos son diferentes: de entrada, no se busca proponer un modelo lingüístico universal, sino más bien **identificar las particulares que hacen distinta una lengua de otra.**

Si se dan estas diferencias, ¿por qué se producen? ¿Son el resultado de una percepción particular de la realidad? ¿Son el resultado de interacciones entre hablantes (aspectos comunicativos) ¿Se dan por concepciones sociales que los hablantes asumen —ya sea de forma explícita o implícita— y manifiestan en sus interacciones? ¿Se dan porque son síntomas de problemas de procesamiento lingüístico: afasias, diferencias de edad, trastornos psicológicos u otros factores (aspectos cognitivos)?



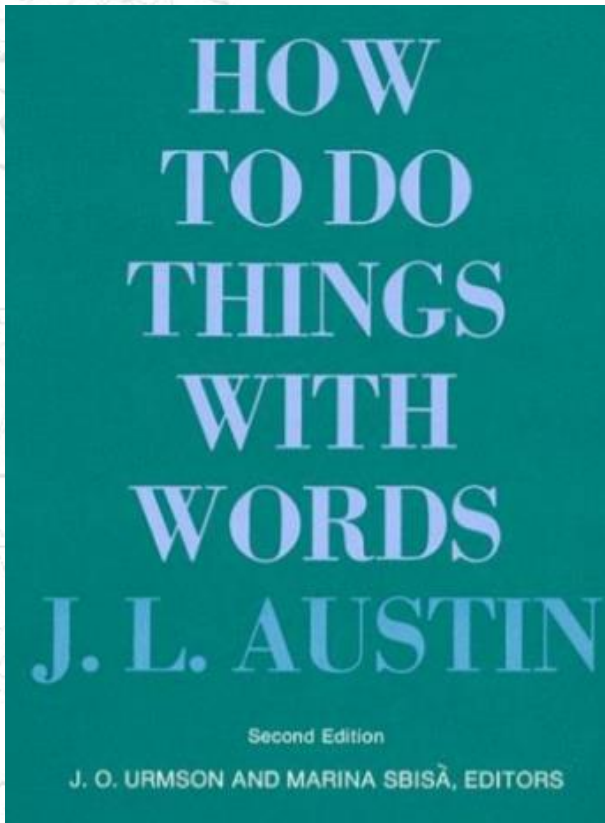
Las goteras en el techo (4)

Si consideramos las preguntas anteriores y tratamos de darles respuesta, nos damos cuenta que los sistemas lingüísticos son más complejos de lo que pensamos, dado que mantienen relaciones dinámicas con otras esferas de la vida humana, por lo que tratamos de hacer investigaciones que tratan de explicar cómo **funciona** un idioma como mecanismo de expresión, qué relación mantiene con otros **sistemas y mecanismos cognitivos** (percepción, motricidad, aprendizaje, visión, etc.), cómo **evolucionan** a lo largo del tiempo (variación lingüística), qué **rasgos** son **particulares** a una **variante determinada** (dialectología), y un largo etc.



La irrupción de la pragmática (1)

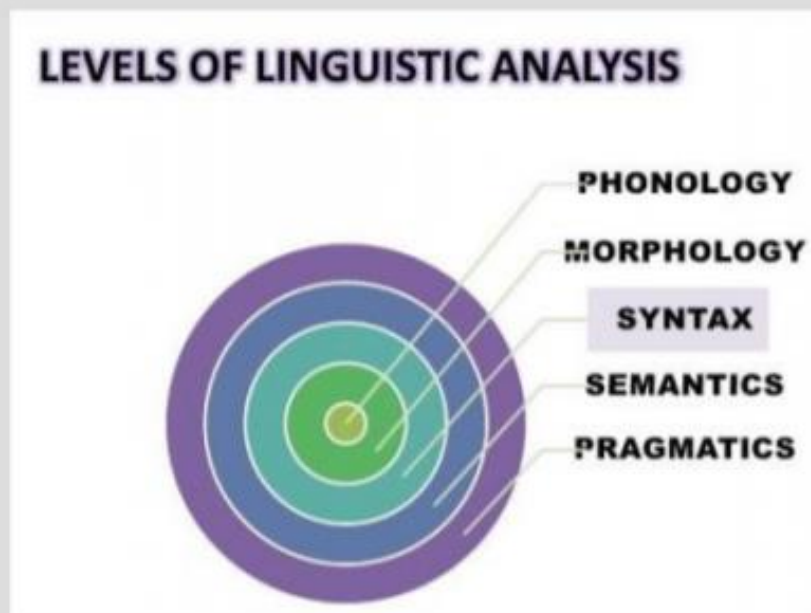
Asimismo, tras varios años de mantener una postura discreta en los círculos lingüísticos y filosóficos, el estudio pragmático cobró nueva fuerza gracias a la **teoría de los actos de habla**, propuesta por **John L. Austin** (1911-1960), y reformulada por **John Searle**.



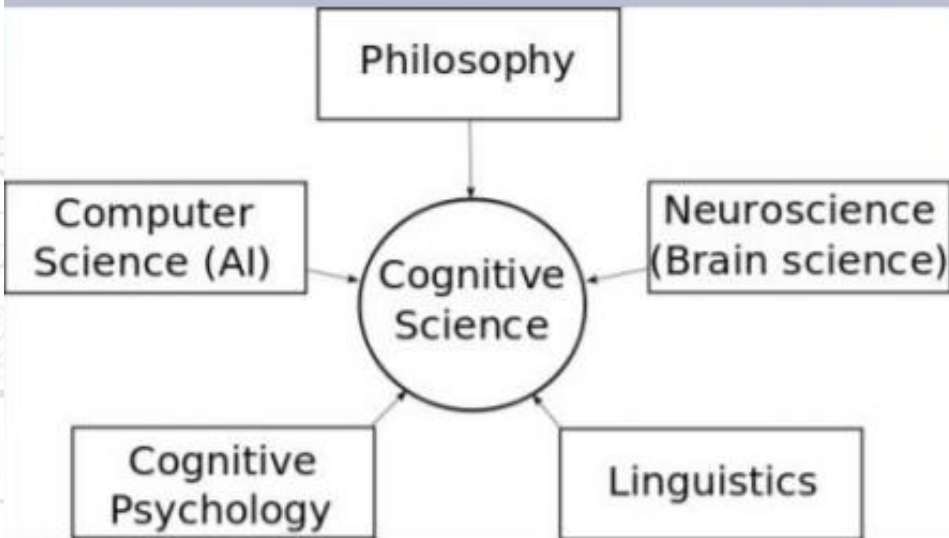
La irrupción de la pragmática (2)

Searle fue consciente de que su trabajo planteaba otros objetivos que no cuadraban con los objetivos que se había planteado estudiar Chomsky y sus numerosos colegas.

Por esta razón, consideró que la gramática generativa se encargaría de dar cuenta de los niveles *axiomáticos* o *duros* del lenguaje (fonética, morfología, etc.); de tal suerte que la pragmática, si bien queda en la periferia, no por ello es un elemento menor: es más bien una **interfaz** que ayuda a los humanos a interactuar con su entorno.



La irrupción de la pragmática (3)



A partir de esta decisión, Searle fue enfocando la investigación pragmática como una línea que no estaba limitada a trabajar en los temas que no les interesaran a los gramáticos generativistas, sino como un proyecto que colindaba de forma cercana con dos campos de investigación novedosos en esa época: las **ciencias cognitivas** y la **inteligencia artificial**.



La irrupción de la pragmática (4)

Searle empezó a plantear preguntas claras y concretas que Chomsky y sus colegas no podían responder sin evitar la polémica, p. e.:

1. ¿Los humanos venimos programados no solo para comprender y producir oraciones, sino también para hacerlas funcionar en un acto de habla determinado?
2. ¿Hay alguna especie animal que sea capaz de generar o inferir actos de habla —ya sean simples o complejos— de forma similar a como lo hacen los humanos?
3. Un robot (o un computador muy avanzado) que sepa analizar y entender una gramática, ¿es ya un buen candidato a ser considerado como un hablante, o necesita conocer el nivel pragmático para decir que realmente lo es?
4. ¿Qué implicación tiene, de acuerdo con estas preguntas, el que una persona posea el conocimiento pragmático que le brinda su lengua?

La irrupción de la pragmática (5)

Y siguiendo con la ilación de estas preguntas, Searle va a plantear un problema que únicamente la pragmática puede responder de forma satisfactoria: supongamos que logramos programar un robot que sabe comprender y producir oraciones con significado de forma coherente. Sin embargo, el único camino que tenemos para comprobar que realmente puede hacerlo es..., ponerlo a interactuar con humanos, de tal modo que si es capaz de comunicarse con ellos de forma *natural*, entonces podremos decir que tal robot realmente tiene un conocimiento lingüístico, además de que sabe usarlo de forma eficiente.



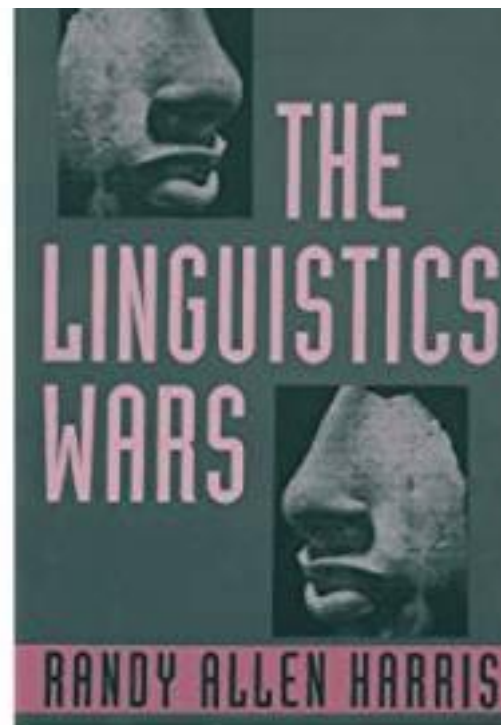
Las guerras lingüísticas (1)

Estos problemas llevaron a una ruptura entre Chomsky y varios de sus colegas, quienes consideraron que no lograban resolver cuestiones como el procesamiento semántico, o la relación entre lenguaje y contexto social.

A este periodo que va (aproximadamente) de 1970 a 1980 se le conoce como *Guerras Lingüísticas*, las cuales fueron documentadas por **Randy Allen Harris**.



Randy Allen
Harris

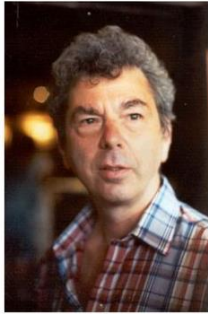


Las guerras lingüísticas (2)

Algunos de ellos decidieron que era el momento de crear un nuevo proyecto alejado de los postulados generativistas, y más focalizado al estudio de las facultades lingüísticas en uso. Así nace la llamada **lingüística cognitiva**.



Charles Fillmore



Leonard Talmy



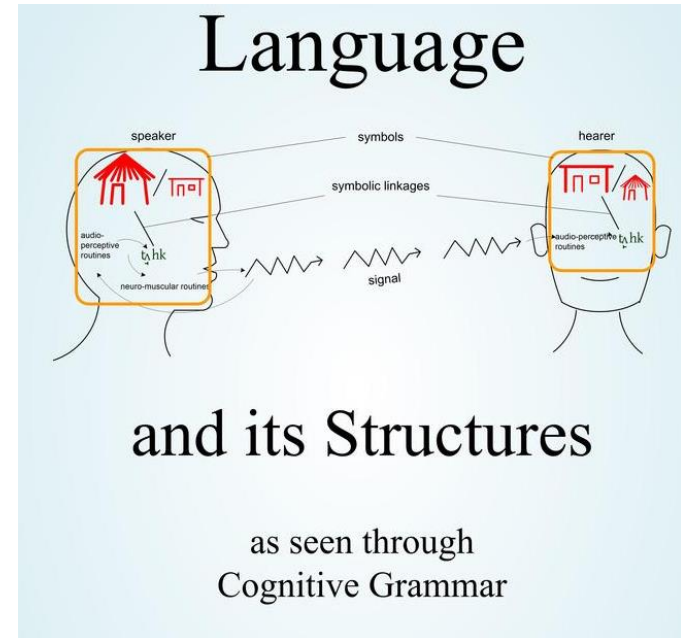
Ronald Langacker



George Lakoff



Mark Johnson

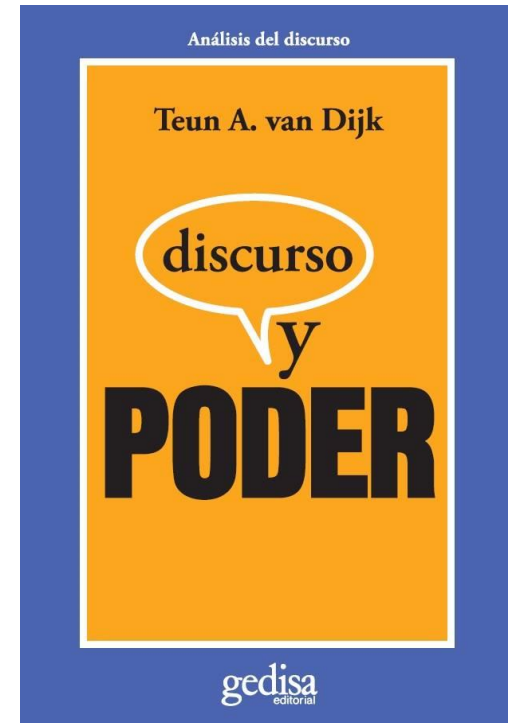
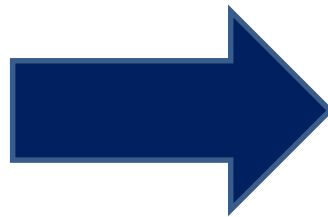


Las guerras lingüísticas (3)

Otros consideraron que los generativistas había olvidado por completo el estudio del **discurso**, considerado como una unidad compleja que pone en juego todos nuestros conocimientos lingüísticos en el plano de la interacción social, particularmente el de las relaciones de poder. En esta línea, uno de los investigadores más destacados es **Teun van Dijk**.



Teun van Dijk
(1943)



Las guerras lingüísticas (4)

En resumen, para la década de los 90, la lingüística estaba dividida entre dos grandes bandos:

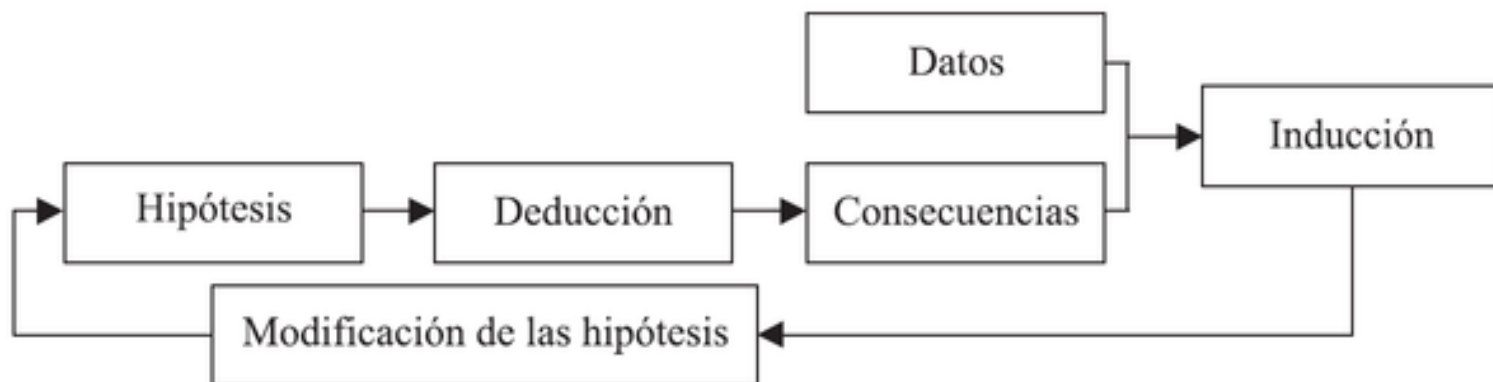
El generativismo, como vimos, parte de la idea de ver el procesamiento lingüístico como una operación ejecutada por un computador, el cual cuenta con dos módulos para realizar tal procesamiento: uno llamado **competencia**, el cual combina reglas sintácticas con patrones de combinación; y otro llamado **actuación**, encargado de generar resultados, adecuados a un contexto determinado.

El enfoque funcional-cognitivo, por su parte, retoma ideas del **relativismo**, y plantea que el lenguaje mantiene estrechas relaciones con otros sistemas humanos, como los cognitivos, los sociales y los culturales, por lo que su comportamiento es el de un sistema **interactivo**. Esta interacción determina su desarrollo y evolución, e igualmente proyecta diferencias entre idiomas.

Dejemos que hablen los datos (1)

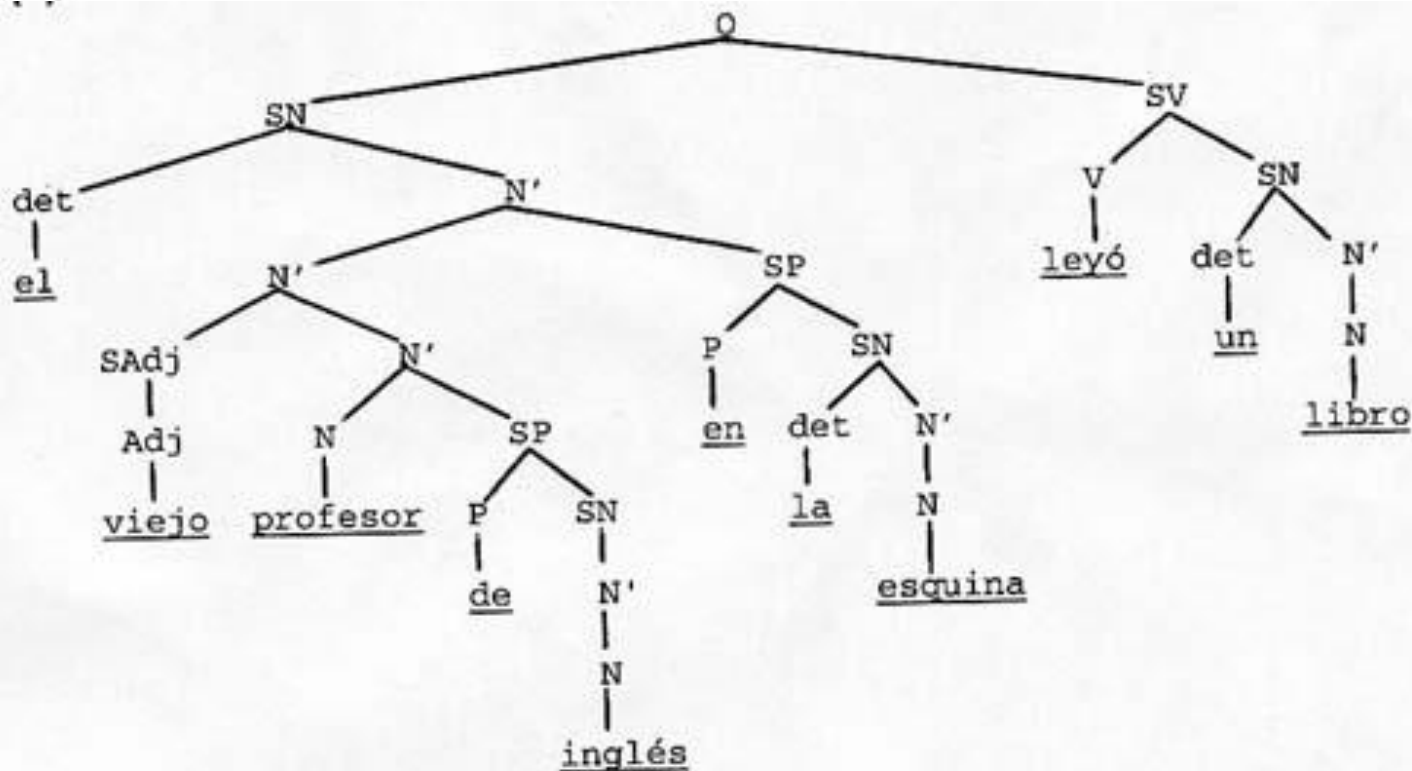
Un aspecto que se puede vislumbrar aquí es que la lingüística, como ciencia, la mayoría de las veces pondera el uso de un método hipotético-deductivo, en una versión reformulada por **Karl Popper** con su propuesta falsacionista.

El generativismo estructuró su base epistemológica sobre esta propuesta.



Dejemos que hablen los datos (2)

Dada esta perspectiva, el generativismo se decantó por un modelo de análisis que emplea la inducción como un mecanismo para generar ejemplos y contra-ejemplos que validen (o invaliden) sus hipótesis.



Dejemos que hablen los datos (3)

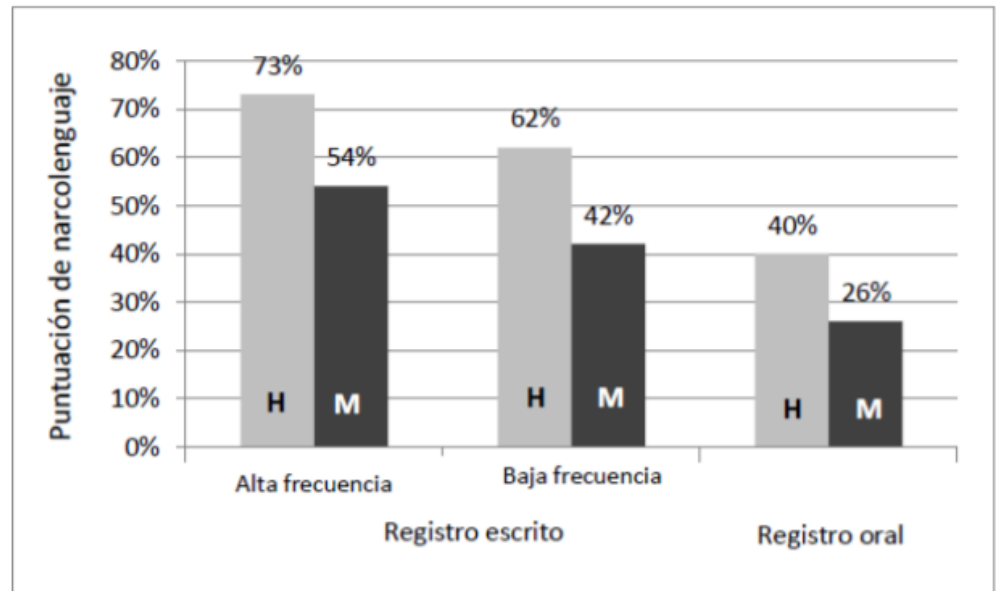
Sin embargo, muchos lingüistas —en particular los sociolingüistas— apostaron por introducir métodos estadísticos que les permitieran sustentar empíricamente sus hipótesis.



William Labov
(1927)



Yolanda Lastra
(1932)



Dejemos que hablen los datos (4)

Un aspecto importante a considerar en esta clase de investigaciones es el **proceso de transcripción** de los datos colectados. Lo que se espera aquí es que tal transcripción describa de la manera más exacta posible el experimento con el cual se tomaron dichos datos.



La transcripción es un proceso que requiere mucha atención, pues es aquí en donde el investigador debe resolver muchas cuestiones que validan (o invalidan) sus decisiones metodológicas.

Además, los procesos de traducción implican también:

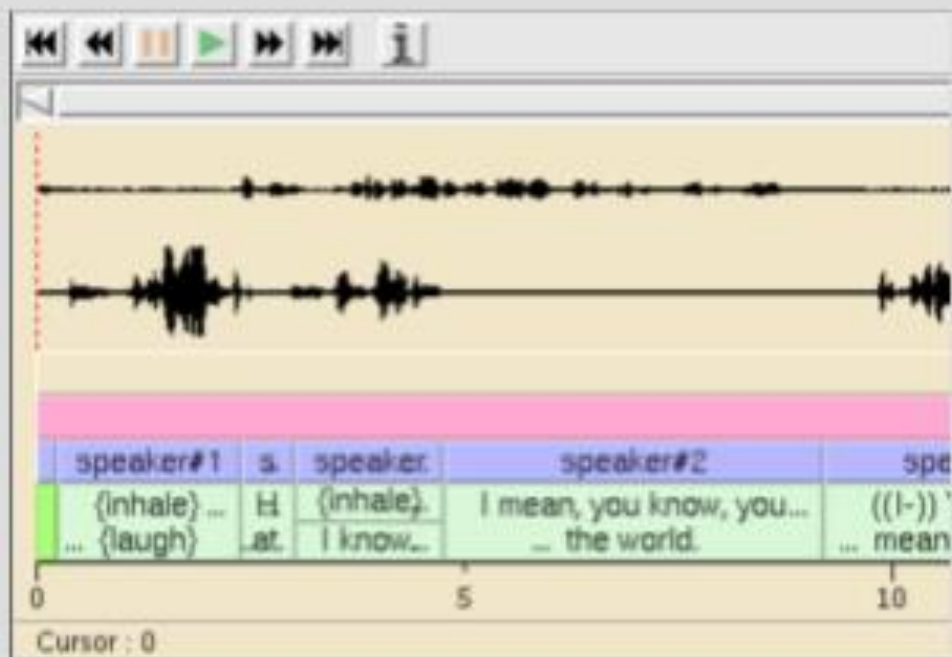
- i. La inversión de horas de trabajo dedicadas a la traducción de datos en registros escritos.
- ii. El entrenamiento de los transcritores, sobre todo para que sean capaces de reconocer el objeto de estudio considerado.

Dejemos que hablen los datos (5)

La transcripción brinda a los investigadores la posibilidad de sistematizar la representación de los fenómenos que se van a analizar, en aras de evitar interpretaciones erróneas.

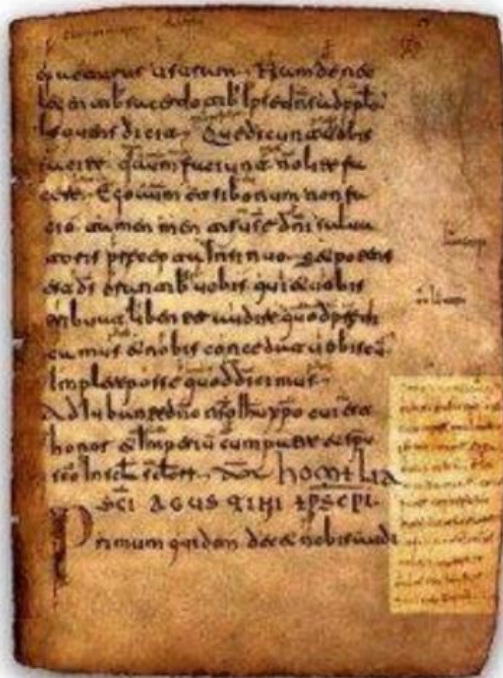
Algunos de los aspectos que deben atenderse para evitar esta clase de interpretaciones son:

- La descripción de interacciones entre el investigador y el sujeto participante
- La mención de pausas o silencios entre ambos, así como la marcación temporal de su duración
- Indicar si hay un soporte acústico o visual de tal interacción



La lingüística de corpus (1)

Los trabajos realizados desde la sociolingüística revitalizaron una herramienta planteada desde el S. XIX, y que había sido obviada por completo por el generativismo: los **corpus lingüísticos**, esto es, colecciones de datos que ofrezcan una representación de un fenómeno lingüístico específico:



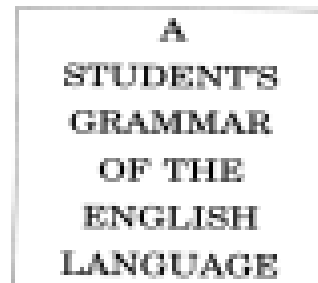
Como autorio de
nuestro dueno dueno
Christo dueno
salhatore qual dueno
get ena honore e qual
duenno tienet cla
mandatione como
Patre como Spiritu
Sancto enos uicculos
de losicculos. Para
nos Deus onnipotes
tal servitio fere ke
denante cla sua face
gaudioso seguimus.
Amen.

A raíz del interés por la filología europea por el análisis de textos sánscritos, se ponderó la recopilación de datos lingüísticos que permitieran la reconstrucción de una “protolengua”, la cual sería la base de todas las lenguas pertenecientes a la familia indoeuropea.

La lingüística de corpus (2)

Si bien estos trabajos no lograron reconstruir esta “lengua original”, muchas de los métodos usados dieron lugar a toda una serie de estudios enfocados a la recopilación y documentación de datos lingüísticos para fines tales como la historiografía lingüística, la dialectología, la lexicografía, la sociolingüística, la estilística, y otras similares.

Un ejemplo es el trabajo de **Sir Randolph Quirk**, enfocado en el estudio y la comparación de distintas variedades y usos del inglés británico. En 1960 elaboró una base de datos de un millón de palabras, la cual le permitió hacer cálculos estadísticos para validar sus ideas sobre variación dialectal.



Sidney Greenbaum
Randolph Quirk

Randolph Quirk (1920-2017)

La lingüística de corpus (3)

La propuesta de Quirk tuvo una buena recepción no solo porque revalorizó el empleo de conjuntos de datos para contrastar sus hipótesis, sino también porque fue la base para crear grandes corpus lingüísticos, tales como el **British National Corpus** (BCN), cuyo diseño se llevó a cabo alrededor de 1990, y se concluyó en 1995.

B RITISH **N** ATIONAL **C** ORPUS

Para saber más, véase:
www.natcorp.ox.ac.uk

La lingüística de corpus (4)

En Estados Unidos desde 1961 en la Universidad de Brown se compiló un corpus con anotación gramatical manual, el cual fue validado por un grupo de lingüistas coordinados por **Henry Kučera** (1925–2010) y **Nelson Francis** (1910–2002), y se conoció como **Corpus Brown**.

BROWN
UNIVERSITY

The **Brown Corpus** of Standard American English was the first of the modern, computer readable, general corpora. It was compiled by W.N. Francis and H. Kucera, Brown University, Providence, RI. The corpus consists of one million words of American English texts printed in 1961.

Para saber más, véase:
[http://clu.uni.no/icame/manuals/
BROWN/INDEX.HTM](http://clu.uni.no/icame/manuals/BROWN/INDEX.HTM)

La lingüística de corpus (5)

Poco a poco las investigaciones con corpus lingüísticos fueron cobrando importancia en todo el mundo, tanto en sus cuestiones prácticas como teóricas.

En el primer rubro, en la Universidad de Pennsylvania se fundó la iniciativa **Linguistic Data Consortium (LDC)** en 1992, la cual tuvo por objetivo crear un enorme repositorio de corpus lingüísticos para la mayor parte de lenguas del mundo. Cabe señalar que uno de los patrocinadores es la **Defense Advanced Research Project Agency (DARPA)**.



Para saber más, véase:
www ldc upenn edu

La lingüística de corpus (5)

En el segundo rubro, se han desarrollado centros de investigación abovocados a la creación y estudio de los mismos, así como también se han generado muchas publicaciones especializadas, p. e., la revista **Corpus Linguistics & Linguistic Theory** es un proyecto que se enfoca sobre todo en sustentar la pertinencia teórica de emplearlos en cualquier investigación lingüística.



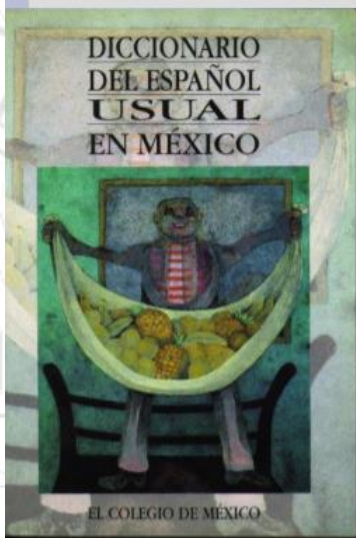
Para saber más, véase:
<http://ucrel.lancs.ac.uk>



Para saber más, véase:
www.degruyter.com/view/j/cllt

La lingüística de corpus (6)

En el caso de Latinoamérica, un proyecto pionero es del **Diccionario del Español de México**, coordinado por el **Dr. Luis Fernando Lara** (CELL-COLMEX), iniciado en 1973. Para desarrollar este diccionario, se compiló el **Corpus del Español Mexicano Contemporáneo (CEMC)**, el cual cuenta con un poco más de un millón de palabras tomadas de textos escritos entre 1921 hasta 1974.



El CEM puede ser considerado como el primer corpus latinoamericano que se compiló para ser procesado a través de sistemas de cómputo, lo que le da a este diccionario un valor especial: sus definiciones se basan en un análisis sobre el uso real de los hablantes sobre su vocabulario.

Para saber más, véase:
<http://cemc.colmex.mx>



Gracias por su atención

Blog del curso:

<http://cesaraguilar.weebly.com/curso-de-procesamiento-del-lenguaje-natural.html>