

# L'anàlisi de contingut en perspectiva qualitativa

Carles-Enric Riba Campos

PID\_00205322



*Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-NoComercial-SenseObraDerivada (BY-NC-ND) v.3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los i transmetre'ls públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), no en feu un ús comercial i no en feu obra derivada. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.ca>*

# Índex

<b>1. Bases metodològiques</b> .....	5
1.1. Coordenades de l'anàlisi de contingut respecte als eixos qualitatiu i quantitatiu .....	5
1.1.1. Anàlisi de contingut qualitativa i quantitativa .....	5
1.1.2. Definicions de l'anàlisi de contingut .....	8
1.1.3. Tipus d'anàlisi de contingut .....	8
1.2. El mostratge en l'anàlisi de contingut .....	10
1.2.1. Mostratge intertextual .....	11
1.2.2. Mostratge intratextual .....	13
1.3. Principis de segmentació i creació d'unitats en l'anàlisi de contingut: text i context .....	15
1.3.1. Text i context .....	15
1.3.2. Unitats textuais .....	16
1.3.3. Unitats contextuais .....	18
1.3.4. Les unitats textuais de registre i la seva codificació .....	23
1.4. Les categories en l'anàlisi de contingut .....	24
1.4.1. La categorització sistemàtica de segments textuais .....	25
1.4.2. Criteris per a la creació de categories i la construcció de sistemes de categories en anàlisi de contingut .....	25
1.4.3. Funcions d'inferència .....	30
1.5. Mesures i índexs bàsics en l'anàlisi de contingut .....	31
1.5.1. Mesures bàsiques de freqüència i regles d'enumeració .....	32
1.5.2. Altres mesures derivades de la freqüència o que la pressuposen .....	35
1.5.3. Índexs .....	37
1.6. Mesures de contingència i seqüencials en l'anàlisi de contingut .....	38
1.6.1. Mesures de coocurrència .....	38
1.6.2. Mesures seqüencials .....	40
<b>2. Exemples desenvolupats d'anàlisi de contingut</b> .....	44
2.1. Exemple 1: Recomptes de freqüència en general .....	44
2.2. Exemple 2: Anàlisi de contingència i anàlisi seqüencial .....	47
2.2.1. Exemple d'anàlisi seqüencial amb una taula de contingència complexa .....	51
2.3. Exemple 3: Anàlisi avaluativa .....	54
<b>3. La qualitat de les dades en l'anàlisi de contingut</b> .....	58
3.1. El marc general de la qualitat de dades en l'anàlisi de contingut: recordatori .....	58
3.1.1. Fiabilitat, precisió, validesa i generalització .....	58

3.2.	Els diferents enfocaments de la fiabilitat en l'anàlisi de contingut .....	59
3.2.1.	Dimensions de la fiabilitat en general (metodologia qualitativa i observacional) .....	60
3.2.2.	Dimensions de la fiabilitat en l'anàlisi de contingut .....	63
3.3.	L'aproximació correlacional a la fiabilitat .....	66
3.4.	L'aproximació probabilística i "punt a punt" a la fiabilitat .....	68
3.4.1.	Acord o coincidència sobre la totalitat del sistema de categories .....	69
3.4.2.	Acord o coincidència sobre una categoria en concret ...	73
3.4.3.	Acord quan no es compleix la condició d'exhaustivitat .....	74
<b>4.</b>	<b>Síntesi de l'anàlisi de contingut.....</b>	<b>76</b>
4.1.	Un exemple integrador: la meditació transcendental mitjançant els seus documents .....	76
4.2.	Sinopsi final en forma de diagrama .....	78
<b>Bibliografia.....</b>		<b>81</b>

## 1. Bases metodològiques

### 1.1. Coordenades de l'anàlisi de contingut respecte als eixos qualitatiu i quantitatiu

La denominació *anàlisi de contingut* denota un conjunt de tècniques d'anàlisi, qualitatives o quantitatives, aplicades a materials significatius, textuais o icònics.

#### 1.1.1. Anàlisi de contingut qualitativa i quantitativa

Com hem afirmat repetides vegades, en la investigació qualitativa l'activitat d'anàlisi neix pràcticament amb la planificació mateixa del mostratge i del registre i pren embranzida amb la recollida factual d'informació, recollida impossible d'acomplir sense una certa segmentació del flux d'esdeveniments, àdhuc sense alguna codificació o categorització del material recol·lectat. I creiem que salta a la vista que les fases de categorització interpretativa (axial o selectiva, en termes de la teoria fonamentada) que tot just hem comentat demanen ja de ple dret un treball autènticament analític per part de l'investigador.

Igualment, ha d'haver quedat prou clar en els mòduls precedents que aquest treball d'anàlisi qualitativa es refereix a textos, textos creats per l'investigador (en qualitat de transcripcions, registres narratius, *memos*, etc.) o generats pels subjectes o les institucions.

Ara bé, les operacions d'anàlisi no solament impliquen divisió o desconstrucció del text, sinó també síntesi i esforç constructiu, els quals són quasi impossibles de dur a terme sense fer comparacions i establir i justificar relacions, a diferents nivells, entre les categories proposades al final de la fase que acabem d'exposar; o sense arribar a aïllar regles o principis que permetin una interpretació holística del material estudiat. Aquest establiment de relacions pot exigir una certa quantificació, que recolza –això sí– en l'anàlisi prèvia, categorial i qualitativa, dels textos.

L'anàlisi de contingut produeix resultats, interpretacions o conclusions com qualsevol altra anàlisi científica. La major part dels motllos metodològics en els quals s'acomoda qualsevol tipus d'investigació psicològica poden acomodar també una anàlisi de contingut. Així, podem fer anàlisi de contingut experimental (amb intervenció) o observacional (sense intervenció), podem obtenir el text mitjançant observació participant o no participant, podem comparar subjectes o grups o no fer-ho, podem fer una investigació longitudinal

#### Vegeu també

Com que alguns d'aquests aspectes no són específics de l'anàlisi de contingut, no tocarem aquí aquests punts, que s'han tocat en altres assignatures del grau.

o transversal, etc. Tanmateix, res d'això és específic de l'anàlisi de contingut, sinó que forma part de la panòpia de recursos metodològics que un psicòleg ha de conèixer.

No obstant això, si consulteu alguns dels textos més coneguts sobre anàlisi de contingut veureu que, sovint, amb un probable afany corporativista, aquesta tècnica es presenta com una metodologia absoluta i dintre d'autèntics tractats en què es repassa la totalitat de la metodologia de les ciències socials i del comportament, com si fos específica o pròpia de l'anàlisi de contingut. Però una cosa és que certs procediments adquireixin una aparença particular quan s'apliquen dintre de l'anàlisi de contingut i una altra, ben diferent, és que aquests procediments hagin de presentar-se i tractar-se com a part constituent d'aquesta tècnica.

### Exemple d'aplicació de l'anàlisi de contingut

En un estudi de 2002, Karsenti i els seus col·laboradors van aplicar anàlisi de contingut convencional als materials obtinguts en tertúlies i correus electrònics d'estudiants que havien rebut un curs per Internet. Naturalment, la millora havia de ser detectada i valorada mitjançant un seguiment del procés d'aprenentatge, obligadament longitudinal. Les quatre categories centrals de l'anàlisi foren:

- "Curs" (missatges relacionats amb els continguts del curs),
- "Esculls" (missatges que comunicaven diverses dificultats de comprensió o execució),
- "Treballs" (missatges sobre les memòries o informes que els estudiants havien d'elaborar i lliurar per la xarxa) i
- "Diversos" (una categoria de calaix de sastre que agrupava missatges que no tenien una relació directa amb el curs).

L'anàlisi es va intensificar al voltant de la categoria "Esculls", per raons òbvies. L'evolució de les freqüències d'aquesta categoria i de les subcategories que la integraven ("Esculls tècnics", "Esculls d'autonomia", "Altres") va ser seguida al llarg de quinze setmanes i va revelar, almenys des d'aquest punt de vista, l'èxit del curs, ja que la seva disminució va ser dràstica a partir de la quarta setmana, tot i un petit repunt en la darrera.

És evident que un procediment de registre i avaluació com aquest s'hauria pogut utilitzar en un estudi transversal, comparant, per exemple, els missatges de diferents grups d'estudiants (nois i noies, posem per cas).

Actualment, l'anàlisi de contingut s'entén com una denominació prou àmplia que aplega totes les tècniques d'anàlisi textual quantitativa, aplicades a documents, transcripcions i fins i tot a missatges no estrictament textuals, però significatius, com els icònics (comunicació no verbal, pintura, còmic, cinema, publicitat, etc.). Sobre aquesta base quantitativa l'anàlisi de contingut pot fonamentar, orientar i perllongar les inferències que l'investigador qualitatiu fa en estudiar els textos. Per tant, l'anàlisi de contingut no és, en rigor, una tècnica purament qualitativa, però pot adoptar orientacions qualitatives i pot també posar-se al servei de moltes investigacions qualitatives, i és el complement de tècniques de registre i codificació com les que hem explicat fins aquí.

#### Lectura recomanada

T. Karsenti, F. Larose, i M. Núñez (2002). La apertura universitaria a los espacios de formación virtual: Un reto a la autonomía estudiantil. *Revista Electrónica de Investigación Educativa*, 4 (1), 1-18.

#### Lectura recomanada

G. Pérez Serrano (1994). *Investigación cualitativa: retos e interrogantes. II, Técnicas y análisis de datos* (pp. 133-134). Madrid: La Muralla.

Mirant les coses així, el debat sobre si l'anàlisi de contingut és una tècnica quantitativa o qualitativa deixa de tenir sentit: tot depèn de quins dels seus aspectes adquireixen més pes en l'anàlisi i quines són les vies de desenvolupament preferides i potenciades.

De vegades, l'anàlisi de contingut té un valor totalment quantitatiu, en tant que tècnica marcada per la voluntat d'objectivitat (en l'accepció més positivista d'aquest terme), centrada en els nombres, en la tabulació del material textual, poc disposada als relativismes, poc atenta als significats o als missatges. Aquesta orientació és extensiva i s'ocupa de grans volums de material textual. Justament, una de les aportacions de la tècnica és la reducció o simplificació de tot aquest material mitjançant variables molt generals, abans de computar freqüències i probabilitats d'aparició i calcular correlacions entre aquestes.

Altres vegades l'anàlisi de contingut revela una intenció del tot contrària, en tant que tècnica genuïnament qualitativa, atenta als matisos de la significació, a la complexitat, profunditat i caràcter relatiu dels missatges textuais, prop, llavors, de la teoria fonamentada, la fenomenologia, l'etnometodologia o l'anàlisi del discurs. Llavors no és tan important la freqüència o altres paràmetres amb els quals apareix un element en un text, com el fet que aparegui o no aparegui en comparació d'altres textos. És una immersió intensiva, en profunditat, en el text, atenta al detall i a la complexitat, al text i al context, i tant a allò que apareix explícitament com a allò que no es fa palès. Aquesta mena d'aproximació exigeix, sovint, que l'anàlisi es limiti a un o pocs casos, de manera que l'anàlisi de contingut qualitativa acaba essent una eina més en el si dels estudis de cas.

### Proposta purista de Philip Mayring

La posició adoptada aquí és bastant eclèctica respecte a la possibilitat de fer anàlisis de contingut qualitatives amb components quantitativs, o anàlisis essencialment quantitatives però amb elements qualitativs. No obstant això, existeixen posicions més *puristes* en relació amb la versemblança d'anàlisis de contingut genuïnament qualitatives, malgrat que ni tan sols aquestes posicions exclourien totalment els recursos quantitativs. Una d'aquestes és la proposta de Philip Mayring, en la qual l'obertura de l'anàlisi es fa compatible amb una certa canalització de les operacions realitzades, gràcies a la ruta teòrica seguida per l'investigador. Aquesta aposta per la combinació entre obertura i teoria, ensems amb una curiosa aplicació seqüencial dels protocols, col·loca aquesta proposta a prop de la teoria fonamentada que hem manejat en el mòdul "Les tècniques qualitatives en la selecció i tractament de la informació", tenint en compte, a més, que la construcció teòrica té el seu nucli generatriu en sistemes de categories com els que també hem vist. Ara bé, Mayring posa sempre un èmfasi molt especial en el caràcter molt sistemàtic i rigorós dels protocols aplicats, els quals, segons ell, farien qualsevol d'aquestes anàlisis replicables, per bé que radicalment qualitatives.

Per altra banda, tot i que les anàlisis no s'haurien de limitar exclusivament a l'estudi de casos, gosem dir que aquest és el terreny en què funciona millor la metodologia de Mayring. Altres trets fonamentals de l'enfocament qualitatiu són igualment previstos per Mayring, que entén el missatge textual en la dimensió comunicativa, en la qual el subjecte és l'emissor o autor i, per consegüent, cal arribar-hi i entendre'l per a interpretar el seu text.

### Lectures recomanades

P. Mayring (2000). Qualitative content analysis [article en xarxa]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1 (2), juny, art. 20. [Data de consulta: 5 d'octubre de 2004]: <<http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-e.htm>>

També podeu consultar: F. Kohlbacher (2006, gener). The Use of Qualitative Content Analysis in Case Study Research [article en xarxa]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 7 (1), art. 21. [Data de consulta: 10 de gener de 2009]: <<http://www.qualitative-research.net/index.php/fqs/article/view/75/154>>

### 1.1.2. Definicions de l'anàlisi de contingut

Una definició prou general de l'anàlisi de contingut, que permet encabir-hi perspectives diferents, podria ser la de Weber.

"L'anàlisi de contingut és una metodologia de recerca que utilitza un conjunt de tècniques per a fer inferències vàlides a partir dels textos."

R. P. Weber (1990). *Basic Content Analysis* (p. 9).

Tanmateix, aquesta mena de definicions són tan àmplies que manquen de substància. És el cas de la que proposen Penalva i Mateo.

"L'anàlisi de contingut és una tècnica d'anàlisi textual que s'utilitza per a observar de manera indirecta una societat. Com a tècnica basada en la documentació, aquesta observació no es du a terme ni mitjançant l'observació del comportament dels subjectes, ni tampoc conversant amb ells, sinó amb l'estudi dels textos que produeix aquesta societat."

C. Penalva i M. Á. Mateo (2006). *Tècniques qualitatives d'investigació*. Alacant: Universitat d'Alacant, Departament de Sociologia II, Psicologia, Comunicació i Didàctica / Universitat d'Alacant, Secretariat de Promoció del Valencià / Conselleria de Cultura, Educació i Ciència.

La definició de Penalva i Mateo sembla més completa però, com està confegida amb una òptica sociològica, hauríem d'afegir a *societat* els termes *persona*, *grup* o *institució*. Noteu el qualificatiu de *indirecta* i la referència a la documentació. Efectivament, en situar-nos respecte a l'anàlisi de contingut, hauríem de recuperar conceptes com el d'*observació indirecta*, *document*, *transcripció* i les tipologies associades a aquests.

#### Context

Caldrà que matisem un altre punt, que completa les definicions anteriors de l'anàlisi de continguts: aquestes observacions i inferències, sigui quin sigui el prisma a través del qual es faci l'anàlisi, es fan per mitjà del context; millor dit: dels distints tipus de context, intratextual o extratextual, tal com aviat confiem aclarir.

### 1.1.3. Tipus d'anàlisi de contingut

Ja coneixem les línies principals d'aplicació de l'anàlisi de contingut. Ara distingirem les orientacions d'aquesta tècnica en un marc menys general i considerant no tant les seves possibilitats teòriques com els desenvolupaments més freqüents i coneguts en la pràctica investigadora.

- **L'anàlisi temàtica.** És l'orientació més immediata de l'anàlisi de contingut a partir de definicions bàsiques com les que hem exposat més amunt, centrada precisament (valgui el joc de paraules) en el *contingut*. L'analista en aquests casos s'ocupa del significats denotatius (significat nuclear o *de diccionari*) o connotatius (associacions, significació perifèrica i més personal) del text. Dintre de l'anàlisi temàtica trobem dues modalitats principals:
  - L'anàlisi categorial que, després de classificar categorialment els materials textuais, en computa la freqüència sobre la base de la hipòtesi asenyada que diu que la freqüència amb la qual apareix un cert tema,

#### Vegeu també

Recordeu que el concepte d'observació indirecta es tracta en l'assignatura *Introducció als mètodes d'investigació en psicologia*. Els conceptes de document i transcripció es tracten en l'assignatura *Mètodes d'investigació qualitativa*.

#### Vegeu també

En l'assignatura *Mètodes d'investigació qualitativa* ja es van presentar les línies principals d'aplicació de l'anàlisi de contingut.

#### Lectura recomanada

Seguirem la classificació força ajustada als nostres propòsits i necessitats que proposen R. Quivy i L. van Campenhoudt (1997). *Manual de recerca en ciències socials* (p. 222 i seg.). Barcelona: Herder.



referència, al·lusió, etc., al llarg del text, es correlaciona d'alguna manera amb la importància que posseeix dintre del sistema de denotacions de l'autor.

- L'anàlisi avaluativa, que perllonga la línia d'anàlisi categorial en la direcció de valorar, no solament temes o referències, sinó els judicis i opinions de l'autor, en clau connotativa a més de denotativa, i tracta d'estimar la intensitat dels continguts a més de la seva presència en el text.
- **L'anàlisi formal.** Adreçada al *com* més que al *què* dels missatges textuais, en la seva aparença expressiva o comunicativa, i tant a escala de les unitats com del seu encadenament. Les orientacions principals són:
  - L'anàlisi de l'expressió, que s'ocupa de la forma del text i del discurs que hi és incorporat. La forma proporciona indicacions globals sobre les característiques de l'autor (des del seu tarannà o estat d'ànim fins a la seva ideologia). L'analista no faria sinó aplicar la vella dita que diu "l'estil és l'home" (o, evidentment, "la dona"), i els indicadors concrets en els quals recolzaria serien l'elecció del vocabulari (més o menys culte, més o menys directe, amb més o menys adjectius, etc.), el grau de complicació de les frases (amb més o menys punts i a part, amb més o menys oracions coordinades o subordinades), els recursos retòrics (interrogants, admiracions), i així successivament.
  - L'anàlisi de l'enunciació, que s'ocupa de la forma del text i del discurs que incorpora en si mateixos, en tant que posseïdors d'una dinàmica pròpia fins a cert punt independitzada de l'estat i trets del seu autor i més subjecta a les convencions culturals i a les regles socials del discurs. Ara l'analista es fixa més en l'ordre general del text (per exemple, en l'organització ritual del discurs d'un polític), en els seus retrocessos i avenços, en les repeticions, canvis de ritme, etc.
- **L'anàlisi estructural.** Posa tot l'esforç a descobrir les regles de construcció del text o del discurs, regles sovint no manifestes. Per tant, aquest camp de l'anàlisi de contingut està totalment compromès en la cerca de relacions, en dues direccions diferents:
  - La de l'anàlisi estructural pròpiament dita, abocada al descobriment dels principis generadors del text, en profunditat, ja sigui des del punt de vista de l'autor (com ha organitzat la tasca de producció del text), o des del punt de vista de l'interpret (com s'ha de representar o "modelitzar" el text per a fer-lo plenament intel·ligible).
  - La de l'anàlisi de contingències, també anomenada *d'associacions o de concomitàncies*. Aquesta, d'entrada, és una anàlisi més superficial que l'anterior, almenys en la seva primera fase, ja que en fases posteriors pot conduir a aprofundiments i a la localització dels principis organitzadors del text. És una anàlisi eminentment seqüencial i consisteix

a relacionar cada unitat d'anàlisi amb elements anteriors o conseqüents, a diferents distàncies dins del text. Com aquesta línia de treball disposa de recursos tècnics força estandarditzats, això ens permet concretar i dedicar més endavant una part d'aquest mòdul a exposar-los amb una mica de detall.

Sovint en els manuals se sol diferenciar entre les anàlisis que es conformen amb el significat *manifest* o evident i les que pretenen cercar-lo en capes més profundes del discurs, en tant que significat latent o ocult. Però aquestes dues opcions s'ofereixen transversalment en totes les modalitats descrites, tot i que, certament, l'accés a les capes generatrius o més profundes de significat està sempre més garantit en aquelles anàlisis les inferències de les quals recolzen en la cerca i establiment de relacions.

Malgrat que és una tècnica originalment i essencialment quantitativa, l'anàlisi de contingut es pot posar al servei de la investigació qualitativa.

## 1.2. El mostratge en l'anàlisi de contingut

La primera operació en l'anàlisi de contingut consisteix en l'establiment d'un corpus de textos i, dintre d'aquest, en l'elecció de la mostra que serà estudiada efectivament.

Una anàlisi de contingut comença, com no podria ser d'altra manera, amb una delimitació de l'objecte d'estudi i amb la selecció i fixació dels materials que es pretenen analitzar. En el llenguatge típic de l'anàlisi de contingut es diu que aquests materials seleccionats poden haver estat recollits mitjançant:

- recopilació de documents ja existents,
- la producció de textos o transcripcions *ad hoc*.

Aquesta distinció coincideix amb la que s'aplica, com sabeu, als documents, que separa aquells que són creats espontàniament dels que són creats a demanda.

Un cop feta aquesta primera selecció de textos, podem optar per estudiar-los totalment, o bé limitar la nostra indagació solament a una part de cada un. Si preferim aquesta segona opció hem de dur a terme operacions de mostratge sobre el material recollit en primer terme. Els criteris per a establir quines parts del text s'analitzen efectivament i quines no poden ser de diverses classes: estadístics, teòrics, pràctics, normatius, etc.

### Vegeu també

A continuació presentarem els trets bàsics d'aquesta tècnica, entesa com una via d'accés als textos, a la seva preparació i anàlisi, complementària a les que ja hem detallat en el mòdul anterior: algunes facetes del mostratge, segmentació i categorització exposats en el si de la teoria fonamentada són compartits per l'anàlisi de contingut sota la seva accepció més general; altres seran afegides ara. Recordem, per altra banda, que l'anàlisi de contingut ja va ser presentada i caracteritzada globalment en l'assignatura *Mètodes d'investigació qualitativa*. El que farem aquí és ampliar el que es deia allà, desenvolupant els aspectes més específics de l'anàlisi. Organitzarem els subapartats següents sobre la base d'un esquema semblant al que ens ha servit de guia fins aquí: abordarem primer el mostratge de materials textuals, continuarem amb la segmentació i la categorització dels documents i transcripcions analitzats i acabarem amb una breu incursió en les mesures i índexs útils en aquest àmbit metodològic.

### Vegeu també

En relació amb la tipologia dels documents, consulteu el mòdul 5 de l'assignatura *Mètodes d'investigació qualitativa*.

## Observació discontinua d'un text

De la mateixa manera que, en una sessió d'observació, el comportament d'un subjecte no ha d'observar-se i registrar-se necessàriament d'una manera contínua, en tot moment o en tota unitat temporal de registre, sinó que sovint es comptabilitza només en determinats intervals o punts de temps prèviament fixats, també un text pot ser "observat" de manera discontinua, centrant el registre en segments privilegiats i prèviament escollits en funció de diferents tipus de criteris. Amb tot, el paral·lelisme entre l'observació del comportament i la de textos té innegables limitacions, a les quals al·ludirem més endavant.

Sembla justificat referir-se a dos tipus d'operacions de mostratge textual que estableixen, al final, les fonts de dades pròpiament dites: per un costat, l'investigador de textos té la possibilitat de fer un **mostratge intertextual**, que comporta la selecció dels escrits dels quals s'ocuparà la investigació en curs; per un altre, pot o no dur a terme un **mostratge intratextual**, que fixa les parts dels textos ja escollits que seran registrades, categoritzades, mesurades i analitzades efectivament, de manera que les segrega d'aquelles que seran deixades de banda.

### 1.2.1. Mostratge intertextual

Quan és necessària, la selecció dels documents que entren dins de l'objecte d'estudi depèn evidentment d'aquest i dels objectius que es marca la recerca en curs. Els interessos i metes de l'investigador acoten l'univers documental en què ha de trobar el seu rumb: les cartes d'un polític, els relats dels participants en una crisi familiar, les entrevistes fetes per a elaborar una història de vida, etc. I els documents separats i escollits dintre d'aquest univers han de ser pertinents o rellevants amb vista a les preguntes que aquell investigador es vol respondre.

El primer que cal fer és establir contacte amb aquest univers documental. Això vol dir que s'han de llegir els textos, probablement no una, sinó diverses vegades. Si els documents han de ser produïts, i no simplement recopilats, la primera feina de l'investigador és, és clar, construir instruments que generin les respostes adients per a obtenir el material buscat. Els més habituals, com ja sabem, són els protocols d'entrevista.

#### Llibertat de selecció dels documents

No és inusual que els documents utilitzats estiguin assenyalats pels objectius mateixos de la recerca o pels patrocinadors d'aquesta. Aquesta és la situació normal en investigació aplicada. Per exemple, una empresa de perfumeria encarrega a un equip d'analistes la comparació entre els textos de la seva publicitat d'una colònia amb els corresponents de la competència. L'encàrrec és provocat per una davallada de vendes. En casos així, òbviament, el científic o el professional no poden escollir els seus materials.

Quan els plantejaments són més amplis, el marge de llibertat –i d'incertesa– és més gran. Suposem que ens demanen quins són els eixos de transformació del discurs oficial sobre educació de la joventut durant els darrers anys. Les possibilitats quant al material disponible són enormes: podem recórrer a decrets i notes de premsa del ministeri o la conselleria, a discursos del ministre o del conseller, als discursos en els repartiments de títols o premis, a articles de diaris i revistes, als llibres de text, etc.

Després de llegir i conèixer els tipus de documents amb els quals cal treballar, ja es pot fer la selecció definitiva sobre la qual es farà de veritat l'anàlisi. Aquesta selecció definitiva donarà lloc a un corpus.

El **corpus** és el conjunt de materials textuais que té en compte el científic social per a respondre la pregunta d'investigació a la qual s'enfronta, els escrits que generen la informació que després aquell processa.

Un cop decidits els criteris selectius que determinen un corpus, el nombre d'elements que aquest conté ha de ser màxim respecte als criteris mencionats. En altres paraules: el corpus ha de ser exhaustiu en el sentit que cap dels elements previsibles sobre la base dels criteris de selecció pot faltar-hi.

Parlem finalment d'aquests criteris de selecció. La mostra de textos final ha de ser representativa respecte a l'univers del qual prové. Aquesta representativitat apunta al nombre de documents que hi haurà en el corpus i que ha de cobrir la variabilitat que aquest exhibeix en diversos eixos.

En alguns casos, però no en tots, la selecció de materials es pot resoldre mitjançant criteris estadístics, els utilitzats en la recerca nomotètica o poblacional. Si coneixem la distribució dels elements de la mostra respecte a alguna variable, podem fer un mostratge a l'atzar, estratificat, per quotes, etc.

Altres vegades, quan l'horitzó de la recerca és qualitatiu, els principis que guien la selecció són de caire més teòric que estadístic, o estan soldats als objectius mateixos de la recerca. Així, si dins d'un epistolari ens interessen només les missives escrites durant les festes de Nadal, les destinades a certes persones, o les que toquen un tema concret, establirem el corpus sobre aquesta base, sense complicacions ulteriors; i si d'una entrevista a un polític tan sols ens interessem per les preguntes i respostes lligades a temes econòmics, llavors ens cenyirem a aquestes.

### **Exemple de mostratge**

No seria impossible, posem per cas, conèixer el nombre d'anuncis de cotxes en els diaris de Catalunya durant el 2009, classificats per la seva grandària, el nom de la publicació, el mes d'aparició, la marca i model del cotxe, etc.; i, tenint en compte totes aquestes fonts de variació, agafar-ne una mostra ajustada a una distribució normal.

Finalment, en un tercer tipus de situació de partida, la representativitat del corpus queda assegurada pel simple fet que univers i corpus coincideixen o –el que és el mateix– perquè l'investigador estudia tots els documents disponibles relacionats amb el tema que l'ocupa. Aquesta solució és característica de la recerca idiogràfica o de l'estudi de casos.

### 1.2.2. Mostratge intratextual

D'antuvi, el text fixat en el corpus ha de ser preparat. Aquesta preparació no és estrictament una operació de mostratge, però la fa possible o la facilita. La preparació ha de ser material i formal.

- La **preparació material** pretén "posar en net" el text quan aquest no reuneix les condicions de claredat o llegibilitat que se li poden exigir. Una carta escrita amb lletra difícil pot ser mecanografiada; el mateix que un autoinforme o un relat manuscrits. Els documents antics solen demanar una preparació considerable abans del desxiframent i l'anàlisi. Les transcripcions d'entrevistes o interaccions gravades en magnetòfon o vídeo poden també incloure's dintre de les operacions de "neteja" a les quals ens referim.
- La **preparació formal** al·ludeix a l'edició dels textos. Així, per comoditat de l'analista, els paràgrafs es poden separar, línia per línia, desglossant les oracions completes en els seus components i numerant-los; o es poden fixar marges molt amples per a codificacions posteriors; o es pot marcar el text de diverses maneres. No cal dir que, avui dia, aquesta preparació es pot fer i es fa per mitjà dels recursos que subministren els programaris d'anàlisi qualitativa.

#### Vegeu també

Sobre els programaris d'anàlisi qualitativa vegeu l'annex 2 del mòdul "Annexos".

Pel que fa al mostratge intratextual pròpiament dit, el principi orientatiu que cal seguir és el que diu que, un cop delimitat el corpus, com més segments de text analitzem dintre d'aquest, molt més valuosos seran els resultats. De fet, l'ideal és estudiar la totalitat de cada document, atès que un text és semànticament coherent però no homogeni al llarg del seu curs, per la qual cosa la selecció de fragments representatius pot ser problemàtica i, per descomptat, no sempre pot obeir a criteris estadístics.

Si, tanmateix, cal fer un mostratge intratextual, l'analista ha de tenir molt present l'estructura del text o el que ja en conegui. El text, sens dubte, exhibeix una estructura.

Un discurs presidencial de presa de possessió mostra una organització ritual: hi ha una introducció i una cloenda ben marcades amb fórmules tradicionals; els afers interns i els econòmics es tracten abans que els internacionals, etc. Una autobiografia sol consagrar més pàgines a la vida adulta que a la infantesa. Un relat palesarà una progressió temàtica molt dependent de la personalitat del seu autor i del tema sobre el qual s'estén: una persona compulsiva introduirà les qüestions centrals de seguida (sobretot si són conflictives), mentre que una de més tranquil·la tal vegada farà un llarg preàmbul de preparació.

Per consegüent, és prudent fer el mostratge dintre de cada sector d'aquesta estructura, que cal conèixer des de bon principi, almenys en una certa mesura. En un discurs, per exemple, cal fer una selecció de fragments en la introducció, una altra en el cos central de l'exposició i una altra en la cloenda.

A més, els fragments escollits han de posseir una certa unitat: la d'un paràgraf o la d'una secció dintre d'un capítol, posem per cas. En cap circumstància és convenient separar oracions o mots aïllats, deslligant-los del seu context, amb la qual cosa la interpretació esdevé difícil, si no impossible.

Semblantment al mostratge intertextual, també en l'intratextual podem recórrer a diferents tàctiques per a fer la selecció final de segments de text:

- **Podem seguir criteris estadístics.** En aquest tipus de mostratge, com el conjunt de textos seleccionats ha de permetre un repertori i una classificació exhaustius de segments de registre i de categories d'anàlisi, l'investigador es pot servir de tècniques habitualment emprades en l'observació del comportament. Efectivament, en aquesta darrera es pot estimar la quantitat d'observacions necessàries perquè el nombre d'unitats de registre i de categories mesurades pugui considerar-se complet i no ampliable. Així, en una primera fase de l'anàlisi, un cop assolida una classificació categorial provisional a partir del repertori de segments de text registrats, podem estimar si hem seleccionat prou material dintre del text en qüestió (prou línies, paràgrafs, pàgines, etc.) amb la fórmula

$$\theta = 1 - \frac{N_1}{T}$$

en què  $N_1$  és el nombre de categories detectades solament una vegada fins al moment de la decisió i  $T$  el nombre de segments o unitats de registre recollides també fins aquest mateix moment. Quan l'expressió pren valors propers a la unitat considerem que el nombre de segments seleccionats per a l'anàlisi ja és suficient.

- **Podem seguir criteris teòrics o pràctics,** derivats de la naturalesa mateixa de l'objecte d'estudi o de la demanda que ha motivat la recerca, tal com ja hem fet notar més amunt.

### **Efectes de final de pàgina**

Suposem aquesta indagació. En els àlbums de còmics clàssics es nota sovint que, originalment, els textos foren lliurats a publicacions periòdiques, pàgina per pàgina o tira per tira. El principal indici d'això en els àlbums publicats complets amb posterioritat es troba en els anomenats *efectes de final de pàgina* o de *final de tira*, recursos textuais o visuals amb els quals els autors pretenien –i aconseguien– mantenir l'interès del lector fins l'endemà, la setmana següent o el dia en què tocava el lliurament subsegüent de material. Es parla també, en relació amb aquests recursos, de l'efecte "continuarà".

Doncs bé, imagineu que algú vol estudiar a fons els efectes de final de pàgina. Òbviament, la selecció de parts del text –en aquest cas, vinyetes– obeirà a aquest objectiu; l'estudi no es fa en tota la superfície de la historieta sinó, com a màxim, en les dues o tres vinyetes finals de cada pàgina. De manera que les operacions de mostratge intratextual no requeriran altra consideració fora d'aquesta –atendre al final de pàgina–, coincident amb els objectius de la recerca.

Vegeu C. Riba (1986). Sobre la utilitzación de la página en Tintín. *Neuróptica*, 4, 100-109.

- Finalment, com ja hem dit, l'investigador pot estudiar la totalitat de cada un dels textos seleccionats, opció practicable si aquests no són molts ni molt extensos. En qualsevol cas aquesta solució, que era típica de les apro-

ximacions idiogràfiques i qualitatives als textos en el terreny del mostratge intertextual, també ho és en el del mostratge intratextual.

Els criteris de mostratge inter- i intratextual poden ser estadístics, teòrics o derivats dels objectius de la recerca.

### **1.3. Principis de segmentació i creació d'unitats en l'anàlisi de contingut: text i context**

La segmentació inherent a l'anàlisi de contingut preveu l'aïllament tant d'unitats textuais com contextuals, ja que les primeres adquireixen sentit dintre de les segones.

#### **1.3.1. Text i context**

Una vegada disposem d'un corpus i d'una selecció de textos dins d'aquest, els principis que guien la diferenciació sistemàtica de segments de text i la creació d'unitats textuais no són molt diferents, en l'anàlisi de contingut, dels que hem mencionat i comentat a propòsit de la codificació en general de qualsevol text, ni tan sols dels principis específics aplicats en la teoria fonamentada. Tanmateix, atesa la subjecció major a protocols de l'anàlisi de contingut, i tenint en compte que, per la seva filiació quantitativa, en aquesta la subjectivitat i creativitat de l'analista poden restar força neutralitzades, els procediments de segmentació estan molt més sotmesos a norma. I, per altra banda, com sol passar en les ciències socials, el dialecte especialitzat d'aquesta tècnica ha introduït termes que no sempre són coincidents amb el que hem vist fins aquí.

Per totes aquestes raons, hem de dedicar unes pàgines a fixar l'accepció d'alguns termes aplicats a la segmentació en l'anàlisi de contingut, els quals no es contradiuen amb els que ja hem fixat en el mòdul anterior, ans els completen amb exigències d'estandardització. Ara bé, com que l'anàlisi de contingut, a pesar de ser una tècnica essencialment quantitativa, pretén fer inferències i interpretacions textuais, no pot deixar aquestes a l'arbitri de l'analista, sinó que tracta d'establir un patró o esquema primari sobre el qual es poden sostenir aquestes inferències i interpretacions.

Aquest patró o esquema es configura sobre la **relació text-context**. Un segment de text és interpretat en funció d'un segment de context corresponent, que és el que li atorga significat relacional. I cada segment de text s'assigna a un o més segments de context segons les regles d'adscripció, que solen ser de tipus metodològic, no teòric (a diferència de la tàctica dominant en la teoria fonamentada). La regla d'assignació sol coincidir, de manera més o menys exacta, amb la de la interpretació contextual clàssica, en la qual cada segment o unitat considerada s'enllaça, com hem vist, o bé amb un element antecedent

(referencial o causal), o bé amb un de consegüent (funcional o intencional). No obstant això, el context també pot ser d'altres tipus, no ancorat en elements anteriors o consegüents, sinó que pot englobar o abraçar l'element per interpretar.

Aquest esquema interpretatiu es pot aplicar tant a seqüències d'accions com a seqüències de text i cal aclarir igualment que, així com en tot el que hem dit fins ara estava implícit que el context pertanyia a la mateixa seqüència que l'element interpretat, també pot esdevenir-se que resti fora d'aquesta. I, finalment, com de seguida precisarem, el context també pot estar fora del text.

En qualsevol cas, la segmentació en l'anàlisi de contingut reclama dos tipus de criteri, un de referit als segments de text, i un altre als segments de context, per bé que els dos tipus de segments són associats o acoblats. Ens caldran **unitats de text**, les unitats de registre pròpiament dites en tant que peces o elements del text; i també **unitats de context**.

### Exemple de contextualització

Incloem un primer exemple de contextualització, com a aperitiu dels que veurem després. L'hem pres d'E. López-Aranguren (1989). El análisis de contenido. A M. García Ferrando, J. Ibáñez i F. Alvira (Comp.), *El análisis de la realidad social. Métodos y técnicas de investigación* (2a. ed., pp. 383-414). Madrid: Alianza.

Considerem aquesta notícia traduïda, apareguda al *New York Times* en temps de la dictadura franquista, el 10 de maig de 1970 (ometem la capçalera):

"El Gobierno anunció formalmente hoy la destitución del jefe de la Escuela Militar de Estado Mayor, que había hablado públicamente de la injusticia social en España.

La orden de destitución del General de Brigada Narciso Ariza García fue firmada por el Generalísimo Franco y publicada en el *Boletín Oficial del Estado*. La orden no daba razón alguna para la destitución.

Sin embargo, fuentes militares han dicho que el general ha sido cesado por su crítica del sistema social español en un discurso pronunciado en la Escuela el pasado domingo."

En aquesta notícia podem distingir dos nivells de contextualització al voltant del nucli de la notícia, que és l'element per interpretar: la destitució d'un general. Prenent com a unitat d'anàlisi el paràgraf, és obvi que, globalment, el tercer paràgraf fa de context consegüent dels altres dos, rebla l'argument introduït en el primer i interpreta explícitament el sentit de la destitució. Ara bé, si prenem com a unitat l'oració (entre comes o punts), tenim tres contextos consegüents, un per paràgraf. Dos estan marcats per la interpretació del fet en termes de càstig, igual que en la interpretació anterior ("[...] que había hablado públicamente de la injusticia social en España", en el primer paràgraf; "[...] fuentes militares han dicho que el general ha sido cesado por su crítica del sistema social español en un discurso pronunciado en la Escuela el pasado domingo", en el tercer). Per contra, el context consegüent, "La orden no daba razón alguna para la destitución", és un context nul, que només permet constatar la manca d'informació, però que queda compensat pels altres dos.

### 1.3.2. Unitats textuales

Les unitats textuales corresponen als segments i formen el nivell més elemental de referència de l'anàlisi.

#### Vegeu també

En relació amb la interpretació contextual, i a tall de recordatori, us remetem al mòdul 2, subapartat 2.3.2, dels materials de l'assignatura *Mètodes d'investigació qualitativa*.



És clar que el límit inferior de la seva grandària és el mot o el monema, límit més enllà del qual hi ha la pèrdua de significació. És a dir, adduint un argument de rellevància o pertinència, una anàlisi de contingut de caire qualitativa no pot prescindir de la significació i, per tant, el mot és el sòl des del qual ha de fer remuntar qualsevol interpretació. Més enllà d'aquesta consideració, les característiques i grandària de les unitats textuais depenen, com és preceptiu, dels objectius de la recerca i del nivell d'anàlisi que interressi adoptar.

Les unitats textuais han de ser perfectament discernibles en tant que segments d'escriptura i han de poder relacionar-se entre si. Les més utilitzades són:

- **Mot.** És la peça mínima de significat. Com a unitat es distingeix per ser una forma gràfica (o sonora, abans de la transcripció), simple o composta. El mot és fàcilment contextualitzable en frases, oracions, etc., i es presta al tractament informatitzat (des de fa anys els ordinadors són capaços de lliurar-nos llistes o recomptes de mots, catàlegs de les frases en què figura un mot, etc.). El problema sorgeix quan un mateix mot en tant que forma gràfica brinda diferents significats o diferents mots posseeixen el mateix; és a dir, quan existeixen homònims o sinònims. Per això el mot es pot tractar de dues maneres:
  - com a unitat **multívoca**, que presenta diferents valors o significats;
  - com a unitat **unívoca**, que està associada a un sol valor o significat.

#### **El mot com a unitat multívoca o unívoca**

El mot *taula* com a unitat multívoca podria proporcionar dues unitats de registre diferents, segons que el seu significat fos 'moble amb potes que suporten una superfície horitzontal i que serveix per a menjar-hi, escriure, etc.', o bé 'format gràfic de classificació de doble entrada, amb files i columnes'. Com a unitat unívoca, en canvi, només en proporcionaria una.

- **Frases amb unitat semàntica**, com les frases fetes (*caure molt baix, donar per suposat*) i els refranys i dites.
- **Oracions.** De qualsevol tipus, simples, coordinades, subordinades; actives o passives. Els criteris de delimitació de l'oració han de ser, inevitablement, sintàctics o gramaticals. Probablement, després del mot aquesta és la unitat de registre més emprada.
- **Temes.** Un tema pot referir-se a fets, situacions, estats d'ànim; pot ser d'abast individual o col·lectiu. Aquesta unitat, a diferència de l'anterior, és més semàntica que gramatical i, per tant, de delimitació més complicada. En general, però, és possible fixar en un text els límits del tractament d'un tema.

#### **Exemple de tema**

Podem dir que en una història de vida l'autor comença a parlar de temes centrats en la seva salut a la línia 4 de la pàgina 5 i acaba a la línia 29 de la pàgina 6; torna a mencionar-los a la línia 31 de la pàgina 11 i deixa de parlar-ne a la línia 4 de la pàgina 13; i així successivament.

- **Paràgrafs.** Aquesta és una unitat formal, encara que també té connotacions semàntiques, ja que un canvi de paràgraf sol correspondre a un canvi de tema o de matís temàtic. Redueix l'esforç de registre, en ser un segment de text generalment gran. En contrapartida, la tasca de categorització de paràgrafs sol ser més arriscada.
- **Text sencer.** Aquesta solució és practicable si el corpus sencer és escàs, com quan s'analitzen les capçaleres de diaris, grafitis o anuncis. Llavors cada text pot ser vist com una unitat de registre, comparable amb les altres, i el conjunt del corpus pot veure's com el seu context de referència.

### 1.3.3. Unitats contextuais

La unitat contextual o context mínim és el menor segment de text necessari per a interpretar una unitat textual.

Les unitats textuais, per si mateixes, poden ser descrites, però no interpretades. Per a ser interpretades o per a atribuir-los un significat caldria posar-les també en context, en relació amb segments de nivell superior en l'entorn textual o fora d'aquest, perquè els contextos no han de pertànyer necessàriament al text. Si hi pertanyen, llavors la interpretació es fa dintre de la cadena de segments textuais, tal com hem indicat més amunt. En cas contrari, la interpretació ha de recórrer a l'entorn o les circumstàncies en les quals va ser fabricat o publicat un text. De manera que hem de distingir dos tipus d'unitats o marcs contextuais: contextos intratextuais (o textuais a seques) i contextos extratextuais (o no textuais).

#### Contextos intratextuais

Els contextos intratextuais són aquells que es vinculen a les unitats textuais dintre del text mateix seleccionat del corpus.

##### Exemple de context intratextual

El gruix d'una notícia de diari sobre la condemna a mort d'un dissident en un estat dictatorial pot contextualitzar-se en el paràgraf precedent, dedicat a explicar com aquesta persona va fer declaracions contra el règim, o en el següent, que descriuria les revoltes populars resultat de la sentència.

Integrem també aquí els anomenats de vegades *contextos intertextuais*, és a dir, aquells textos, no inclosos en el corpus, que podrien servir com a contextos d'interpretació del text analitzat. En qualsevol cas aquests contextos continu-

en essent textos, que és el que ara ens ocupa, ja que la possibilitat alternativa, com de seguida mencionarem, és marxar de l'univers textual i traslladar-nos al món "real".

### **Exemple de context intertextual**

Una notícia d'un diari d'una cadena mediàtica entesa en el context d'un altre diari, més important, de la mateixa cadena mediàtica, però no analitzat en el mateix estudi, és un exemple de context intertextual.

Aquests contextos són, doncs, unitats inserides en l'entorn escrit, documental, que les conté. Poden ser antecedents o consegüents, en conformitat de la pauta global d'interpretació contextual a la qual ja hem al·ludit diverses vegades, que en l'argot de l'anàlisi de contingut es designa com a *unitat-mot-en-contexte*<sup>1</sup>.

<sup>(1)</sup>En anglès, *key-word-in-context* o KWIC.

### **Exemple**

En un registre de grup focal, cada intervenció representa un torn i està flanquejada per la intervenció anterior d'un altre membre del grup i per una altra de posterior, del mateix membre que havia intervingut abans o d'un altre. En una transcripció d'entrevista, el context precedent de cada resposta del subjecte pot ser la pregunta prèvia de l'entrevistador, mentre que el consegüent pot localitzar-se en la pregunta següent. En un diari personal, cada fragment corresponent a una data es troba entre el fragment de la data prèvia i el de la data posterior, tot i que també podríem contextualitzar paràgrafs o oracions de cada anotació diària entre els paràgrafs o oracions precedents i consegüents de la mateixa data.

Aquests contextos, per tant, tenen una amplitud variable. El context d'un mot pot ser el mot que el precedeix i el que el segueix; o els dos mots que el precedeixen i els dos que el segueixen; o els tres mots que el precedeixen i els tres que els segueixen; o les frases o oracions que el precedeixen i el segueixen, també en nombre variable. A propòsit del context d'una frase o d'una oració, podríem repetir el mateix exercici: una frase podria estar contextualitzada entre la frase antecedent i la consegüent; o entre les dues antecedents i les dues consegüents; etc.

De totes maneres, l'amplitud del context està en part determinada per la grandària de la unitat textual considerada. Hi ha una correspondència notòria entre l'escala en la qual es defineixen les unitats textuales i l'escala en la qual es defineixen les unitats de context intratextual. El context d'un mot (altres mots, una frase, una oració) tendeix a ser de mida menor que el d'una oració (altres oracions, paràgrafs), i el d'una oració tindrà segurament menys abast que el d'un paràgraf (pàgines, seccions).

Ultra això, la distància a la qual l'analista intèrpret pot situar el context també pot ser variable. Un cop establert un determinat nivell de segmentació, un segment en particular pot vincular-se amb segments situats immediatament després (o abans), o bé un segment més enllà, dos segments més enllà, etc.

### **Trastorn límit de personalitat**

Pareu compte en el text següent, que informa sobre una conferència al voltant del trastorn límit de personalitat, una malaltia que cada vegada acapara més l'atenció de psicòlegs clínics i psiquiatres, a causa de la gran incidència en la societat actual.

"El Turó va organitzar ahir una conferència per a conèixer el trastorn límit de la personalitat. Aquest trastorn afecta un 2% de la població i encara està poc investigat.

L'Associació de familiars i amics de persones amb malaltia mental "El Turó" va organitzar ahir una xerrada sobre el trastorn límit de la personalitat, un trastorn mental amb importants repercussions clíniques i socials que ha experimentat un augment de la incidència en la nostra societat aquests últims anys.

La xerrada anava a càrrec de Carme Sánchez Gil, psicòloga clínica dels Centres Assistencials Doctor Emili Mira, que treballa en el Programa "Trastorn límit de la personalitat" de Santa Coloma de Gramenet. Sánchez va explicar als assistents que el trastorn límit de la personalitat provoca en la persona malalta una inestabilitat que es manifesta en diversos àmbits de la seva conducta, el seu pensament i les seves relacions socials. Per a ells, les relacions són més intenses i tenen por de ser abandonats. Tendeixen a fer-se mal a ells mateixos i, fins i tot, al suïcidi.

El problema d'aquest tipus de trastorns és que afecten molta més gent de la que ens pensem, fins a arribar al 2% de la població. Això està provocant que cada cop es destinin més recursos a pal·liar els seus efectes en les persones que el pateixen. Les persones que pateixen un trastorn límit de la personalitat necessiten un tractament intensiu i de llarga durada, i també han d'estar a mans d'un equip pluridisciplinari."

Hem extret aquests paràgrafs, modificant lleugerament la distribució del text, d'una notícia presentada per Martínez Goretti (29 d'abril de 2008). El trastorn límit de personalitat afecta cada cop més persones. *RTVVilafranca.Cat, Diari Multimèdia de l'Alt Penedès*. [Data de consulta: 4 de maig de 2009]: <<http://www.rtvvilafranca.com/noticies-rtvvilafranca/societat/16385-el-trastorn-limit-de-la-personalitat-afecta-cada-cop-mes-persones.html>>

Suposem que ens interessa analitzar i, per tant, contextualitzar el mot *trastorn*. D'antuvi, la primera qüestió a considerar podria ser si el contextualitzem per oració, frase o unitat temàtica; o si ho fem per línia o per paràgraf. Escollim aquestes darreres opcions que, ara mateix, impliquen menys dificultats. Dintre dels paràgrafs distingim línies. Numerem les línies i paràgrafs mitjançant dos dígitos per a disposar de referències clares sobre la posició de cada element textual que pugui interessar (el primer dígit és el paràgraf i el segon la línia).

L'arranjament per paràgrafs i línies podria aparèixer així:

1.1	"El Turó va organitzar ahir una conferència per a conèixer el <b>trastorn</b> límit de la personalitat.
1.2	Aquest <b>trastorn</b> afecta un 2% de la població i encara està poc investigat.
2.3	L'Associació de familiars i amics de persones amb malaltia mental "El Turó" va organitzar ahir
2.4	una xerrada sobre el <b>trastorn</b> límit de la personalitat, un <b>trastorn</b> mental amb importants
2.5	repercussions clíniques i socials que ha experimentat un augment de la incidència en la nostra
2.6	societat aquests últims anys.
3.7	La xerrada anava a càrrec de Carme Sánchez Gil, psicòloga clínica dels Centres Assistencials
3.8	Doctor Emili Mira, que treballa en el Programa " <b>Trastorn</b> límit de la personalitat" de Santa Coloma
3.9	de Gramenet. Sánchez va explicar als assistents que el <b>trastorn</b> límit de la personalitat provoca
3.10	en la persona malalta una inestabilitat que es manifesta en diversos àmbits de la seva
3.11	conducta, el seu pensament i les seves relacions socials. Per a ells, les relacions són més

3.12	intenses i tenen por de ser abandonats. Tendeixen a fer-se mal a ells mateixos i, fins i tot, al
3.13	suïcidi.
4.14	El problema d'aquest tipus de <b>trastorns</b> és que afecten molta més gent de la que ens pensem,
4.15	fins a arribar al 2% de la població. Això està provocant que cada cop es destinin més recursos
4.16	a pal·liar els seus efectes en les persones que el pateixen. Les persones que pateixen un
4.17	<b>trastorn</b> límit de la personalitat necessiten un tractament intensiu i de llarga durada, i també
4.18	han d'estar a mans d'un equip pluridisciplinari".

Limitant primer la contextualització de *trastorn* a les línies on apareix:

1	"El Turó va organitzar ahir una conferència per a conèixer el <b>trastorn</b> límit de la personalitat.
2	Aquest <b>trastorn</b> afecta un 2% de la població i encara està poc investigat.
4	una xerrada sobre el <b>trastorn</b> límit de la personalitat, un <b>trastorn</b> mental amb importants
8	Doctor Emili Mira que treballa en el Programa " <b>Trastorn</b> límit de la personalitat" de Santa Coloma
9	de Gramenet. Sánchez va explicar als assistents que el <b>trastorn</b> límit de la personalitat provoca
14	El problema d'aquest tipus de <b>trastorns</b> és que afecten molta més gent de la que ens pensem,
17	<b>trastorn</b> límit de la personalitat necessiten un tractament intensiu i de llarga durada, i també

Lògicament, el fet de treballar a una escala o a l'altra determina l'abast del context considerat. Si ens servim d'un context tan global que exhaurixi tot el marc de referència, ja sigui la línia o el paràgraf, aleshores el context serà, precisament, la línia sencera o la frase sencera. Si recorrem a antecedents o conseqüents, amb una estratègia de tipus KWIC, hem de fixar el nombre d'unitats que utilitzarem com a context.

En el cas de treballar per línies podríem establir, posem per cas, tres mots per davant o per darrere, com a context, en el supòsit que la nostra unitat fos el mot. Així, en la primera línia, el context precedent de trastorn seria *a conèixer el* i el conseqüent seria *límit de la*; en la segona l'antecedent seria *aquest*, mentre que el conseqüent seria *afecta un 2%*; etc.

Posat ara que treballéssim la contextualització dins de cada paràgraf, podríem continuar definint el context mot per mot o bé línia per línia, a un altre nivell. Així, en el paràgraf 4, línia 17, el mot *trastorn* podria ser contextualitzat pels tres mots conseqüents (*límit de la*) però per cap precedent, ja que és al principi de la línia; però per altra banda podria contextualitzar-se entre la línia 16, del damunt (*a pal·liar els seus efectes en les persones que el pateixen. Les persones que pateixen un*), i la 17, de sota (*d'estar a mans d'un equip pluridisciplinari*), etc.

Aturem-nos una mica més en les contextualitzacions de tipus KWIC, canviant d'exemple. Vegeu una petita part d'una llista de contextualitzacions del mot *rights*, generada per ordinador, a partir de documents de les plataformes electorals del partit republicà l'any 1980, amb Reagan de candidat. Aquest és el mateix tipus de contextualització de l'exemple precedent, de nou l'esquema interpretatiu *unitat-mot-en-context*. Aquí, més precisament, es tractaria d'una unitat de registre en forma de *mot-entre-mots* ja que els contextos anteriors i conseqüents estan formats per 8 mots cada un, tot i que el mot inicial o final pot aparèixer tallat. En la selecció que hem fet de la llista, el mot *rights* és una unitat textual unívoca, atès que és portadora d'un sol significat (el corresponent a drets humans, dels ciutadans, etc.).

1	young people want the opportunity to exercise the	rights	and responsibilities of adults. The Republican Party
2	of their school systems. We will respect the	rights	of state and local authorities in the management
3	and that new laws with guarantees of individual	rights	are possible and will work. Republicans believe
4	it's rethoric. The most flagrant offenders of human	rights	including the Soviet Union will not be ignored
5	an economic security. Hispanics seek only the full	rights	of citizenship - in education or in law enforcement
6	t reaffirm our party's historic commitment to equal	rights	and equality for women. We also recall you
7	xemption from the military draft. We support equal	rights	and equal opportunities for women, without taking
8	their emigration is a fundamental affront to human	rights	and the decline in exit visas
9	policy must be based on the primacy of	rights	and responsibilities.
	...	...	...

Font: R. P. Weber (1990). *Basic content analysis*, p. 45.

Vegeu també C. Grbich (2007). *Qualitative data Analysis. An Introduction* (p. 116 i seg.). Los Angeles: Sage.

Les unitats de context, per tant, tenen un estatut relatiu dintre de la jerarquia d'escals possibles. Això permet que, canviant l'escala o el nivell d'anàlisi, una unitat de context intratextual pugui convertir-se en una unitat textual. Així, l'oració pot ser la unitat contextual del mot però, al seu torn, pot emprar-se també com a unitat de registre, enquadrada llavors en el paràgraf –posem per cas– com a unitat de context.

Aquests diferents nivells o grandàries de context atorguen matisos de significat també diferents a les unitats de text que engloben. No és el mateix assignar significat a un mot per la seva posició dins d'una oració que fer-ho a partir del seu valor dins de la totalitat d'un escrit.

### Aplicació del context intratextual

El context intratextual és el que prefereix l'estructuralisme i el que utilitzen les investigacions lingüístiques i semiòtiques i, en general, les aproximacions qualitatives.

En l'oració *jo mai no deixo de comprar el regal d'aniversari per a la meva mare*, el valor del mot *comprar* es fixa principalment en relació amb la destinatària de l'acció (la mare) i la condició de l'acció (l'aniversari). Es tracta d'una compra marcada pels vincles familiars i per afectes o obligacions. Però suposem que aquest mateix mot, dintre de la mateixa oració, hagués d'interpretar-se en el si del diari d'un addicte al consum. Llavors, és clar, la interpretació canviaria i en el context general del document el mot *comprar* adquiriria altres connotacions.

## Contextos extratextuals

Els contextos extratextuals són aquells als quals es vinculen les unitats textuals des de dintre a fora del text.

Els contextos extratextuals corresponen a les circumstàncies en què va ser conegut el text, a les seves condicions de producció. En aquestes darreres, es materialitzen els factors pragmàtics que incideixen sobre el text, el seu entorn comunicatiu i social, incloent-hi aspectes com la personalitat de l'autor i les vicissituds que sofria quan va escriure el document o quan va verbalitzar el discurs que s'analitza, la personalitat o el perfil del destinatari, les motivacions subjacents a la creació del document, el seu rerefons històric; etc.

### Aplicació del context extratextual

El context extratextual és el preferit pels estudis sociològics i per la crítica social i ideològica.

Els contextos extratextuals s'adiuen amb unitats textuals relativament molars o grans. Això és de sentit comú. En general, no sembla gaire prometedor relacionar l'ús d'un mot, per exemple, amb la crisi familiar enmig de la qual algú el va posar sobre el paper; sembla més fructífer lligar aquesta crisi a unitats temàtiques o a documents sencers. Tampoc semblaria gaire lògic voler interpretar una frase d'un personatge públic en el marc de la seva afiliació a un partit polític. Seria el text sencer el que s'adaptaria millor a un marc tan ampli.

### 1.3.4. Les unitats textuals de registre i la seva codificació

Recapitulant, direm que, normalment, una unitat textual de registre en l'anàlisi de contingut és una unitat de *text-en-context*, com la de *mot-entre-mots* que vèiem en l'exemple del subapartat 1.3.3. El context pertinent en perspectiva qualitativa és l'intratextual. Dit d'una altra manera, la unitat de registre és una unitat textual interpretada o en vies d'interpretació gràcies a la seva contextualització, ja sigui en contextos anteriors, en conseqüents, o en tots dos.

Aquestes unitats de registre enteses així són les que es codifiquen seguint un procés semblant al que hem il·lustrat en el mòdul "Les tècniques qualitatives en la selecció i tractament de la informació", i formen la base sobre la qual s'edifica el sistema de categories que permet prosseguir l'anàlisi de contingut en una determinada direcció.

Toca revisar ara, per tant, el concepte de categoria en l'àmbit de l'anàlisi de contingut.

### Exemple de codificació de les unitats textuais de registre

En l'exemple del subapartat 1.3.3 dels textos de les plataformes electorals de Reagan podem vincular el mot *rights* amb diversos marcs de referència: l'emigració, la dona, l'exèrcit, les lleis, el castrisme, i així successivament. Suposem que el vinculem amb mots que denotin igualtat –*equal, equality, etc.*–, interpretant llavors la referència als drets en termes d'igualtat, de les dones respecte als homes, dels emigrants respecte a la població autòctona, etc. Hi ha dues unitats de registre que poden ser codificades o precategoryades així, la 6 i la 7, ja que les claus contextuais esmentades hi són presents, tant en el context anterior com en el següent. En el registre podrien ser marcades amb una Q (d'*equal* o *equality*).

### Activitat

Llegiu primerament amb atenció el text següent, extret de C. Lévi-Strauss (1991), *El pensament salvatge*, Barcelona: Edicions 62 [traducció de M. Martí i Pol], centrat en la relació entre màgia i ciència.

"Això no obstant no regressedem a la tesi vulgar –i d'altra banda admissible en la perspectiva estreta en què se situa– segons la qual la **màgia** seria una forma tímida i balbucejant de la ciència: car hom es privaria de tot mitjà de comprendre el **pensament màgic**, si es pretenia de reduir-lo a un moment, o a una etapa, de l'evolució tècnica i científica. Més aviat l'ombra que anticipa el seu cos, és, en un sentit, completa com ell, tan acabada i coherent, en la seva immaterialitat, com l'ésser sòlid precedit només per ella. El **pensament màgic** no és un principi, un començament, un esbós, la part d'un tot encara no realitzat: forma un sistema ben articulat... En comptes, doncs, d'oposar **màgia** i ciència, valdria més de posar-los en paral·lel..."

A partir d'aquest text escolliu la unitat de registre "màgia" prenent com a context els 9 mots anteriors i posteriors; després escriviu els registres resultants. A continuació feu el mateix amb contextos de 8, 7 i 6 mots, abans i després. Heu de notar que la dificultat d'interpretació o de comprensió creix a mesura que reduïm el marc contextual. Repetiu el mateix exercici amb la unitat de tipus frase "el pensament màgic", prenent ara contextos anteriors i següents de 5, 4 i 3 mots.

### Orientacions i pistes per a la solució de l'activitat

En els dos casos podeu agafar com a exemple orientatiu el de Weber, exposat en el text. Considerant que el mot *màgia* surt dos cops en l'extret de Lévi-Strauss, lògicament us han de sortir també dues contextualitzacions per a cada un dels marcs contextuais de 9, 8 i 7 mots que sol·licitem.

En el cas d'"el pensament màgic" us han de sortir igualment dos registres per a cada una de les contextualitzacions de 5, 4 i 3 mots que feu.

Els contextos pertinents respecte a una anàlisi de contingut, poden ser intratextuals (o intertextuals) i extratextuals.

## 1.4. Les categories en l'anàlisi de contingut

A través del prisma de l'anàlisi de contingut, la categorització és sobretot un procediment per a reduir i sistematitzar el material textual escollit, operacions que tanmateix impliquen una interpretació.



### 1.4.1. La categorització sistemàtica de segments textuais

La informació ja segmentada i codificada (cada classe de segment amb el seu codi) d'un registre textual, en tant que recull unitats de text dintre del seu context, és heterogènia i complexa. Sobre aquesta diversitat o riquesa l'analista difícilment podria definir i mesurar variables, ni fer inferències sense risc d'equivocar-se. Per tal de salvar aquest obstacle, la informació en qüestió ha de ser classificada o, el que és el mateix, categoritzada a partir de les seves similituds i diferències. Ara, però, hem de redefinir aquesta operació amb una òptica més ajustada a les necessitats de l'anàlisi de contingut.

Segons aquesta òptica, la categorització no és tant un recurs flexible per a construir un model teòric d'allò que s'investiga i aprofundir o desvetllar significats, com un procediment que permet condensar el material textual i reduir-lo sistemàticament, alhora que fa les dades accessibles i tractables. Així, tal vegada perdrem informació de detall, a petita escala, però, en rescabament, guanyem capacitat de fer generalitzacions a més gran escala, acomodant la informació recollida en una espècie de taula amb les seves files i columnes, en un sistema de categories, en definitiva.

Dintre del caràcter sistemàtic i relativament tancat de les tècniques d'anàlisi de contingut, convé ara recordar i completar els conceptes relatius als sistemes de categories textuais.

Ja sabem que els materials verbals tenen una naturalesa substancialment diferent de la dels no verbals. Són materials que ja estan interpretats pel seu emissor o –també– pel seu receptor, abans que el científic n'assagi la interpretació.

- Una acció no verbal ha de ser definida verbalment, en un codi i en un medi que no són els que li pertocuen; no són els que modelen originalment l'acció en tant que moviment, canvi, etc.
- Per contra, una acció verbal o una frase ja presenten la seva definició o caracterització pròpia o –almenys– són definibles en el mateix llenguatge en què foren pronunciades o escrites.

Tot això, al costat d'altres factors que citarem, afegeix dificultats específiques a la categorització de textos, dificultats que solen tractar-se en el si de l'anàlisi de contingut, tot i que poden ser compartides per altres orientacions de l'anàlisi.

### 1.4.2. Criteris per a la creació de categories i la construcció de sistemes de categories en anàlisi de contingut

Ens trobem amb tres punts de partida ben diferenciats:

#### Vegeu també

Ja hem tractat la classificació de la informació en el mòdul "Les tècniques qualitatives en la selecció i tractament de la informació" d'aquesta assignatura.

#### Vegeu també

Els sistemes de categories textuais només han estat mencionats d'esquella manera en el mòdul "Les tècniques qualitatives en la selecció i tractament de la informació" d'aquesta assignatura, el qual estava principalment centrat en el procés de la codificació i no tant en el seu producte final: un sistema de categories.

**1) Existeix ja un sistema de categories o classificació adequat al material que es vol analitzar.** La seva aplicació correspon, per tant, a una estratègia *up-down* o hipoteticodeductiva. El sistema ha sorgit d'estudis previs sobre materials semblants als que ara es volen categoritzar. De la mateixa manera que no cal construir un test nou cada vegada que es vol mesurar un tret de personalitat, amb el benentès que ja n'hi ha de disponibles –i idonis– en el mercat, tampoc no cal bastir un sistema de categories quan ja n'hi ha d'adients a l'abast de l'investigador.

Sovint aquests sistemes ja consagrats han rebut una certa estandardització, amb l'avantatge que els torna molt rendibles quan es desitgen fer comparacions entre els resultats de diferents investigacions que se n'han servit.

### **Diccionari categorial**

Avui dia aquests protocols estandarditzats solen estar implementats en programes informàtics, i llavors es denominen *diccionaris* o *índexs*. Un diccionari categorial està organitzat com un tesaurus: les entrades es concentren en conceptes clau i, dintre de cada una, apareixen els mots o expressions relacionats d'alguna manera amb els conceptes.

**2) No existeix un sistema de categories adequat al material codificat.** Aleshores l'analista l'ha de construir, i s'ha de col·locar indefectiblement en una perspectiva analiticoinductiva.

**3) Una part del text ja està categoritzada, mentre que una altra no.** El cas més corrent és el de la transcripció d'una entrevista.

### **Transcripció d'una entrevista**

Per una banda les preguntes de l'entrevistador poden veure's com a categories de la seva activitat verbal, sobretot si formen part d'un qüestionari preparat. Tocant a aquestes preguntes, doncs, estaríem en la primera situació de partida.

Per altra banda, les respostes de l'entrevistat han de ser encara contextualitzades, codificades i –finalment– classificades. Així, tocant a les respostes estaríem com en la segona situació de partida, és a dir, sense sistema de categories.

Ara bé, si les preguntes corresponguessin a una entrevista oberta, en la qual no estan fixades prèviament, ens acostaríem de nou a la segona situació de partida, aquella en la qual tota la feina està per fer.

Els **criteris de classificació** utilitzats per a categoritzar un text estan dictats, com sol ocórrer, pels objectius de l'anàlisi, però també per l'entorn d'interessos, incloent-hi els teòrics, en què es mou el científic. Abans hem dit que les categories en l'anàlisi de contingut no són eines per a l'aprofundiment teòric, com, per exemple, sí que ho són en el cas de la teoria fonamentada. No obstant això, això no impedeix que a l'hora d'escollir criteris per a fixar d'una vegada per totes el sistema de categories, l'investigador recorri a un marc teòric o, si la recerca és inductiva, a les microhipòtesis i suposicions que sorgeixen en el curs de la investigació.

Amb tot, des d'un punt de vista general, aquests criteris poden considerar-se semiòtics o comunicacionals, malgrat que aquesta adscripció no sempre es fa explícita. Però és indiscutible que un text posseeix sempre un valor expressiu

o comunicatiu i, en qualsevol cas, de transmissió intencional d'informació (el text no apareix per generació espontània), valors que es realitzen en el moment que un lector o un intèrpret s'hi posa al davant. Ens concentrarem, doncs, en aquest aspecte fonamental del text i de les categories que se'n poden derivar, i distingirem entre criteris de categorització:

- **Sintàctics.** En el pla comunicacional correspondrien a la forma o estructura del missatge. Dintre d'aquests trobem les distincions gramaticals o sintàctiques estrictes (subjecte/predicat/complement/...; nom/adjectiu/verb/adverbi/...). També cauen en aquest epígraf les categories anomenades *de forma*, referides més aviat a aspectes retòrics o estilístics. Lasswell, per exemple, distingia entre declaracions d'identificació ("jo sóc comunista"), de preferència ("els comunistes tenen raó de voler fer la revolució") i de fet ("els comunistes conqueriran el món"). En termes retòrics podem diferenciar argumentacions que recorren a autoritats d'altres argumentacions que fan valer la lògica.
- **Semàntics.** En el pla comunicacional correspondrien al contingut del missatge. Aquí s'inclouen les classificacions basades en temes i en conceptes, i també les de tipus lèxic (agrupacions de sinònims, homònims). En la literatura aquests criteris s'apliquen a categories anomenades *de matèria*, que preveuen temes de conversa (la moda, l'esport, la política, homes enfront de dones, la salut, els diners), temes de programes radiofònics (noticiaris, xafarderies, música, tertúlies, magazins), televisius (fulletons, telenotícies, pel·lícules, concursos, telescombraries), etc. Igualment podem encabir en aquest apartat les categories dites d'"apreciació" o de "valors", les quals engloben dimensions com "aprova-desaprova", "afirma-nega", "bé-mal", i així successivament.
- **Pragmàtics.** En el pla comunicacional es referirien a l'emissor, a la situació d'emissió, al canal de transmissió, al receptor, als efectes copsats en aquest darrer. Entre aquests criteris comptaríem els que afluïren quan hi ha diferents utilitzacions del mateix text en diferents contextos (per exemple, diferents usos d'un mateix missatge segons els propòsits de l'emissor o el seu punt de vista). Les tàctiques seguides per l'autor del text per a aconseguir els efectes buscats també donen lloc a categoritzacions pragmàtiques, com "propaganda", "negociació", "exhibició de poder", "seducció", "suborn", "adulació", etc. Així mateix incloem aquí les classificacions fetes sobre la base de la intensitat dels efectes en el receptor. En aquest sentit, hi ha textos o missatges més potents que altres. Finalment, formen part igualment dels criteris pragmàtics aquells que resideixen en els trets diferencials dels actors o protagonistes del text (els personatges de la narració) i els que permeten establir perfils d'emissors o de fonts de procedència, i també de receptors.

## Classificacions de Lasswell i d'Stewart

En l'àmbit dels estudis clàssics trobem la classificació proposada per Lasswell per a l'anàlisi dels fenòmens polítics en els textos. La llista fa gala d'una clara orientació pragmàtica.

- 1) Actors o persones dedicats a la política ("polítics", en clau popular).
- 2) Comunitats de referència que figuren en els textos (nord-americans, jueus, comunistes, partits, sindicats).
- 3) Organitzacions presents en els textos (Congrés, Parlament, Senat, Tribunal Suprem).
- 4) Estats o situacions de la política (guerra/pau, crisis, reformes).
- 5) Ideologies o sistemes polítics (democràcia, dictadura, liberalisme).

En els estudis –també clàssics– de Stewart sobre la premsa americana, es distingeix d'antuvi entre fets i comentaris segons que el prisma sigui més semàntic o més pragmàtic i, dintre de cada una d'aquestes classes, se separen diferents subcategories:

### 1. Fets

- 1.1. Fets presentats com a certs
- 1.2. Fets presentats com a possibles
- 1.3. Fets la certesa dels quals es valora amb un judici
- 1.4. Fets la incertesa dels quals és objecte de valoració

### 2. Comentaris

- 2.1. Personals de l'autor o de la font directa de procedència
- 2.2. Indirectes d'emissor o font determinats
- 2.3. Indirectes d'emissor o font no determinats
- 2.4. Cita directa de font determinada
- 2.5. Cita directa de font no determinada

Vegeu M. Duverger (1975). *Métodos de las ciencias sociales* (8a. ed., pp. 182-183). Barcelona: Ariel.

Pel que fa a l'estructura lògica dels sistemes de categories d'anàlisi de contingut, no difereix essencialment de la que presenta un sistema de categories conductual o de la que hem mostrat en els exemples de teoria fonamentada. No obstant això, en l'anàlisi de contingut, les exigències formals es posen en primer pla i per això convé recordar-les de nou i amb un cert èmfasi:

- L'organització lògica de les categories textuais té la forma d'un arbre classificatori. L'exigència bàsica que s'aplica a la classificació és que aquesta satisfaci les condicions semàntiques o conceptuals d'exclusivitat i exhaustivitat. La condició d'exhaustivitat estableix que el sistema de categories ha de cobrir tot el text seleccionat per a l'anàlisi, per la qual cosa cap segment d'aquest no pot escapar a la classificació. La condició d'exclusivitat exigeix que les definicions de cada categoria no s'encavalquin o es confonguin, la qual cosa en la pràctica equival a demanar que els trets, components o indicadors de cada categoria, o la seva combinació, siguin diferents per a cada una.
- En canvi, l'organització temporal de les categories al llarg del registre, vessant important en l'ús d'un sistema de categories conductual, no és en general pertinent en un sistema de categories textual. Les exigències d'exclusivitat i exhaustivitat temporals no s'apliquen al text. Entenem que la seqüència de segments textuais és "neta" una vegada fixat el text, i això

### Vegeu també

Recordeu que teniu disponibles exemples de teoria fonamentada en el mòdul "Les tècniques qualitatives en la selecció i tractament de la informació".

<sup>(2)</sup>Deixar espais buits representaria tenir una pàgina o un espai en blanc, però la lectura passaria a la pàgina o espai següent i seria contínua.

equivale a dir que els segments del text no es poden encavalcar en el temps (no poden ser simultanis), ni poden deixar espais buits<sup>2</sup>. Certament, en la transcripció d'una conversació o d'una entrevista pot ocórrer que dos interlocutors parlin a la vegada, però, a escala textual, la simultaneïtat tampoc no existirà, si no es vol, i si la notació del registre indica aquest encavalcament, es referirà al fenomen en si, no a la transcripció en forma de text: els segments es poden considerar d'un en un.

#### Vegeu també

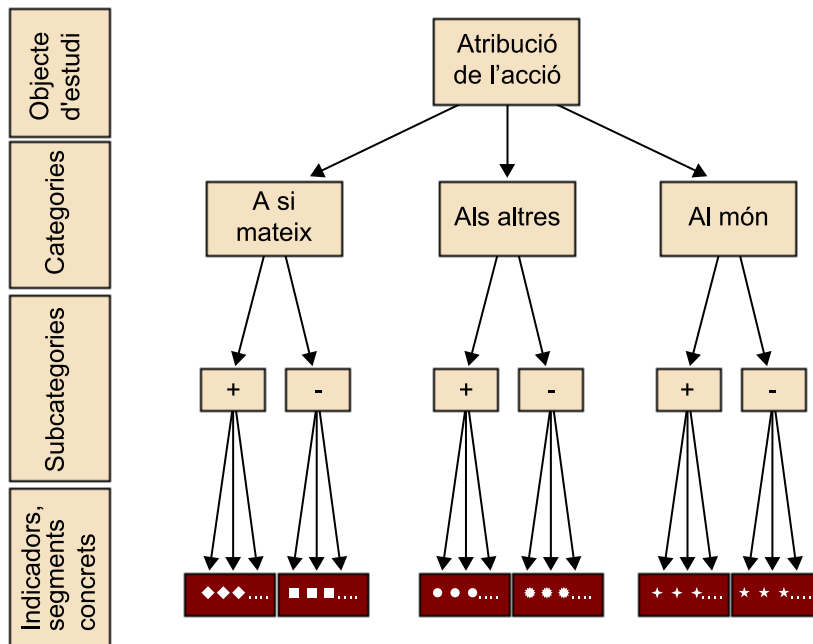
Totes aquestes qüestions poden ser consultades en els mòduls 2 i 5 de l'assignatura *Mètodes d'investigació qualitativa*.

En l'interior d'un sistema de categories d'anàlisi de contingut trobem, com en qualsevol altra classificació, una sèrie de pisos o estrats corresponents a diferents nivells d'organització. Ja hem dit que l'enllaç entre aquests nivells successius no pot prescindir, ni tan sols en les orientacions més quantitatives, de l'anàlisi de contingut, d'alguna aposta teòrica o hipotètica que li doni suport, almenys provisional.

Per tant, la construcció d'un sistema així implica inevitablement una sèrie de decisions prèvies quant al tipus de relacions verticals que enllacen els diferents nivells jeràrquics d'organització. Aquestes relacions són eminentment semàntiques, atès que impliquen atribucions de significat. Fent un recorregut de dalt a baix, partint del nivell superior corresponent a l'objecte d'estudi, davallarem fins al nivell de les categories pròpiament dites, després baixarem al de les subcategories i finalment al dels segments o subsegments, en tant que unitats de registre o components de les categories on estan classificats. A una unitat de registre se li atorga un significat que la classifica dintre d'una certa categoria o subcategoria; i al seu torn les categories posseeixen un significat que configura l'objecte d'estudi. Per això, en determinats àmbits, aquestes unitats amb un vector de significat adreçat a la categoria s'anomenen *indicadors*. Sigui com sigui, el que fem sempre és assignar una sèrie de significats, a diferents nivells, al segment o unitat de registre.

#### Categories extretes de missatges electrònics

A la primera figura hem il·lustrat aquests punts amb l'exemple de tres categories extretes de missatges electrònics d'un adolescent. Aquests, segons a qui atribueixi la responsabilitat o l'agència de l'acció l'autor del text, poden reduir-se a tres: les atribuïdes a l'escriptor mateix del missatge, les atribuïdes a qualsevol altra persona o les que remetent al món físic o a la natura en general. Cada una d'aquestes categories, en atribuir l'acció, pot tenir una valència positiva (recolzada en indicadors –mots, frases, oracions– de mèrit, èxit, benefici, millora, judici moral o ideològic positiu, etc.) o negativa (indicadors de demèrit, fracàs, dany, empitjorament, judici moral o ideològic negatiu, etc.), però no neutra. La categoria "Al món" podria estimar-se una categoria zero com les que hem caracteritzat en el final del mòdul "Les tècniques qualitatives en la selecció i tractament de la informació", ja que l'interès del subjecte, i de l'investigador, apunta a la dimensió de socialització ("jo / els altres") i no tant al món en general. És indiscutible que la distribució de segments per subcategoria i categoria i la forma com aquestes omplen l'objecte d'estudi obeeix a judicis de valor de l'investigador, judicis sempre revisables però que a partir de la constitució del sistema de categories en guien l'anàlisi, per molt quantitativa que sigui.



### 1.4.3. Funcions d'inferència

És obvi que un dels objectius metodològics de l'investigador ha de ser el de validar aquestes relacions entre indicadors i categories, abans de trobar regularitats o confirmar hipòtesis. Tinguem present que la validació d'aquestes relacions és, de fet, una validació de la interpretació, ja que les categories són sempre interpretatives i són les que configuren l'objecte d'estudi; i sovint, és també una confirmació del significat latent o ocult assignat al text o a parts del text, atès que les categories validades precisament incorporen aquesta interpretació més arriscada.

Al marge d'això, l'investigador qualitatiu té davant de si diversos camins de validació:

- Un, de típicament qualitatiu, és el de consensuar amb l'autor del text o amb altres analistes la categorització definitiva, amb una **validació èmica**. Això no sempre és possible, atès que l'autor pot ser absent, anònim o renuent a complir aquest paper d'informant.
- Un altre, també genuïnament qualitatiu, i en l'esperit de les tècniques explorades en el mòdul "Les tècniques qualitatives en la selecció i tractament de la informació", demana un itinerari espiral, en el qual l'autor torna una i altra vegada al text, modificant i ajustant les categories segons la coherència de les conclusions a les quals hagi arribat.
- Un altre és el de recórrer a tècniques multivariables similars a les que serveixen per a dotar als tests de validesa discriminant i consistència estructural (anàlisi factorial, components principals, escalament multidimensi-

onal, agrupacions o clústers, alfa de Cronbach, etc.). Aquest és típic de les aproximacions més "dures" de l'anàlisi de contingut.

Sobra afegir que les tres vies són, en principi, compatibles.

Un cop garantida l'estructura del sistema de categories cal plantejar-se una qüestió ben diferent, com és la de quines inferències es poden fer a partir de les mesures fetes. Com veurem aviat les mesures (per exemple, la freqüència d'ocurrència) es poden prendre:

- Amb els segments o unitats de registre. La seva funció és justificar el sistema en tant que instrument consistent i coherent d'anàlisi, en línia amb el que acabem de comentar en els paràgrafs anteriors.
- Amb les categories o subcategories. La seva funció és la de propiciar explicacions i inferències en comparar parts d'un text o diversos textos entre si.

### **Exemples de línies d'inferència**

Us oferim alguns exemples de línies d'inferència, precedides o no per hipòtesis, compatibles amb l'anàlisi de contingut.

- A partir de retalls de premsa, en fase de conflicte prebèl·lic, deduir i predir si un país *A* atacarà un país *B*.
- Sobre la base de la transcripció d'una entrevista, saber si el subjecte ha intentat seduir, agradar, a l'investigador o, al contrari, li ha fet el boicot i ha mantingut distància.
- Establir els trets psicològics de l'autor d'un diari (per exemple, el de Sylvia Plath, poetessa anglesa que es va suïcidar el 1963).
- Detectar el nivell de sinceritat en un autoinforme.
- Reconstruir les línies mestres d'una estratègia de negociació mitjançant l'epistolari entre dos polítics, empresaris, etc.
- Esbrinar en quins aspectes de l'expressió textual es manifesta el racisme o el sexisme en articles periodístics, novel·les, llibres de text, etc.
- Detectar la percepció que un infant té dels immigrants a partir de relats escrits que se li han demanat a l'escola.

El sistema de categories creat es pot validar amb els autors dels documents o els subjectes registrats, amb anàlisis reiterades del text o amb procediments estadístics.

## **1.5. Mesures i índexs bàsics en l'anàlisi de contingut**

La mesura bàsica de l'anàlisi de contingut és la freqüència i es pot aplicar a qualsevol codi utilitzat: unitats de registre o segments i categories.

### 1.5.1. Mesures bàsiques de freqüència i regles d'enumeració

Acabàvem el subapartat 1.4 dient que l'anàlisi de contingut fa mesures en dos nivells ben definits: el de les unitats de registre i el de les categories. I en cada un d'aquests nivells la funció de la mesura és diferent: en el primer serveix per a ponderar la importància o incidència de cada molècula textual, per a legitimar-ne la incorporació en una o altra categoria; en el segon, recolza les inferències o permet confirmar hipòtesis de treball.

Les autèntiques variables són, doncs, les categories, ja que és a partir d'aquestes que es generen els resultats de la recerca. Són, això sí, variables qualitatives, dicotòmiques o politòmiques, la qual cosa equival a dir que no prenen valors continus, sinó binaris de presència/absència, sí/no, 1/0. Són les variables que apareixien en els exemples que hem vist. Un autor, locutor, emissor, etc., "enuncia fets" o "no enuncia fets"; "fa comentaris" o "no fa comentaris"; "traça una imatge ridícula del seu pare" o "no traça una imatge ridícula del seu pare".

En el si d'aquesta poderosa restricció, la mesura més adient i més rendible és la freqüència absoluta d'ocurrència o freqüència a seques, un recompte dintre de la sèrie de nombres naturals. Comptabilitza el nombre de vegades que una categoria està present en un text o en zones seleccionades del text.

#### Harmonia entre l'estil qualitatiu i el quantitatiu

Adoneu-vos que l'ús conjunt del següent:

- variables categorials bastides sobre la dicotomia *presència/absència*, i
- mesures freqüencials d'aquestes variables,

és el nucli d'una harmoniosa sintonia entre l'estil qualitatiu i el quantitatiu. Efectivament, l'ús de taules de presència/absència, per un costat, i de mesures contínues, per un altre, és un dels trets que marquen el contrast principal entre els mètodes qualitius i els quantitius, per la qual cosa la seva utilització conjunta ha d'interpretar-se com un avenç envers la integració metodològica.

Tanmateix, aquesta promesa d'integració segueix sense eliminar la limitació essencial dels mètodes quantitius i, en particular, de la quantificació freqüencial i altres mesures. I és que no són capaces d'assignar valor a la categoria o al fenomen que no està present, que "brilla per la seva absència". L'enfocament quantitatiu s'enfronta a enormes dificultats per a donar sentit al zero.

Per altra banda, la preeminència de la freqüència també és un tret de la metodologia observacional en general, la qual també fa descansar les seves mesures sobre categories. Així doncs, com a metodologia d'observació indirecta o documental, l'anàlisi de contingut es caracteritza per la utilització, no solament de la freqüència com a mesura principal, sinó d'un lot de mesures derivades d'aquesta o edificades a sobre, de les quals la freqüència relativa o proporció d'ocurrències és la capdavantera.

Per comoditat, a partir d'ara, expressarem la freqüència d'una categoria de contingut  $C$  com a  $f_C$  i la seva freqüència relativa (que permet estimar probabilitats d'ocurrència) com a  $p_C$ :



La **frequència**,  $f_C$ , és el nombre de vegades que una unitat de registre codificada o una categoria està present en un text o secció de text.

La **frequència relativa**,  $p_C$ , és la proporció de vegades que una unitat de registre codificada o una categoria està present en un text o secció de text respecte a la freqüència total d'ocurrències. Per tant,

$$\hat{p}_C = \frac{f_C}{f_T}$$

El denominador és el total d'ocurrències. El l'accent circumflex sobre la  $p$  indica que la freqüència relativa es pot fer servir com a estimador de probabilitat.

Si un anàlisi de contingut es fa sobre un sistema de categories  $S$  tal que  $S = \{W, X, Y, Z\}$ , llavors la freqüència total  $f_T = f_W + f_X + f_Y + f_Z$ . Així, si hem obtingut  $f_W = 56$ ,  $f_X = 14$ ,  $f_Y = 23$  i  $f_Z = 102$ , obtindriem  $f_T = 195$ . Llavors,

$$p_W = 56 / 195 = 0,29$$

$$p_X = 14 / 195 = 0,07$$

$$p_Y = 23 / 195 = 0,12$$

$$p_Z = 102 / 195 = 0,52$$

Naturalment,  $p_W + p_X + p_Y + p_Z = 1$ , com podeu comprovar.

Òbviament, si multipliquem  $p_C$  per 100 obtindrem el percentatge d'ocurrència corresponent a cada categoria (en aquest exemple, 29%, 7%, 12% i 52%, respectivament).

En canvi, per raons que ja hem adduït, l'anàlisi de contingut no fa ús de mesures de temps, ja que el text és estàtic des del punt de vista microtemporal. És a dir, malgrat que alguns textos poden transformar-se al llarg de mesos, anys o segles (d'esborranys a textos definitius, d'una versió a una altra, d'una formulació de lleis o estatuts a una altra, d'un cos doctrinal a un altre), es donen a l'analista d'un sol cop, no progressivament, com es dóna el comportament al llarg d'una sessió.

Per consegüent, tampoc no té sentit dir que una categoria ha aparegut en un text 15 vegades per minut o que ha ocupat un 24% del temps de registre, afirmacions que sí que tenen cap i peus en un registre observacional de sessions: un infant pot canviar d'activitat 2 vegades per minut, de mitjana; un obrer pot soldar 26 peces per hora.

La freqüència és una classe de mesura que no té valor si no és emmarcada en un espai o en un temps de referència.

### Importància de l'espai i del temps per a la freqüència

No vol dir gairebé res dir que un infant ha canviat d'activitat dues vegades. Ens preguntarem de seguida en quant temps ha fet aquests dos canvis d'activitat, perquè no té el mateix significat que els hagi fet en un minut, en una hora o en un dia. Tampoc no té gaire valor l'afirmació *a l'espectacle van assistir 200 persones*, atès que aquestes 200 persones poden atapeir un local justament amb aquesta capacitat, o bé omplir tan sols un sector d'una platea amb 1.000 seients. En l'anàlisi de contingut, descartat el temps, és obligada la mateixa exigència en relació amb l'espai. Si en la transcripció d'una entrevista terapèutica comptabilitzem 6 desqualificacions de la parella, caldrà situar aquest nombre en el context de l'entrevista, saber quina ha estat l'extensió i com s'han distribuït les desqualificacions al llarg de tot el desenvolupament.

Advertiu, doncs, que és capital contextualitzar la mesura, posar-la en un context. Ara, però, aquest no és un context d'interpretació, sinó de referència de la mesura; o, simplement, uncontext de mesura. En el vocabulari de l'anàlisi de contingut diem que no solament cal establir les mesures d'un text sinó també les seves regles d'enumeració. Aquestes es refereixen al fons o marc en el si del qual cobra significat la mesura.

### Freqüència de trastorns

En l'anàlisi de contingut les regles d'enumeració més corrents donen, com a marc de referència de la freqüència i altres mesures, la línia i la pàgina. De vegades les mesures també se situen respecte a l'oració, el paràgraf, el capítol.

Podem dir així que una categoria  $C$  ha aparegut dues vegades en una línia o 7 vegades en una pàgina, amb la qual cosa, més que amb una freqüència, treballem amb una taxa (tanmateix, continuarem mantenint el mateix símbol):

$$f_{C/línia} = 2 \text{ o } f_{C/pàgina} = 7$$

Més habitualment, podem afirmar que una categoria  $C$  ha aparegut una mitjana de  $f_C$  vegades per línia o de  $f_C$  vegades per pàgina. En l'exemple utilitzat en el subapartat 1.3.3, la freqüència de *trastorn* era 8 i les seves freqüències mitjanes, per línia o paràgraf, eren:

$$\bar{f}_{C/línia} = \frac{8}{18} = 0,44 ; \bar{f}_{C/paràgraf} = \frac{8}{4} = 2$$

Podríem haver donat aquestes freqüències i mitjanes per oració, pàgina o capítol però, tot i que aquests denominadors poden estar justificats en certes condicions, no permeten una estandardització de la mesura tal com la fan possible marcs de referència fixos, com la línia o la pàgina. Si considerem la freqüència respecte a contextos d'extensió variable com l'oració, la taxa resultant és difícil d'interpretar.

Finalment les freqüències, de la classe que siguin, també es poden calcular respecte a grans trams de text, delimitats de manera arbitrària o objectiva (segons el nombre de mots, línies) o d'acord amb l'estructura que s'hi endevina. Per exemple, el text d'una narració pot dividir-se en 10 parts iguals, des del principi fins al final, i computar la freqüència de les diverses categories previstes en cada una d'aquestes divisions. Això ajudarà, en una aproximació inductiva, a esbrinar com es reparteixen les categories al llarg i ample del text, abans d'advertir hipòtesis sobre la seva organització.

En un exemple tret del llibre de Grbich (2007, p. 119), les freqüències de la categoria arrelada en el mot *love*, al llarg de les 10 parts d'un text de Shakespeare són:

Part 1: 17	*****
Part 2: 8	*****
Part 3: 11	*****
Part 4: 10	*****

Part 5: 5	*****
Part 6: 25	*****
Part 7: 9	*****
Part 8: 5	*****
Part 9: 5	*****
Part 10: 4	****

Òbviament, per raons dramàtiques, el mot s'utilitza sobretot en el 60% inicial de l'obra, i l'ús decau en picat en el 40% final (Grbich, 2007, p. 119).

Com és natural podem calcular també, si ens interessa, les desviacions típiques, variàncies, etc.

### 1.5.2. Altres mesures derivades de la freqüència o que la pressuposen

#### Freqüència ponderada

L'ús de la freqüència com a mesura treu partit d'una sèrie de supòsits, dels quals un dels més importants és el següent: en principi totes les unitats o categories són equivalents i virtualment tenen la mateixa probabilitat d'aparèixer en el text. És com si diguéssim que totes les categories arrenquen de la mateixa línia de partida en una cursa i cap s'avança a l'altra fins que no comença la cursa. Sobre la base d'aquest supòsit inicial, la rellevància d'una unitat o categoria creix a mesura que creixi la seva freqüència.

Aquest supòsit, però, topa sovint amb diverses limitacions. Ja n'hem citat una: de vegades la circumstància que un element del text tingui freqüències baixes o no aparegui és tan significativa com el seu contrari.

Un altre obstacle, que comentarem a continuació, és que aquesta condició d'equivalència pot, senzillament, no complir-se, atès que hi ha mots, frases, expressions, temes que, en un determinat tipus de text, tenen més presència *a priori* que altres. Àdhuc, podem fer una estimació teòrica de la incidència d'aquests elements en el text, ja que existeixen, sortosament, registres d'ús en moltes llengües que ens diuen quina és la freqüència amb la qual podem esperar trobar-los. Aleshores, la incidència d'un mot o d'una categoria en un text s'ha de posar en relació amb la seva incidència estadística mitjana en materials estàndard del mateix tipus.

Per raons com la que acabem d'exposar i d'altres semblants, s'utilitzen sovint mesures de freqüència ponderada.

Les mesures de freqüència ponderada són mesures de freqüència en què cada unitat o categoria registrades no té el mateix "pes" o rellevància, sinó que unes en posseeixen més que altres. El pes és assignat multipliant cada freqüència empírica per un valor decidit prèviament, a partir de consideracions teòriques o de dades anteriors.

### Exemple d'ús de freqüències ponderades

Si tenim un sistema de categories  $S = \{X, Y, Z\}$  i considerem que  $Y$  té el doble de pes que  $X$ , i  $Z$  el triple, aleshores les freqüències ponderades  $f_{\text{pond}}$  que comptabilitzarem seran:

$$f_{\text{pond } X} = f_X; f_{\text{pond } Y} = 2(f_Y); f_{\text{pond } Z} = 3(f_Z)$$

Si  $f_X = 61$ ,  $f_Y = 20$ ,  $f_Z = 14$ , aleshores  $f_{\text{pond } X} = 61$ ;  $f_{\text{pond } Y} = 20 \times 2$ ;  $f_{\text{pond } Z} = 3 \times 14$ .

### Intensitat

Un desafiament seriós dins de l'anàlisi de contingut és la mesura de la intensitat, essencialment vinculada a les categories semàntiques. Aquest desafiament, ara per ara, és difícilment afrontable fora de procediments psicòmètrics.

Assumir la mesura de la intensitat vol dir, en el fons, transcendir el plantejament dicotòmic o binari de la quantificació categorial i passar, com a mínim, a una quantificació politòmica. És a dir, que les coses, si es desitja mesurar la intensitat d'una categoria, no són blanques o negres, sinó que poden oferir una gamma de grisos, gamma graduada en forma d'escala ordinal, si més no.

### Exemple de mesura de la intensitat

A tall d'il·lustració, compareu aquests tres enunciats corresponents a tres unitats de registre classificables en una categoria que podríem etiquetar com "rebuig de la política presidencial" per part de polítics de l'oposició:

- a) "Podríem arribar a considerar la possibilitat de desaprovar la política del president".
- b) "Cal denunciar, amb pena, l'actual orientació política del president".
- c) "Certament, en el passat hem estat sovint en desacord amb la política del president".

És clar que convindria situar cada una d'aquestes afirmacions en una escala d'intensitat quant al grau de rebuig que expressa o vol expressar. Un procediment artesanal (però força emprat) seria assignar ponderacions a cada ítem  $a$ ,  $b$  i  $c$ , tractant-los com si fossin subcategories ordenades d'una categoria general de rebuig. Els criteris en què recolzaríem podrien ser diversos i combinables: semàntica dels verbs tal com els defineix el diccionari o els usuaris, temps d'aquests mateixos verbs, adverbis –especialment els de mode– que reforcen o suavitzen el contingut verbal, adjectius qualificatius, signes d'admiració... En l'exemple anterior podríem assajar la ponderació  $a = 2$ ,  $b = 3$ ,  $c = 1$ .

Més endavant, ens aturarem en un exemple d'anàlisi avaluativa en què s'utilitzen ponderacions d'intensitat. Més enllà de procediments com aquest, la construcció d'una escala d'intensitat obliga a aplicar totes les fases i requisits d'un escalament psicomètric.

### Direcció

Bona part del que hem dit a propòsit de la intensitat és també aplicable a la direcció. Moltes interpretacions del sentit o valor d'un text es tornen més riques i completes si se situen en una dimensió de mesura, entre un mínim i un màxim. Per exemple, en un estudi sobre integració d'infants immigrants en la població autòctona podríem servir-nos d'una categoria de "Desig de convivència" i mesurar aquesta entre un desig mínim, zero o negatiu, i un de màxim, amb un nombre de valors intermedis.

Històricament, aquestes mesures han estat resoltes a partir d'escala d'estimació, de Likert, i en general qüestionaris als quals el subjecte ha de respondre escollint un valor dins d'una sèrie: d'un a cinc, d'un a set, d'un a deu, etc. Charles Osgood ha estat un dels paladins d'aquesta classe d'estratègia.

#### **Altres resolucions de les mesures de direcció**

Una altra via de resolució de les mesures de direcció és la creació d'índexs amb un àmbit de variabilitat adequat a aquest tipus de recerca.

### **1.5.3. Índexs**

#### **Diversitat**

Un índex molt popular és el de **diversitat** d'unitats o categories. Tradicionalment s'ha emprat una proporció senzilla entre el nombre d'unitats o categories diferents (*Dif*) i el nombre total d'unitats o de categories (*T*):

$$I_d = \frac{Dif}{T}$$

Si invertim el quocient, tindrem un **índex de pobresa** o **monotonia**,  $I_m$ :

$$I_m = \frac{T}{Dif}$$

#### **Càlcul de la diversitat i de l'índex de pobresa**

Considerem el paràgraf següent: "Un índex molt popular és el de diversitat d'unitats o categories. Tradicionalment s'ha emprat una proporció senzilla entre el nombre d'unitats o categories diferents i el nombre total d'unitats o categories".

En aquest segment, el valor de  $I_d$  per a unitats textuals de tipus "mot" seria  $I_d = 23/35 = 0,657$  (*s'* i *d'* han estat comptats com un mot). El valor de  $I_m$  seria  $I_m = 35/23 = 1,52$ .

#### **Índex i teoria de la informació**

Actualment, aquesta classe d'índexs se solen calcular traient partit de la teoria de la informació. En l'àmbit de l'observació de camp, es pren com a fórmula de la relació entre *D* i *T* l'equació  $D = CT^z$ , en què *C* i *z* són les constants que, precisament, cal esbrinar en cada investigació. Aquesta relació pot transformar-se en lineal sobre escala logarítmica, ja que

$$\log D = \log C + z \log T.$$

Per tant, en una gràfica logarítmica la relació entre el nombre de categories o unitats diferents i el nombre total d'aquestes registrades fins a un cert moment és una recta, la qual es pot dibuixar mitjançant tècniques de regressió simple.

## Densitat

L'índex de densitat fa una avaluació diferent. La densitat d'un text és la suma de les freqüències de totes les categories o codis ( $F_C$ ) dividida per la suma total de totes les paraules, segments o unitats bàsiques d'anàlisi comptabilitzades ( $F_U$ ) (aquest denominador coincidiria pràcticament amb  $I_d$ ). Si el sistema de categories té una estructura tal que l'exclusivitat es compleix en tots els seus nivells, aleshores aquest quocient serà 1, ja que a cada segment o unitat correspondria una sola categoria. Però, com sabem, sovint la relació entre categories i segments és de caire combinatori, de manera que una mateixa unitat o segment pot pertànyer a diferents categories i la identitat de cada categoria s'estableix a partir de la constel·lació particular de segments que la caracteritzen. Llavors el valor de la densitat pot ser diferent d'1.

$$D = \frac{F_C}{F_U}$$

### Frequència d'ús de diferents categories gramaticals

Vegem dos exemples d'índexs dedicats a quantificar la relació entre la freqüència d'ús de diferents categories gramaticals. El quocient adjectius/verbs (amb denominador habitualment fixat en 100) ha permès distingir la locució esquizofrènica de la normal o no patològica. Per a alguns autors el quocient (noms + verbs)/(adjectius + adverbis) és un bon indicador del tarannà del discurs polític, dinàmic o descriptiu segons si el trencat supera la unitat o queda per sota.

Una altra família popular d'índexs és la dels dedicats a quantificar la relació entre la freqüència d'ús de diferents categories gramaticals.

Les freqüències i altres mesures derivades d'aquesta s'han de contextualitzar mitjançant regles d'enumeració.

## 1.6. Mesures de contingència i seqüencials en l'anàlisi de contingut

Les mesures de contingència i seqüencials estableixen relacions entre unitats textuais (codis, categories) o entre unitats textuais i de context.

### 1.6.1. Mesures de coocurrència

La **freqüència de contingència** o **coocurrència** és una variant de la freqüència d'ocurrència, per la qual coneixem el nombre de vegades que dues o més unitats de text coincideixen en una mateix context d'interpretació, o el nom-

bre de vegades que una unitat de registre codificada o una categoria coincideix en un mateix context de mesura (línia, paràgraf, pàgina, capítol; frase, oració, tema).

Aquesta mesura és la clau de volta d'una determinada línia d'anàlisi, ja citada: l'anàlisi de contingències, associacions o concomitàncies. Quan la coincidència es refereix particularment a la contigüïtat seqüencial o successió entre antecedents i consegüents (en relació text-context o sota altres òptiques), aleshores l'anàlisi de contingències és un altre nom de l'anàlisi seqüencial, de la qual ens ocuparem de seguida.

### **Exemple d'us de la freqüència de contingència**

Podem utilitzar la freqüència de contingència per a fer afirmacions com que, en un material com el de l'exemple vist de Weber en el subapartat 1.3.3, la unitat textual "drets" i la unitat textual "igualtat" han coincidit 21 vegades en el context d'oracions que són enunciats de lleis. Formalment, si "drets" l'anomenem  $w$  i "igualtat" l'anomenem  $x$ , i aquest tipus d'oració el designem per  $L$ , tindrem que

$$f_{w,x/L} = 21.$$

O, igualment, podríem servir-nos d'aquesta freqüència per a dir que la categoria  $Y$  i la categoria  $Z$  han coincidit 17 vegades en una línia o 34 vegades en una pàgina, ara referint-nos al context de mesura en lloc del d'interpretació.

### **Freqüència de contingència del trastorn límit de personalitat**

Tornem a l'exemple del trastorn límit de personalitat del subapartat 1.3.3 d'aquest mateix mòdul. Si busquéssim la concurrència, per paràgraf, de la categoria centrada en "trastorn" (referències a la malaltia en qüestió) amb una altra categoria que apuntés al nivell d'incidència poblacional de la malaltia i a la seva importància social (referències al 2% d'incidència, a l'augment de la incidència, etc.), ens trobaríem que les dues coincideixen:

- en el primer paràgraf ("trastorn" i "afecta un 2% de la població");
- en el segon paràgraf ("trastorn" i "augment de la incidència en la nostra societat...");
- i en el quart ("trastorn" i de nou "al 2% de la població").

Podríem afirmar, doncs, que pel que fa a aquesta minúscula mostra de 4 paràgrafs la freqüència relativa de contingència del parell "trastorn - incidència poblacional de la malaltia" és de 3/4. Si la mostra fos més gran, podríem fins i tot estimar que la probabilitat que totes dues categories es presentin juntes és justament de 3/4, és a dir, 0,75. Lògicament, per a obtenir aquesta xifra dividim el nombre de casos favorables, aquells en els quals realment s'ha presentat l'associació buscada, pel nombre de casos possibles, proporcionat pel nombre de paràgrafs considerat. Aquesta és una probabilitat condicionada, ja que subordinem l'ocurrència de "trastorn" a la d'un altre fet, a saber, que hi hagi un paràgraf que l'aculli. Evidentment, podríem repetir aquest exercici amb totes les categories utilitzades en l'anàlisi del text, i examinar les diferents combinacions en parells, trios, etc. Així mateix, podríem prendre com a referència la línia, la pàgina, etc.

Aquesta orientació de l'anàlisi té moltes més possibilitats que les suggerides aquí però, per tal de no repetir-nos, les desenvoluparem a continuació i en relació amb l'enfocament seqüencial. Aquest constituiria l'altra cara de la moneda de l'anàlisi que acabem de presentar: en comptes de computar i interpretar la coincidència de dos o més elements en un mateix espai del text, computaríem i interpretaríem els nexes precedent-consegüent amb diferents distàncies entre tots dos. La mecànica del procediment pot ser la mateixa i ara la detallarem una mica més.

### 1.6.2. Mesures seqüencials

Les mesures seqüencials són de fet mesures de l'ordre o de l'organització de la seqüència textual. L'ordre amb què es manifesten seqüencialment unitats de registre o categories és un altre vessant de la informació que admet quantificació. En aquest terreny, l'anàlisi de contingut comparteix recursos, un cop més, amb la metodologia observacional, en el si de la qual s'han desenvolupat una munió de tècniques destinades a la mesura i modelització de l'ordre dels elements d'una sèrie o seqüència.

En un enfocament d'anàlisi de contingut pot interessar-nos conèixer el grau d'organització seqüencial d'un text, les regles subjacents, el patró nuclear que li confereix la seva forma aparent. En perspectiva plenament qualitativa, l'investigador pretén sovint descobrir la matriu de ritual o cerimònia que conforma el material verbal o textual, tal com es posa especialment de manifest en els enfocaments etogènics i etnometodològics. Per tal d'arribar a escatir aquest ordre, cal començar trobant quins lligams hi ha entre els elements (unitats, codis, categories) que componen la cadena del text.

El punt de partida de la cerca de patrons d'ordre consisteix en el còmput de freqüències de transició. Aquestes són les freqüències amb les quals cada una de les categories d'un sistema segueix en la seqüència textual la resta de categories del sistema, incloent-hi ella mateixa. En general, doncs, són freqüències de parells d'elements contigus,  $f_{ij}$ . De fet, es podrien computar tant freqüències de precedència ( $I$  precedeix  $J$ ) com de seguiment ( $J$  segueix  $I$ ), o totes dues a la vegada. Ja sabem que, des del punt de vista qualitatiu, la categoria que precedeix una altra categoria pot entendre's com el seu context antecedent; i la que segueix aquesta mateixa categoria, com el seu context consegüent. Copsareu que les freqüències de transició són una variant de les freqüències d'associació o contingència, tal com hem declarat abans.

Ara bé, tot i que en les línies prèvies hem pensat sempre en precedències o seguiments immediats, d'elements contigus, recordem que això no ha de ser així necessàriament: les dependències entre els elements d'una cadena textual poden establir-se a més distància, entre una categoria i la que la precedia o seguia dues, tres o quatre posicions més endarrere o més endavant. El nombre d'unitats o passos que separen cada element de l'antecedent o consegüent amb el que es vol relacionar rep el nom de *retard* (en anglès, *lag*) i s'incorpora a la freqüència de transició com un tercer subíndex  $K$ . L'expressió completa d'una freqüència de transició és, doncs,  $f_{ijk}$ . Si un parell de categories  $AB$  han estat associades com a antecedent i consegüent immediats (retard 1), 17 vegades en un text, aleshores  $f_{AB1} = 17$ .



Semblantment al procediment explicat en el cas de les freqüències de contingència, ara aquestes freqüències de transició també poden ser transformades en freqüències relatives de transició  $p_{j|i}$ , les quals, al seu torn, poden servir d'estimadors de probabilitat condicionada. Aquestes probabilitats són les que ens permeten establir pròpiament les regularitats de la seqüència.

### Exemple

Suposem que la categorització i buidat d'un text, a partir d'un sistema de categories  $S = \{A, B, C, D\}$  forneix una seqüència com aquesta:

$D, A, B, C, D, A, B, C, B, A, B, A, C, A, D, C, B, D, B, C, B, D$

Llavors, amb retard 1 (el símbol del qual ometem),

$f_{AB} = 3; f_{BA} = 2; f_{CB} = 3; f_{BC} = 3; f_{AD} = 1; f_{DA} = 2$ , etc.

En forma tabular i més sintètica, posem les quatre categories tant en files com en columnes i en les caselles podem consignar la freqüència de transició de cada parell. Naturalment, podríem cercar associacions entre tríos o quartets, però aquí ens limitarem als parells. Així, com ja sabíem,  $AB$  ha aparegut 3 vegades, mentre que  $BA$  només 2. Els totals de fila  $f_{i+}$  equivalen al nombre de vegades que cada categoria ha estat en posició antecedent i els totals de columna  $f_{+j}$  al nombre de vegades que cada categoria ha figurat en posició conseqüent. En aquesta taula, la diagonal principal és 0 perquè les repeticions  $AA$ ,  $BB$ , etc., no estan permeses. En una entrevista en què cada intervenció d'entrevistador o entrevistat donés lloc a una sola categoria, no hi podria haver dues respostes, ni dues preguntes, seguides. Però el contrari és perfectament factible i veurem després taules en què aquesta diagonal té valors distints de 0.

		Conseqüent (retard 1)				
	$f_{ij}$	A	B	C	D	Total ( $f_{i+}$ )
A		0	/// 3	/	/	5
B		// 2	0	/// 3	// 2	7
Antecedent	C	/	/// 3	0	/	5
(Retard 0)	D	// 2	/	/	0	4
Total	( $f_{+j}$ )	5	7	5	4	21

És clar que les freqüències de transició, per si mateixes, no ens diuen gran cosa. Que dues categories hagin estat enllaçades un cert nombre de vegades no informa sobre regles d'ordre de la cadena, atès que aquesta freqüència depèn de la freqüència global de cada element en solitari dins de la cadena esmentada. Tampoc no ens serviria una probabilitat absoluta o incondicionada. Per exemple, podríem calcular la probabilitat del parell  $CB$ , que seria  $p_{CB} = 3/21$ , és a dir, la seva freqüència dividida pel nombre total de parells. Però aquest resultat no ens informaria sobre la fortalesa de l'enllaç entre  $C$  i  $B$ , perquè també estaria supeditat a la freqüència global de  $C$  i de  $B$  per separat.

El que ens convindria seria saber amb quina probabilitat es dona  $B$  suposant que abans s'hagi donat  $C$ . Aquesta probabilitat pot ser estimada com a freqüència relativa de transició, dividint el nombre de vegades que ha figurat  $B$  darrere de  $C$  pel nombre de vegades que ha aparegut  $C$  en posició antecedent. Aquest darrer nombre no sempre és igual que la freqüència total de  $C$  en la cadena, ja que  $C$  podria estar en l'últim lloc de la seqüència i llavors ja no seria estrictament un antecedent. En definitiva,  $p_{B|C}$  estimaria la probabilitat

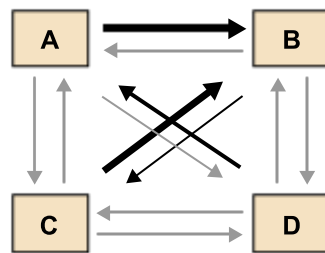
que en la cadena es produeixi  $B$  sempre que abans s'hagi produït  $C$ , i es calcularia pel quocient  $f_{CB}/f_{C+} = 3/5$ .

Com podeu apreciar, aquest càlcul es pot fer directament a partir d'una taula com l'anterior, ja que el numerador del quocient és la freqüència de cada parell i el denominador el total de fila corresponent. Llavors obtindríem una taula de probabilitats condicionades com la que segueix:

		Consegüent (retard 1)				
		$f_{j/i}$	A	B	C	D
Antecedent		A	0	0,60	0,20	0,20
		B	0,29	0	0,43	0,29
		C	0,20	0,60	0	0,20
(Retard 0)		D	0,50	0,25	0,25	0

Aquestes xifres ja ens subministren informació més aprofitable sobre els lligams entre categories. Així, veiem que el lligam entre  $A$  i  $B$  és molt més fort que el que hi ha entre  $B$  i  $A$  o que l'enllaç entre  $C$  i  $B$  és molt més consistent que el que hi ha entre  $D$  i  $B$ . Ultra això, la regla d'associació que ens forneix una probabilitat condicionada permet construir un model de relacions en el corpus o fragment analitzat. A fi de trobar aquest model, l'aproximació correcta seria la de trobar la significació estadística d'aquests enllaços aïllats, punt que tocarem en un exemple proper. Ara, però, de moment, per tal d'il·lustrar aquesta possibilitat, recorrerem a una opció no tan correcta però que ajuda tanmateix a visualitzar el sistema de relacions que expressen les probabilitats.

L'opció consisteix a marcar un mínim per damunt del qual es tenen en compte les relacions. En aquest exemple podria ser 0,4. Això vol dir que les associacions per damunt d'aquest valor serien incorporades al model, mentre que les altres no. Òbviament, els parells  $AB$ ,  $BC$ ,  $CB$  i  $DA$  serien els beneficiats. Si representéssim aquestes relacions mitjançant un diagrama de flux i cada interval de 10 punts (entre 0,40 i 0,49; entre 0,50 i 0,59; i entre 0,60 i 0,69) amb fletxes d'un gruix diferent, obtindríem un esquema com el de la següent figura.



Podem sospitar que la categoria  $B$  és candidata a ser un "atractor", per la fortalesa de la convergència de  $A$  i  $C$  cap a ella.  $A$  sembla que té un lligam moderat i específic amb  $D$ , que tendeix cap a ella. La relació de  $B$  amb  $C$  ja és de menor importància. En aquest esquema hem indicat en gris les relacions descartades per la seva feblesa.

No cal dir que aquest exemple és totalment insuficient amb vista a extreure'n conclusions. L'exercici que acabem de fer s'ha de dur a terme amb cadenes textuales molt més llargues i amb freqüències molt més altes de categories.

Per la via de l'anàlisi seqüencial podem arribar a saber, posem per cas, que en una entrevista a una mare cada pregunta sobre els seus fills tendeix a anar seguida, en primera o segona opció, de categories de resposta classificables com a "afectives" i, després, en tercera posició, per una negació de les anteriors. En aquest cas, la constatació seria merament descriptiva i abraçaria els retards 1, 2 i 3.

Per la via de l'anàlisi de contingències podem esbrinar que, en un diari adolescent, les anotacions que obren cada entrada del diari solen ser expressives quan el text continua després narrant conflictes o drames (familiars, escolars, sentimentals), mentre que s'inclinen a la descripció quan els esdeveniments que omplen la pàgina d'aquell dia són més aviat neutres.

Les mesures de contingència i seqüencials poden arribar a coincidir quan els elements o unitats relacionades són antecedents i consegüents.

## 2. Exemples desenvolupats d'anàlisi de contingut

### 2.1. Exemple 1: Recomptes de freqüència en general

En l'anàlisi de contingut és important saber com es distribueixen les freqüències de les categories al llarg del text i com s'associen a les altres categories o als diferents contextos d'ocurrència.

Us proposem un relat etogènic que il·lustra aquesta perspectiva documental. Es tracta d'un paràgraf en el qual un marit propietari d'un pis expressa la seva estranyesa respecte a l'incompliment de les expectatives de venda de l'immoble, expectatives que, tant ell com la seva muller, s'havien fet després de la visita d'una parella de suposats compradors.

"Aquesta gent, quan l'agència ens va avisar que ens els enviaven, van venir de seguida. Gairebé no vam tenir temps d'endreçar el pis. Semblaven molt interessats i semblaven bons compradors. Al principi jo hauria jurat que estaven a punt de decidir-se. De fet fins i tot van parlar (ells dos) de canviar algunes coses del pis. Van comentar que les finestres les farien totes noves. Vam pensar que tenien diners llargs i que serien bons pagadors. Van estar-s'hi molta estona, i quan van marxar semblava que ja s'hi trobaven com a casa. Jo vaig dir a la Cíntia (la meva dona) que truqués a l'agència perquè els collés i estiguessin a l'aguait. Però no van dir res en tot el dia. Ni l'endemà, ni la resta de la setmana. No enteníem res, ni l'agència, tampoc. Al final els vaig trucar personalment i a sobre em penso que els va molestar. Ja no els interessava. Qui ho entén!"

Arranjat en línies numerades:

- 1 Aquesta gent, quan l'agència ens va avisar que ens els enviaven, van venir de seguida. Gairebé
- 2 no vam tenir temps d'endreçar el pis. Semblaven molt interessats i semblaven bons
- 3 compradors. Al principi jo hauria jurat que estaven a punt de decidir-se. De fet fins i tot van parlar
- 4 (ells dos) de canviar algunes coses del pis. Van comentar que les finestres les farien totes noves.
- 5 Vam pensar que tenien diners llargs i que serien bons pagadors. Van estar-s'hi molta estona, i
- 6 quan van marxar semblava que ja s'hi trobaven com a casa. Jo vaig dir a la Cíntia (la meva
- 7 dona) que truqués a l'agència perquè els collés i estiguessin a l'aguait. Però no van dir res en tot el dia.
- 8 Ni l'endemà, ni la resta de la setmana. No enteníem res, ni l'agència, tampoc. Al final els vaig trucar
- 9 personalment i a sobre em penso que els va molestar. Ja no els interessava. Qui ho entén!

Adoneu-vos primer de tot que el text presenta dues parts ben diferenciades: una en la qual el narrador expressa o creu confirmar diverses expectatives de venda, i una altra, a partir de la línia 7, en la qual l'autor palesa el seu desencís en veure aquestes expectatives frustrades, introduït amb la conjunció adversativa *però*. Això es pot copsar a simple vista, tot i que els instruments d'anàlisi també ho haurien de palesar.

Les unitats de registre podrien ser **mot**s o frases en context d'oracions. Els mots són indicadors categorials que prenen tot el seu sentit en el context de les oracions. Les oracions (oblidant ara les línies) estan delimitades o definides

#### Vegeu també

Recuperem per a aquesta exemplificació el relat etogènic que il·lustrava aquesta mena de perspectiva documental en el mòdul 5 de l'assignatura *Mètodes d'investigació qualitativa*.

operacionalment, a) quan són simples, pel fet d'estar compreses entre punts i seguit consecutius, o entre inici i punt i seguit, o entre punt i seguit i punt final; b) quan estan encabides dins d'una coordinada copulativa, han d'estar compreses entre un punt i el connector *i* o viceversa (altres tipus d'oracions no es veuen aquí).

Les categories estan informades per diferents mots o frases –els seus indicadors– i són de dos tipus: categories d'expectativa confirmada,  $E^+$ , i categories d'expectativa frustrada,  $E^-$ .

Així, les frases *molt interessats* i *bons compradors*, en els contextos respectius de *semblaven molt interessats* i *semblaven bons compradors* es codifiquen com a  $E^+$ . *No van dir res* en el context de *però no van dir res en tot el dia* es codifica com a  $E^-$ . Etc.

Tot comptat arribem a la llista següent de codificacions i, per tant, de categories (indicadors en cursiva):

- Aquesta gent, quan ens va avisar l'agència que ens els enviaven, *van venir de seguida*:  $E^+$
- Gairebé *no ens van donar temps* a posar el pis en ordre:  $E^+$
- Semblaven *molt interessats*:  $E^+$
- semblaven *bons compradors*:  $E^+$
- Al principi jo hauria jurat que *estaven a punt de decidir-se*:  $E^+$
- De fet fins i tot van parlar (ells dos) de *canviar algunes coses del pis*:  $E^+$
- Van comentar que les *finestres les farien totes noves*:  $E^+$
- Vam pensar que *tenien diners llargs*:  $E^+$
- (vam pensar) que *serien bons pagadors*:  $E^+$
- Van estar *molta estona* a casa:  $E^+$
- quan van marxar semblava que ja *s'hi trobaven com a casa seva*:  $E^+$
- Jo vaig dir a la Cíntia (la meva dona) que truqués a l'agència *perquè els collés*:  $E^+$
- (Jo vaig dir a la Cíntia (la meva dona) que truqués a l'agència perquè) *estiguessin a l'aguait*:  $E^+$
- Però *no van dir res* en tot el dia:  $E^-$
- (...no van dir res en tot el dia) *(Ni) l'endemà*:  $E^-$
- (...no van dir res en tot el dia) *(ni) la resta de la setmana*:  $E^-$
- *No enteníem res*:  $E^-$
- l'agència *tampoc*:  $E^-$
- Al final *els vaig trucar personalment*:  $E^-$
- a sobre crec que *els va molestar*:  $E^-$
- Ja *no els interessava*:  $E^-$

- *Qui ho entén!:*  $E^-$

A partir d'aquesta codificació de categories podem emprar diverses mesures:

$$f_{E^+} = 13; f_{E^-} = 9; p_{E^+} = 13 / 22 = 0,59; p_{E^-} = 9 / 22 = 0,41$$

Podem establir direccionalitat mitjançant un índex senzill com

$$I = \frac{f_{E^+} - f_{E^-}}{f_{E^+} + f_{E^-}}; \text{ de manera que } I = \frac{13 - 9}{13 + 9} = 0,18$$

L'índex està comprès entre  $-1$  i  $1$ , és a dir:  $-1 > I > +1$ . Entenem, doncs, que en aquest segment de text, sense considerar l'ordre seqüencial, les expectatives positives dominen lleugerament sobre les negatives.

Quant a l'índex de diversitat  $I_d$ , referit a categories, és fàcil de computar, atès que només tenim 2 categories i ja coneixem el nombre total de categories comptabilitzades,  $f_{E^+} + f_{E^-} = 13 + 9 = 22$ . Així  $I_d = 2/22 = 0,09$ .

Pel que fa a l'ordenació seqüencial veiem, a vista d'ocell, que aquest paràgraf presenta una organització molt rígida i dicotomitzada. Totes les categories  $E^+$  s'agrupen en les 8 primeres línies, mentre que les  $E^-$  es concentren en les 3 finals. Això es deu a la retòrica de "fer-se la víctima" que utilitza el narrador i que consisteix a insistir molt al començament en la seva confiança en un curs d'esdeveniments positius (per a ell) per a passar després, abruptament, a exposar la seva desil·lusió terminal. Així el lector rep al final la mateixa "dutxa d'aigua freda" que ha rebut ell. És obvi que l'anàlisi ha d'anar orientat a l'autor o emissor del missatge, que deixa conspicuament la seva empremta en el text.

Vista aquesta ordenació tan rígida, el sentit comú qualitatiu ens diu que sobra fer cap còmput de freqüències de transició. No obstant això, per a amortitzar l'exemple, vegeu com es computarien aquestes freqüències  $f_{ij}$  mitjançant una taula de contingència seqüencial:

	Categoria consegüent			
		$E^+$	$E^-$	Total
Categoria precedent	$E^+$	12	1	13
	$E^-$	0	8	8
	Total	12	9	21

La lectura d'aquesta taula ens diu que la transició o parell seqüencial  $E^+E^+$ , s'ha produït 12 vegades,  $E^+E^-$ , 1 vegada,  $E^-E^+$ , 0 vegades i  $E^-E^-$ , 8 vegades. L'única transició  $E^+E^-$  articula les dues parts del text a les quals fèiem al·lusió més amunt.

Les probabilitats condicionades serien

0,92	0,08
0	1

Aquestes, per si calgués, ens diuen que, en aquesta mostra, és impossible una transició  $E^-E^+$  mentre que, amb  $E^-$  com a antecedent, és obligada la transició cap a  $E^-$ , i que la transició  $E^+E^+$  és altament probable, mentre que la  $E^+E^-$  ho és molt poc. Però, com ja hem dit, les regles d'associació ja eren evidents en la taula original de freqüències.

En una anàlisi d'associacions o de relacions seqüencials, les probabilitats condicionals són les que proporcionen les regles d'aparició o d'ordre.

## 2.2. Exemple 2: Anàlisi de contingència i anàlisi seqüencial

Les anàlisis de contingència i les seqüencials es basen en la valoració estadística de residuals, és a dir, de les diferències entre les freqüències o probabilitats d'associació observades i les esperades.

Suposem ara que ens enfrontem amb les transcripcions de les respostes d'un pacient anorèctic a les preguntes del terapeuta. Es tractaria ara d'esbrinar les associacions de dues úniques categories, "Crítica" i "No crítica" (s'entén que la crítica és negativa), corresponents a segments de tipus oració, amb dos possibles contextos d'ocurrències corresponents a paràgrafs (marcats per pauses orals), en què el tema dominant era, o bé la figura de la mare, o bé la figura del pare. Per tant, es tractaria d'una anàlisi de contingència o de concurrència, i caldria conèixer la distribució de les freqüències de les dues categories entre els dos contextos previstos. La taula de contingència seria la següent:

	Contextos d'ocurrència			
		Mare (M)	Pare (P)	Total
Categories	Crítica (C)	27	4	31
	No crítica (NC)	18	11	29

	<b>Contextos d'ocurrència</b>			
	Total	45	15	60

Davant d'aquesta distribució de freqüències, una primera indagació que hauríem de fer és si pot ser deguda a l'atzar o no, amb un determinat risc d'error. Com sabeu, aquesta informació la podem derivar de la prova de  $\chi^2$ .

Per a aplicar la fórmula, el primer que hem de fer és calcular les freqüències esperades si la distribució per caselles hagués estat a l'atzar. Amb aquest objectiu, entenem que els quatre esdeveniments  $C$ ,  $NC$ ,  $M$  i  $P$  són independents i que, per tant, la seva trobada, per parells, en cada una de les quatre caselles depèn d'un producte de probabilitats. Així, la probabilitat que es doni l'associació  $CM$  és igual que la probabilitat que es doni  $C$  multiplicat per la probabilitat que es doni  $M$ . Per consegüent,  $p_C \times p_M = 31/60 \times 45/60 = 0,3875$ . Lògicament, a fi d'estimar la freqüència esperada amb aquesta probabilitat cal multiplicar-la pel nombre de casos observats, és a dir, 60. En definitiva, la freqüència esperada de  $CM$  és  $60 \times 0,3875 = 23,25$  (aquesta fórmula se sol presentar mitjançant la recepta mecànica: total de files  $\times$  total de columnes / total general).

Fent, doncs, aquests càlculs, obtindrem la taula següent de freqüències esperades:

	<b>Contextos d'ocurrència</b>			
		Mare ( $M$ )	Pare ( $P$ )	Total
<b>Categories</b>	Critica ( $C$ )	23,25	7,75	31
	No critica ( $NC$ )	21,75	7,25	29
	Total	45	15	60

Substituint en la fórmula de  $\chi^2$ ,

$$\sum_i \frac{(O_i - E_i)^2}{E_i}$$

El numerador és el *residual* de les diferències entre les freqüències observades i les esperades per casella, elevat al quadrat. Tindrem que

$$\chi^2 = [(27 - 23,25)^2 / 23,25] + [(4 - 7,75)^2 / 7,75] + [(18 - 21,75)^2 / 21,75] + [(11 - 7,25)^2 / 7,25] = 0,60 + 1,81 + 0,65 + 1,94 = 5$$

#### Vegeu també

Heu de conèixer la prova de  $\chi^2$  de l'assignatura *Tècniques d'anàlisi de dades quantitatives*.



La significació d'aquest valor s'ha de jutjar amb 1 grau de llibertat, ja que  $\eta = (2 - 1) (2 - 1) = 1$ . Vistes les taules trobem que és estadísticament significatiu amb un risc de 0,05. De manera que podem concloure que, tal com sospitàvem, la distribució de freqüències obeeix alguna regla o model propis de les verbalitzacions de la pacient en situació terapèutica; no és deguda a l'atzar.

Conegut aquest punt, podem recular a la taula inicial i fer-ne una lectura fàcil. Sabem que:

$$f_C = 31; f_{NC} = 29; f_M = 45; f_P = 15.$$

I, a més, tocant a freqüències relatives,

$$p_C = 31/60 = 0,52; p_{NC} = 29/60 = 0,48; p_M = 45/60 = 0,75; p_P = 15/60 = 0,25.$$

Mitjançant aquestes senzilles proporcions corroborem el que les freqüències ja ens deien, és a dir, que les categories de "Crítica" i "No crítica" pràcticament es reparteixen al 50%, i que hi ha tres vegades més referències a la mare que al pare, independentment de com es distribueixen les crítiques. També podríem establir que les probabilitats absolutes de dirigir crítiques o no dirigir-les, a pare i mare, són, respectivament,  $p_{CM} = 27/60 = 0,45$ ;  $p_{CP} = 4/60 = 0,07$ ;  $p_{NCM} = 18/60 = 0,3$ ;  $p_{NCP} = 11/60 = 0,18$ .

Com ja hem remarcat unes pàgines abans, aquestes xifres permeten estimar amb quina probabilitat es donaran els parells d'associacions en aquest exemple, però no ens permeten inferir *una regla* d'associació concreta. Amb aquesta finalitat, haurem de calcular probabilitats condicionades: les probabilitats que, quan que la pacient parli del pare o de la mare, les seves frases estiguin acompanyades, o no, d'una crítica negativa. Recordeu que aquestes probabilitats condicionades s'expressen com  $p_{M/C}$ ,  $p_{M/NC}$ ,  $p_{P/C}$  i  $p_{P/NC}$  i es computen dividint el valor de freqüència observada de cada casella pel total de la fila corresponent. Així,

	Mare (M)	Pare (P)
Crítica (C)	$27/31 = 0,87$	$4/31 = 0,13$
No crítica (NC)	$18/29 = 0,62$	$1/29 = 0,38$

Ara, en comptes de recórrer a un criteri arbitrari com el que hem emprat en l'exemple del subapartat 2.1, estimarem la significació de cada una d'aquestes quatre probabilitats mitjançant la  $z$  binomial.

Ens servirem de la modalitat de computació que es basa directament en probabilitats, i no en freqüències.

#### Vegeu també

Recordeu que heu estudiat la  $z$  binomial en l'assignatura *Anàlisi de dades*.

$$z = (\text{probabilitat observada} - \text{probabilitat esperada}) / \text{desviació típica}$$

El numerador és també un residual, ara entès com a diferència de probabilitats.

Si designem per  $I$  la categoria o condició i per  $J$  el context associat a aquesta condició, la probabilitat observada serà  $p_{JI}$ , el que és el mateix, la probabilitat condicionada que consta en la taula anterior.

La probabilitat esperada per a cada casella s'estima a partir del valor de  $p_j$ , és a dir, de la probabilitat que es produeixin cada un dels dos contextos previstos independentment que s'associïn a una o altra categoria. Així, en la primera casella, corresponent a l'associació  $CM$ , la probabilitat esperada que es doni  $M$  és  $p_M = 45/60 = 0,75$  (la seva freqüència relativa), atès que fem l'estimació sobre la base del nombre total d'ocurrències de  $M$  respecte al total de 60, i sense tenir en compte la concentració major de freqüències en la casella  $CM$ .

### Desviació típica hipergeomètrica

De les diverses expressions de  $z$  utilitzem aquí la hipergeomètrica, la menys conservadora, en un format adaptat a taules de contingència. Vegeu: R. Bakeman i J. M. Gottman (1989). *Observación de la interacción: introducción al análisis secuencial*, pp. 196-197. Madrid: Morata.

La desviació típica de  $z$  és la següent:

$$\sqrt{\frac{p_J (1 - p_J)(1 - p_I)}{N p_I}}$$

Acabem d'explicar com es computa  $p_{ji}$  per tant  $1 - p_j$ . L'altra probabilitat que apareix en la fórmula,  $p_i$ , correspon a l'altra cara de la moneda: a la probabilitat que aparegui cada una de les dues categories previstes independentment que s'associïn a un o altre context. Exemplificant-ho de nou amb la primera casella, aquesta probabilitat és  $p_C = 31/60 = 0,5166 @ 0,52$  (de nou, la seva freqüència relativa).

Si fem la totalitat del càlcul de  $z$  per a la primera casella tenim que:

$$\text{Probabilitat observada} - \text{probabilitat esperada} = 0,87096 - 0,75 = 0,12096$$

$$\sqrt{\frac{0,75 (1 - 0,75)(1 - 0,5166)}{60 \times 0,5166}} = 2,24$$

Aquesta  $z$  és significativa, ja que supera el valor crític de 1,96 amb un risc d'error menor de 0,05. A més, és positiva; en conseqüència, l'associació corresponent  $CM$  ha d'interpretar-se de la mateixa manera;  $C$  i  $M$  tendeixen a associar-se, "s'atreuen" en tant que categoria i context d'ocurrència.

Calculant la resta de  $z$  pel mateix procediment, obtindríem quatre valors que fem constar a la mateixa taula:

	<b>Mare (M)</b>	<b>Pare (P)</b>
<b>Critica (C)</b>	2,24	-2,24
<b>No critica (NC)</b>	-2,24	2,24

Traduint el que és obvi: la pacient té una tendència clara a criticar la mare i no criticar el pare en situació de teràpia o, a l'inrevés, inhibeix qualsevol lloança o comentari neutre respecte a la mare i qualsevol comentari crític respecte al pare.

#### **Nota**

Les tècniques del tipus que acabem d'explicar o d'altres semblants serveixen tant per a l'anàlisi d'associacions en forma de concurrències o coocurrències, com per a l'anàlisi d'associacions seqüencials entre antecedents i conseqüents. Ara bé, aquest exemple ha estat relativament còmode de presentar perquè només hem considerat dues categories i dos contextos. En el subapartat 2.1 hem desenvolupat l'exemple del relat etogènic en el qual rastrejàvem els lligams seqüencials de solament dues categories. Potser no cal afegir que en la pràctica real de l'anàlisi el nombre d'elements que integren una taula de contingència com les que hem vist pot ser molt més gran.

### **2.2.1. Exemple d'anàlisi seqüencial amb una taula de contingència complexa**

Afegim, per acabar, un exemple d'anàlisi seqüencial recolzat sobre una taula amb vuit categories de referència. Es tracta ara del diari d'un nen de 13 anys, escrit durant el curs, i analitzat d'oració en oració (entre punts seguits o entre punts a part). Les categories són de caire pragmàtic i es basen en la persona a la qual fa referència l'oració en tant que unitat textual, sobre el criteri que figuren com el seu subjecte gramatical.

La taula de freqüències observades en el diari, entenent que la successió antecedent-conseqüent té retard 1, és la següent:

	Consegüents									
		Mestre	Monitor	Mare	Pare	Ell mateix	Germà	Avi	Àvia	Total
Antecedents	Mestre	3	0	0	0	4	0	0	2	9
	Monitor	0	0	0	1	0	1	0	0	2
	Mare	0	0	161	38	68	1	2	20	290
	Pare	1	0	11	11	184	4	2	22	235
	Ell mateix	2	1	81	140	541	15	6	108	894
	Germà	0	0	7	6	12	4	1	7	37
	Avi	0	1	2	3	1	1	6	10	24
	Àvia	3	0	33	36	84	11	7	37	211
	Total	<b>9</b>	<b>2</b>	<b>295</b>	<b>235</b>	<b>894</b>	<b>37</b>	<b>24</b>	<b>206</b>	<b>1.702</b>

Calculem les probabilitats condicionades en cada casella, que ja ens orientaran respecte als nexes seqüencials que uneixen les categories entre si.

	Consegüents									
		Mestre	Monitor	Mare	Pare	Ell mateix	Germà	Avi	Àvia	Total
Antecedents	Mestre	0,33	0,00	0,00	0,00	0,444	0,00	0,00	0,222	<b>1</b>
	Monitor	0,00	0,00	0,00	0,50	0,00	0,50	0,00	0,00	<b>1</b>
	Mare	0,00	0,00	0,55	0,131	0,235	0,003	0,007	0,069	<b>1</b>
	Pare	0,004	0,00	0,047	0,047	0,783	0,017	0,009	0,094	<b>1</b>
	Ell mateix	0,002	0,001	0,091	0,157	0,605	0,017	0,007	0,121	<b>1</b>
	Germà	0,0	0,00	0,189	0,162	0,324	0,108	0,027	0,189	<b>1</b>
	Avi	0,0	0,042	0,083	0,125	0,042	0,042	0,25	0,417	<b>1</b>
	Àvia	0,014	0,00	0,156	0,171	0,398	0,052	0,033	0,175	<b>1</b>

I calculem les  $z$  mitjançant el mateix algorisme que abans. Ara, per a assegurar-nos més contra el risc que la vegada anterior fixarem un criteri de significació de 2,58, corresponent a un risc  $p < 0,01$ . Els valors significatius estan marcats en negreta.

	Conseqüents								
		Mestre	Monitor	Mare	Pare	Ell mateix	Germà	Avi	Àvia
Antecedents	Mestre	13,6	-0,1	-1,4	-1,2	-0,5	-0,4	-0,4	0,9
	Monitor	-0,1	0,0	-0,6	1,5	-1,5	4,6	-0,2	-0,5
	Mare	-1,4	-0,6	18,9	-0,4	-10,9	-2,3	-1,1	-3,0
	Pare	-0,2	-0,6	-5,5	-4,4	8,5	-0,5	-0,8	-1,4
	Ell mateix	-1,8	-0,1	-9,5	2,3	6,9	-1,5	-2,7	0,0
	Germà	-0,4	-0,2	0,3	0,4	-2,5	3,6	0,7	1,3
	Avi	-0,4	5,8	-1,2	-0,2	-4,8	0,7	9,9	4,5
	Àvia	1,9	-0,5	-0,7	1,5	-4,0	3,2	2,5	2,6

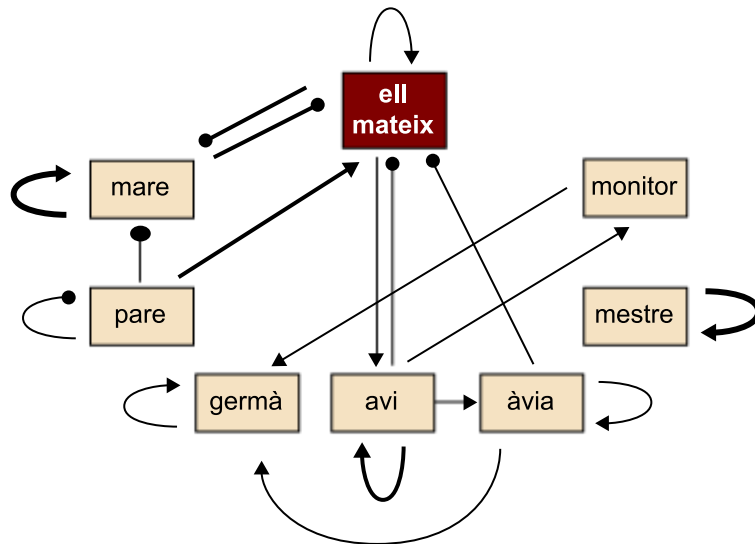
## Activitat

Davant de la taula de les  $z$ , i ja amb tarannà plenament qualitatiu, tracteu d'esbrinar quines són les regles que influencien la concatenació de categories en el diari del nen. Així, per iniciar l'anàlisi, és força evident que qualsevol referència al mestre tendeix a estar seguida d'una altra referència al mestre segons un enllaç que caldria estudiar més per menut. O que quan parla d'ell mateix, l'oració sol anar precedida per comentaris al voltant del pare o sobretot d'ell, però no de la mare ni dels avis, els quals sembla que s'inhibeixen davant d'aquesta categoria (o a l'inrevés: les referències a ell mateix tendeixen a suprimir-se després de les escrites envers la mare o els avis). Podeu continuar aquest exercici a partir del valor "cimentador" entre categories successives que posseeixen les  $z$  positives (i d'acord amb la seva magnitud), o del valor "inhibidor" de les  $z$  negatives.

Podeu tractar de dibuixar igualment un diagrama de flux a partir de les probabilitats condicionades amb valors significatius de  $z$ , seguint la pauta del que hem dibuixat unes pàgines abans.

## Solució

Traslladant simplement les  $z$  significatives al gràfic obtindríem un dibuix com el de la figura següent, que és un model icònic de les regles que ordenen el text amb retard 1, prenent com a elements les categories utilitzades en aquest exemple. Les  $z$  entre 2,58 i 7 han estat representades amb un traç més fi; les  $z$  entre 7,1 i 12, amb una línia una mica més gruixuda; finalment les  $z$  superiors a aquest valor tenen el traç més gruixut. Les relacions positives es representen amb fletxes, segons la convenció més habitual; les relacions negatives o d'inhibició amb baquetes.



Cal destacar que, com ja hem remarcat, qualsevol referència al mestre és doble en el sentit antecedent-consegüent i que, curiosament, el mateix passa amb la mare. Les successions mare-nen (ell mateix) o nen-mare estan força inhibides. Les referències al pare no solen estar seguides per una segona referència, a diferència de les de l'avi, però el vincle pare-nen és bastant consistent en el text. Les referències del nen a ell mateix no tenen cap seguiment significatiu: darrere de les referències que fa a si mateix no introdueix de manera sistemàtica cap al·lusió a altres membres del seu cercle, tot i que sí que les continua sovint amb al·lusions a ell mateix, etc.

Naturalment, aquest ball de xifres només és el suport tècnic sobre el qual l'analista ha d'explorar el substrat d'aquests vincles, gràcies a un aprofundiment més gran dels continguts dels segments implicats en els enllaços. Caldria saber, posem per cas, per què les cites del monitor d'esport s'associen al germà. Se'ns poden acudir moltes hipòtesis pel que fa a això, però qualsevol hauria de ser validada sobre continguts del diari o sobre dades complementàries obtingudes del nen en qüestió.

Una anàlisi de contingència o seqüencial pot mostrar la presència de lligams positius o negatius entre categories, o entre categories i contextos, segons que les unitats analitzades tendeixin a associar-se o a repel·lir-se.

### 2.3. Exemple 3: Anàlisi avaluativa

L'anàlisi avaluativa implica el judici de l'analista ja en la ponderació mateixa de les freqüències.

En aquest darrer exemple, tractarem de treure profit de les freqüències ponderades. La ponderació pot ser una de les portes d'entrada de l'empremta qualitativa en la mesura i l'anàlisi, atès que implica una assignació de valor, un judici d'intensitat, de direcció o de totes dues coses per part de l'analista i, per tant, una interpretació subjectiva per part d'aquest en el sentit metodològicament més favorable del terme, el d'amortitzar l'experiència i els coneixements del científic en l'avaluació del fenomen estudiat. En els exemples anteriors,

aquesta experiència i coneixements s'aplicaven a les xifres generades per les operacions estadístiques o probabilístiques; en la ponderació, en canvi, ja deixen la seva empremta en l'operació mateixa de mesura.

Un dels terrenys en què s'utilitza amb més sentit la ponderació de freqüències és en l'anàlisi avaluativa, a la qual ja hem al·ludit en enumerar les línies principals de treball de l'anàlisi de contingut. Aquesta línia palesa característiques particulars, tant en el seu llenguatge tècnic com en els seus punts d'aplicació. Davant d'un text, l'investigador avaluatiu només s'ocupa d'aquelles oracions que tenen un significat valoratiu, i és respecte a aquest significat que du a terme l'avaluació.

En general:

- L'avaluació es refereix a actituds o opinions valoratives en relació amb certs "objectes" o temes presents en el text, les quals poden estar nodrides per emocions, interessos, ideologies, etc. Aquests objectes se solen simbolitzar en les operacions d'anàlisi mitjançant OA (*objectes d'actitud*).
- Les actituds es posen de manifest en unitats textuais de significació, habitualment oracions o frases. Aquestes unitats se simbolitzen mitjançant EV, és a dir, *expressions valoratives*. *La psicologia no té gaire de científica* és una EV aplicada a l'OA de la psicologia. Però, naturalment, també podríem haver escollit *la psicologia és una disciplina científica*.
- Els OA i les EV es relacionen mitjançant connectors verbals, ja sigui amb la forma d'oracions substantives o simples, sense complements (com *l'àrbitre és un inútil*, *aquest jugador promet*), ja sigui amb la forma d'oracions completes, amb complements, (*l'àrbitre va xiular un penal injust contra els locals*, *un actor com aquest mereixia des de fa anys l'Oscar*). L'anàlisi també es pot centrar en oracions compostes, coordinades o subordinades, però llavors aquestes es descomponen de la manera que indicarem més avall. Cal puntualitzar igualment que la ponderació dels verbs, per una banda, i dels predicats o complements, per una altra, es fa per separat i després es multiplica, com també veurem.
- A partir dels exemples precedents, s'ha d'entendre que les valoracions es mouen en un espai dimensional bipolar, entre dos extrems: positiu-negatiu, a favor-en contra, alt-baix, etc. L'avaluació, doncs, té una direcció, dintre d'una polaritat o d'una oposició. Aquest és el primer component de l'avaluació.
- Finalment, dintre de l'oració hi ha d'haver elements gramaticals i semàntics que permetin ponderar efectivament la intensitat de l'actitud, per mitjà del judici que l'expressa. Aquest és el segon component de l'avaluació. És evident que *la psicologia és una disciplina científica*, *la psicologia és indubtablement una disciplina científica* i *la psicologia tal volta és una disciplina ci-*

### Lectures recomanades

L'anàlisi avaluativa enfonsa les seves arrels en l'obra del psicolingüista americà Charles E. Osgood. Un llibre ben conegut en l'àmbit de la docència acadèmica, en què trobareu un exemple intel·ligible d'aquesta anàlisi, és el de:

R. Mayntz, K. Holm i P. Hübner (1975). *Introducción a los métodos de la sociología empírica* (pp. 211-214). Madrid: Alianza.

També L. Bardin (1996). *El análisis de contenido*. Madrid: Akal.

*entífica* són enunciats que incorporen càrregues diferents d'opinió tocant a la fermesa de qui l'expressa o a la intensitat de l'actitud. En una pràctica força habitual, la ponderació es fa sobre una escala de 3 valors, positius o negatius segons la direcció en la qual s'apliquen. En aquest cas podríem assignar, en la direcció positiva, +1 a *la psicologia tal volta és una disciplina científica*, +2 a *la psicologia és una disciplina científica* i +3 a *la psicologia és indubtablement una disciplina científica*. Per contra, la frase abans citada de *la psicologia no té gaire de científica* mereixeria una ponderació -1, i frases com *la psicologia no és científica* i *la psicologia no té gens de científica* un -2 i un -3, respectivament.

La temporització del procés d'anàlisi és la següent:

1) En primer lloc, s'identifiquen els OA en el text. Ja hem indicat que l'anàlisi avaluativa no practica una exploració exhaustiva del text, sinó que es concentra en les unitats que poden ser considerades valoratives segons determinats criteris.

2) Després se substitueixen els OA per símbols "neutrals", habitualment lletres majúscules, amb la finalitat que el desconeixement de quin OA es tracta en cada cas impedeixi biaixos en la ponderació de l'analista. Així, la frase *la psicologia no té gens de científica* es transformaria en *X no té gens de científica*, i s'evita d'aquesta manera que psicòlegs de diferents tendències, o altres científics socials, es projectin en el contingut de la frase.

3) A continuació, es fa una llista amb totes les frases amb OA, i les convertim en enunciats més o menys estandarditzats a fi d'ajudar en la ponderació. Ja hem dit, que aquests enunciats són bàsicament de dos tipus: oracions substantives, intransitives o transitives però sense complements –en els quals l'OA es connecta solament amb un predicat nominal o un verb–, i oracions que porten complements amb diferents funcions gramaticals (directes, indirectes, circumstancials). Aquesta fase comporta també la descomposició de les oracions compostes en simples.

4) Per a cada OA s'elabora una "tarja de codificació" independent, en la qual es fan constar les ponderacions assignades a connectors verbals, per un costat, i a predicats o complements, per un altre. Com també hem dit, l'assignació se situa entre +3 i -3, segons la direcció de la valoració de l'OA. L'assignació depèn dels arguments que hi apliqui l'investigador. Una manera de dotar de consistència la ponderació és fer-la entre dos o més analistes, intentant després arribar a l'acord entre ells mitjançant tècniques que exposarem en apartats propers.

#### **Exemplificació de les fases de l'anàlisi avaluativa**

Il·lustrarem les facetes i fases de l'anàlisi amb un exemple minúscul. Atengueu al text següent, producte suposat d'un informe en el context d'un conflicte sociolaboral:



"Malgrat que el cap de secció, que intenta fer-se el just, va donar suport aparentment a Rovira en la seva denúncia insistent contra uns serveis de seguretat sospitosos, no confiava plenament en ell".

Substituïm els OA per símbols:

- Cap de secció = X
- Rovira = Y
- Serveis de seguretat = Z

Descomponem les oracions:

- X intenta fer-se el just.
- X va donar suport aparentment a Y en la seva denúncia.
- Y ha denunciat insistentment uns Z.
- Z són sospitosos.
- X no confiava plenament en Y.

Elaborem les targes de ponderacions:

OA	Verb	Ponderació (direcció i valor)	Predicat o complement	Ponderació (direcció i valor)	Producte total
X	intenta	+1	fer-se el just	+1	+1
X	va donar suport	+2	aparentment Y	+2	+4
X	no confiava	-1	plenament en Y	+3	-3
Y	ha denunciat	-3	insistentment uns Z	+3	-9
Z	són	+3	sospitosos	-2	-6

La càrrega de ponderació és responsabilitat de la sensibilitat al llenguatge de l'analista, tot i que decisions com les de les taules anteriors estan inspirades pel sentit comú. És justificable ponderar *intenta* amb +1 quan l'autor del text disposava de verbs molt més potents com *pretén*, *vol* o *es proposa*, que mereixerien un +2 o un +3. *Va donar suport* rep un +2 perquè també té espai semàntic per damunt, el que correspondria a connectors com *va promoure* o *va impulsar* o *va fer*, que rebrien un +3. *No confiava* es queda amb un -1, atès que hi ha verbs molt més contundents en la direcció negativa, com *vigilava*, *criticava* (-2) o *perseguia*, *atacava* (-3).

En la columna dels predicats i complements podríem esgrimir arguments semblants. *Fer-se el just* ha de restar en puntuació per sota de *serjust* (+3) o *actuar justament* (+2). *Aparentment* introduiria un matís reductor respecte a *decididament* o *incondicionalment*, però romandria per sobre de *tèbiament* o *fins a cert punt*. Etc.

### Nota

No podem ocultar que un exercici d'aquesta mena requereix un complet domini semàntic i pragmàtic del llenguatge, a banda d'estar immers en la cultura que ha produït el text, tot i que aquestes competències estan en certa mesura sancionades per l'acord o el consens entre analistes.

L'anàlisi avaluativa comporta la detecció dels *objectes d'actitud* en el text i de les *expressions de valoració* que permeten avaluar-los, habitualment de +3 a -3, en direcció positiva o negativa.

### 3. La qualitat de les dades en l'anàlisi de contingut

#### 3.1. El marc general de la qualitat de dades en l'anàlisi de contingut: recordatori

Els tres fronts en què es planteja la qualitat de les dades en l'anàlisi de contingut són els mateixos que planteja la metodologia qualitativa en general: validesa, fiabilitat i precisió.

##### 3.1.1. Fiabilitat, precisió, validesa i generalització

Ja hem tractat els punts capitals de la qualitat de les dades en la investigació qualitativa. L'únic tema que cal reprendre ara és el de la fiabilitat, el qual, de manera corresponent, és el més tractat en la literatura metodològica sobre anàlisi de contingut. I, ben a prop de la fiabilitat, ens referirem d'esquitllada a la precisió, dos conceptes que són només variants d'una mateixa manera d'entendre la qualitat de la informació obtinguda.

La **validesa**, en perspectiva qualitativa, remet a la definició de les categories d'anàlisi; en general és –convé que sigui– una validesa "de contingut", més que de "de constructe". Això significa que el problema de si realment mesurem o comptem el que volem per mitjà de les categories aplicades a l'anàlisi es pot girar del revés: el que mesurem o comptem és definit i comprès dintre de les categories; no pretenem començar la casa per la teulada; no volem estudiar el que pensa o imagina la gent, sinó el que verbalitza o escriu. Amb això no volem insinuar que l'analista renunciï a fer inferències, però aquestes es fan a partir dels registres codificats per les categories, mentre que les categories han d'estar arrelades al text.

Efectivament, en el cas de l'anàlisi de contingut el que hi ha dintre de les categories són segments concrets de text als quals posem una etiqueta i assignem –això sí– una significació més o menys arriscada, interpretant-los sobre la base d'una teoria, d'hipòtesis de treball, de codis culturals, etc. Tanmateix, com més ens allunyem de la concreció del text, més problemes de validesa haurem d'afrontar. Si definim una categoria d'"Agressivitat" a partir de verbs i oracions de contingut agressiu adreçats a persones o institucions (*pegar, matar, disparar, trinxar, destrossar, afusellar*, etc.) o a partir de la menció d'armes, instruments de tortura, presons, també amb implicació de subjectes socials, aleshores el significat d'aquesta categoria no plantejarà gaires problemes de validesa. Si no fem això i intentem justificar la definició de les categories en funció de continguts ocults o latents, o en funció d'un desxiframent de frases més enllà dels

#### Vegeu també

Els punts capitals de la qualitat de les dades en la investigació qualitativa foren tractats a bastament en el mòdul 6 de l'assignatura *Mètodes d'investigació qualitativa*. Allà s'establia el marc general en el qual cal considerar els problemes de qualitat de les dades en aquest àmbit metodològic, i s'assenyalaven les distincions necessàries entre fiabilitat, precisió i validesa, per un costat, i les qüestions relatives a la generalització, per un altre.

seu significat aparent, aleshores ens complicarem la vida. Certament –insísim– podem i hem de fer sovint inferències en profunditat; però aquestes han de fer-se sobre la base del registre o la mesura, no en l'acte mateix de registre o mesura. En definitiva, s'han de fer mitjançant la comparació, combinació i anàlisi de la informació categorial, tal com hem explicat en el mòdul "Les tècniques qualitatives en la selecció i tractament de la informació", dintre dels procediments de la teoria fonamentada. Ara bé, la informació atribuïda a cada una de les categories emprades en l'anàlisi, en si mateixa, hauria d'estar lligada al text de la manera més immediata possible.

A banda d'aquestes consideracions, cal no oblidar tampoc que l'investigador qualitatiu es permet l'opció de la *validació èmica*, tant fora com dintre de l'anàlisi de contingut. Els subjectes, grups o institucions que són autors dels textos, en el seu paper d'informants, poden contribuir a crear o fixar el sentit de les categories proposades per l'investigador. Aquesta és sens dubte una manera ben qualitativa d'atorgar validesa a les unitats d'anàlisi emprades, les categories, i per consegüent, a la freqüència, que constitueix la modalitat natural de mesura que permet quantificar-les.

Quant a la **generalització**, planteja problemes metodològics d'un abast que desborda de llarg l'àrea de l'anàlisi de contingut. Podríem dir que l'anàlisi de contingut té els mateixos problemes o condicions de generalització que qualsevol altra tècnica o metodologia qualitativa, ni més, ni menys.

La **fiabilitat** constitueix el vessant de la qualitat de les dades en què s'han desenvolupat més tècniques.

### 3.2. Els diferents enfocaments de la fiabilitat en l'anàlisi de contingut

En l'anàlisi de contingut, la fiabilitat es pot enfocar en les tres mateixes direccions que preveu la metodologia qualitativa en general: intraobservador, interobservadors i precisió.

En termes generals, la fiabilitat al·ludeix al grau en el qual s'agrupen, coincideixen o concentren dues o més mesures del mateix fenomen, tot i que aquesta idea remet inevitablement, d'una manera o altra, a la d'estabilitat o consistència de la mesura en qüestió.

#### Vegeu també

Tot el que es va discutir en la part final del mòdul 6 de l'assignatura *Mètodes d'investigació qualitativa* és directament aplicable a l'anàlisi de contingut, particularment el concepte de *cas* i les estratègies inductives que permeten generalitzar una constel·lació de casos.

També en consonància amb els plantejaments generals de la qualitat de dades en metodologia qualitativa, l'anàlisi de contingut pot entendre la fiabilitat de diferents maneres. En aquest context, no obstant això, es barregen les sistematitzacions i classificacions *ad hoc* que han fet autors com Krippendorff –sistematitzacions fetes específicament per a l'anàlisi de contingut– amb les sistematitzacions que, en un pla molt general (i que, per tant, abraça l'anàlisi de contingut com una tècnica més), fa dècades que duen a terme la metodologia observacional i la metodologia qualitativa. Com que les diferents terminologies i els matisos dels conceptes en una tradició i l'altra poden generar confusions fàcilment, exposarem les classificacions des dels dos punts de vista, encara que, en la pràctica, convergeixen en gran mesura (aspecte que també ens encarregarem de subratllar).

### 3.2.1. Dimensions de la fiabilitat en general (metodologia qualitativa i observacional)

La metodologia qualitativa i observacional estàndard distingeix tres tipus de fiabilitat. Recuperem aquí alguns paràgrafs de l'assignatura *Mètodes d'investigació qualitativa*, a fi de connectar aquesta tipologia general amb l'específica de l'anàlisi de contingut, dintre d'aquest apartat.

#### Fiabilitat intraobservador

- L'agrupació o coincidència de la mesura es valora sobre diferents judicis o enunciats d'un mateix observador, relatius al registre, a la mesura o a la interpretació.
- En conseqüència, aquests judicis o enunciats es donen al llarg del temps.
- Els diferents judicis o enunciats es comparen respecte al seu valor mitjà o respecte als elements comuns que comparteixen; mai respecte a un valor fix o patró de referència. Per tant, es pretén assolir l'agrupació de registres, mesures o interpretacions en aquesta dimensió, o reduir o eliminar-ne la variabilitat; no es pretén encertar un valor vertader.
- És típica –tot i que no exclusiva– de la psicometria, que l'ha explorada a fons en la direcció test-retest (la fiabilitat o estabilitat d'un test se sol estimar a partir de la correlació entre els resultats de dues –o més– utilitzacions successives de la prova, separades per un cert interval temporal).
- Formalment, aquesta concepció de la fiabilitat es representa mitjançant una sèrie de registres o mesures d'un mateix observador  $m_1, m_2, \dots, m_m$ , les quals han de tendir a ser iguals o no significativament diferents entre si. En un registre categorial qualsevol la fiabilitat intraobservador es manifestaria en la coincidència dels registres repetits que un mateix observador fes d'un mateix corpus de dades en ocasions successives.

#### Lectures recomanades

K. Krippendorff (1990). *Metodologia del anàlisi de contingut*. Barcelona: Paidós.

És un dels textos de referència essencials en l'anàlisi de contingut d'orientació quantitativa. Tot i que aquest autor és una cita obligada en qualsevol manual d'anàlisi de contingut, nosaltres no recollarem directament en ell, ja que la seva posició metodològica no és precisament qualitativa, tot i que no podem ignorar la seva opinió en relació amb alguns temes. Si sentiu curiositat per copsar un discurs sobre fiabilitat i validesa més a prop del tarannà qualitatiu, consulteu l'obra següent:

J. M. Kirk i M. Miller (1986). *Reliability and Validity in Qualitative research*. Beverly Hills: Sage.

Cal entendre que, en perspectiva qualitativa, les tècniques que exposarem són recursos o eines que l'investigador pot emprar opcionalment, en clau de complementarietat o d'oportunitat, però no formen el nucli de la metodologia. Com ja s'afirmava en l'assignatura *Mètodes d'investigació qualitativa*, la variació en els judicis d'un analista o la disparitat dels judicis de diferents analistes no impliquen necessàriament un llast en un procés de recerca qualitatiu; al contrari, poden contribuir a enriquir-lo. Això no exclou, en nom de l'eclecticisme i del sentit comú, que pugui ser útil disposar d'una valoració quantitativa d'aquesta variació o disparitat.

- Formalment, la fiabilitat intraobservador es pot representar com una sèrie de mesures al llarg del temps entre les quals qualsevol diferència ha de considerar-se error.

*Observador:*  $m_1$  .....  $m_2$ .....  $m_3$  ..... .....  $m_n$ ;

*Error =  $m_n - m_{n-k}$ :* és a dir, la diferència entre dues o més mesures dintre de la sèrie

### Fiabilitat interobservador

- L'agrupació o coincidència de la mesura es valora sobre diferents judicis o enunciatos de diferents observadors, judicis relatius a operacions de registre, mesura o interpretació.
- Aquests judicis o enunciatos es relacionen en el pla sincrònic, és a dir, en un mateix punt o tall temporal, punt on justament es busca la coincidència o acord. Malgrat tot, pot interessar un seguiment d'aquestes coincidències sincròniques al llarg del temps.
- Igual que en la fiabilitat intraobservador, els diferents judicis o enunciatos es comparen respecte al seu valor mitjà o respecte als elements comuns que comparteixen; mai respecte a un valor fix o patró de referència. Per tant, també en aquest cas es pretén assolir la màxima agrupació de registres, no un valor vertader de mesura.
- Ha estat tradicionalment utilitzada per la recerca observacional i de camp, generalment en la modalitat de dos observadors que, un cop garantida la seva equivalència com a instruments de registre, s'ocupen de sectors diferents de la recollida de dades. En aquest marc, es parla de concordança o acord entre observadors, els quals s'estimen, com de seguida detallarem, mitjançant proporcions d'acord i índexs probabilístics.
- En termes formals, hem representat aquesta concepció de la fiabilitat limitada a dos observadors. Així tenim dues sèries de mesures de dos observadors 1 i 2:  $m_{11}, m_{12}, \dots, m_{1n}; m_{21}, m_{22}, \dots, m_{2n}$ . Interessa que  $m_{11} @ m_{21}; m_{12} @ m_{22}$ , etc. Els valors obtinguts pels dos observadors en cada punt de temps 1, 2, ...,  $n$ , han de ser iguals o no significativament diferents.

*Observador 1:*  $m_{11}$  .....  $m_{12}$ .....  $m_{13}$  ..... .....  $m_{1n}$

*Observador 2:*  $m_{21}$  .....  $m_{22}$ .....  $m_{23}$  ..... .....  $m_{2n}$

*Comparacions:*  $m_{11} \leftrightarrow m_{21}; m_{12} \leftrightarrow m_{22}$ ; etc.

*Error:*  $m_{1n} - m_{2n}$

## Precisió

- L'agrupació o coincidència de la mesura es valora sobre judicis o enunciats procedents de *diferents observadors/analistes o de diferents sistemes de registre, un dels quals dóna el valor acceptat com a vertader*.
- Habitualment, aquests judicis o enunciats es relacionen en el pla sincrònic, és a dir, en un *mateix punt o tall temporal*, punt on justament es busca l'aproximació al valor de referència. Malgrat tot, pot interessar un seguiment de l'evolució d'aquestes aproximacions al llarg del temps (fins i tot en la perspectiva intraobservador o d'un sol observador que repeteix la mesura per atansar-se al valor vertader).
- Al contrari que en les dues modalitats anteriors, ara els diferents judicis o enunciats *es comparen respecte al valor o patró de referència* que proporciona un dels observadors o un dels registres. Reiterem, doncs, que es pretén que registres, mesures o interpretacions s'acostin al màxim possible a aquest patró.
- Ha estat sovint utilitzada per la recerca *observacional i de camp*. Les dues modalitats més corrents han estat 1) la de comparar el registre, mesura o interpretació d'un observador expert o estàndard amb els d'un observador no expert o fal·lible, i 2) la de comparar el registre, mesura o interpretació d'un observador fal·lible amb els obtinguts a partir de gravacions o de tractaments informatitzats del material. L'aproximació a aquests criteris s'estima a partir d'índexs semblants als emprats en la fiabilitat interobservadors.
- La representació formal és molt semblant a la de la fiabilitat interobservadors. L'única diferència amb el plantejament precedent resideix en el fet que ara no comparem dues sèries de valors de dos observadors equivalents o del mateix estatus, sinó dos registres de diferents estatus: un de fet per un expert, o de manera automàtica, i que es jutja lliure d'error (representat amb subíndex 0), i un altre subjecte a error (representat amb subíndex 1).

Observador 0:  $m_{01}$  .....  $m_{02}$ .....  $m_{03}$  ..... .....  $m_{0n}$

Observador 1:  $m_{11}$  .....  $m_{12}$ .....  $m_{13}$  ..... .....  $m_{1n}$

Comparacions:  $m_{01} \leftrightarrow m_{11}$ ;  $m_{02} \leftrightarrow m_{12}$ ; etc.

Error:  $m_{0n} - m_{1n}$

### 3.2.2. Dimensions de la fiabilitat en l'anàlisi de contingut

Com ja hem avançat, la metodologia especialitzada en anàlisi de contingut proposa respecte a la fiabilitat unes distincions molt semblants a les anteriors.

#### Estabilitat

L'estabilitat al·ludeix a la consistència de la representació de l'objecte d'estudi en diferents aplicacions d'un instrument de mesura o en diferents mesuraments per part d'un mateix analista. Per tant, aquesta consistència recolza sobre la igualtat de les mesures en qüestió o sobre les seves diferències minimitzades. En la pràctica, es procura que la desviació típica del conjunt de mesures fetes sigui al més petita possible. Se sobreentén que, en aquesta perspectiva, l'objecte d'estudi roman immutable durant les operacions de mesura i que, en conseqüència, qualsevol diferència entre les mesures successives d'aquest mateix objecte constitueixen errors de mesura, no variacions en l'estat de l'objecte.

Dit això, és obvi que l'estabilitat correspon, fil per randa, a la fiabilitat intraobservadors.

Tot i que en la definició d'estabilitat hem agregat algun matís més, la seva definició coincideix en l'essencial amb el concepte de *fiabilitat intraobservador*. És cert que l'aplicació de la idea d'estabilitat al buidat i anàlisi d'un text exigeix alguna puntualització, conceptual i tècnica. En efecte, la fiabilitat intraobservador demanava una comparació de mesures al llarg del temps: una mesura feta en un temps  $t_1$ , una mesura feta en un temps  $t_2$ , etc. Per a poder repetir una vegada i una altra la mateixa mesura amb l'objectiu de provar-ne l'estabilitat, l'observador de camp havia de gravar o filmar els subjectes, conductes o situacions mesurades si volia assegurar la invariabilitat dels fenòmens de referència (tret que, com en el cas dels tests clàssics, se suposés que certs trets del subjecte eren invariables). Altrament, en tornar a accedir a l'objecte o al fenomen, el seu estat ja hauria canviat. I, com acabem d'indicar, és important no confondre les variacions pròpies del fenomen amb les variacions degudes a errors de mesura.

Quan es tracta de textos, aquest problema no existeix. El text, per definició, és invariable, una vegada ha estat inclòs i registrat en el corpus que cal estudiar. El temps afectarà l'observador en el seu transcurs temporal i vital, però no el text, que hauria de restar immutable en el paper o en la pantalla.

#### Lectures recomanades

K. Krippendorff (1990). *Metodología del análisis de contenido*. Barcelona: Paidós.

Per a un bon resum de l'obra de Krippendorff podeu llegir l'obra següent:

E. López-Aranguren (1989). El análisis de contenido. A M. García Ferrando, J. Ibáñez i F. Alvira (Comp.), *El análisis de la realidad social. Métodos y técnicas de investigación* (2a. ed., pp. 383-414). Madrid: Alianza.

En definitiva, amb la tècnica d'anàlisi de contingut podem conèixer l'estabilitat d'una mesura repetint-la diverses vegades sobre el mateix corpus o sobre els mateixos segments de text, separant prudentment cada nova prova per a evitar efectes de pràctica. Després avaluarem la constància del resultat successiu o, en l'altre pol, la seva desviació típica.

## Reproduïbilitat

També batejada com a *objectivitat*, la **reproduïbilitat** es refereix al grau en el qual dos o més analistes, proveïts del mateix procediment de mesura o recompte, obtenen els mateixos resultats en mesuraments fets sobre el mateix material de registre textual, dins del mateix corpus. Els resultats es jutgen, llavors, **reproduïbles**.

Intuïtivament, tendim a assimilar ràpidament aquesta perspectiva amb la de la fiabilitat interobservadors.

En principi, doncs, la reproduïbilitat correspon a la fiabilitat interobservadors.

No obstant això, ara cal precisar els conceptes encara més que en el cas de la fiabilitat intraobservador. I això com a conseqüència que, en la literatura sobre anàlisi de contingut, aquesta idea d'objectivitat suscita molts malentesos, que intentarem desfer aquí.

Per començar, la comparació entre els resultats de diferents analistes que mesuren el mateix corpus s'ha d'entendre en unes determinades condicions. Aquestes exigeixen que la metodologia utilitzada en el cas comparat sigui la mateixa. És a dir, que el mostratge, segmentació, codificació, categorització i, naturalment, la mesura (habitualment la freqüència) siguin idèntics, i que això sigui garantit per la utilització dels mateixos protocols o plans d'anàlisi. Així, els únics factors que resten no controlats són els analistes, en tant que virtuals fonts d'error.

### Congruència

Quan la validesa convergent és plantejada específicament en textos sobre anàlisi de contingut se li sol donar el nom de *congruència*.

Sovint aquesta condició no apareix clara en molts manuals. Si comparem mesures o resultats de dos observadors diferents però que empren segmentacions, codis o categories diferents, aleshores no estarem fent la comparació en el pla de la fiabilitat –o reproduïbilitat– entre observadors, sinó, ras i curt, en el de la *validesa convergent*. En altres paraules: estarem comparant metodologies de re-



gistre i graelles d'anàlisi diferents amb la finalitat de comprovar si condueixen als mateixos resultats i conclusions; no estarem comparant dues mesures fetes amb el mateix protocol i tècnica de mesura, que és al que s'hauria de referir la idea de reproduïbilitat si la situem en el marc de la fiabilitat.

Per altra banda, l'exigència de sincronia temporal que, en el context de la fiabilitat interobservadors, s'imposava per a poder comparar en un punt de temps els resultats dels observadors, ara no té gaire sentit. Insistim de nou que el text, una vegada incorporat al corpus, no canvia i per això el moment temporal en què aquells observadors-analistes poden recomptar o classificar codis i categories no té gaire importància. Si un analista fa el buidat del text avui i l'altre al cap d'una setmana això no afectarà la comparabilitat dels resultats (sempre que el seu estat de salut física i mental no afecti l'acte de mesurar!).

### **Establiment dels punts temporals de comparació**

Òbviament exclouem d'aquestes consideracions aquells estudis en els quals, justament, el que interessa és l'evolució d'un text al llarg d'un període de temps. Llavors, és clar, els punts temporals de comparació han d'establir-se amb cura i ser respectats.

En definitiva, en la tècnica d'anàlisi de contingut podem proposar-nos atorgar un cert grau de reproduïbilitat a la mesura, i fer que diferents analistes ben entrenats assolixin una coincidència considerable en els resultats trets d'un mateix text, ja fixat, amb independència del moment en el qual facin les mesures. En mots més tècnics, interessa que l'acord entre els resultats sigui màxim o que la diferència (o desviació típica, si són molts els analistes comparats) entre ells sigui mínima.

### **Exactitud**

Entenem per exactitud el grau en el qual la mesura s'acosta a un valor ideal que es defineix convencionalment com un patró o referència estàndard de la mesura.

El problema, ja conegut, es planteja quan cal decidir qui o què estableix aquest valor. En física, sovint, el valor ideal prové de l'aplicació d'un aparell de mesura de qualitat en condicions òptimes, malgrat que també es pot afegir a aquestes condicions que el mesurador sigui un expert. En psicologia, en el cas d'un test, els aprenents de psicòleg poden millorar la seva interpretació i diagnòstic comparant els seus resultats amb els d'un expert, clínic o educatiu, que sigui qui prengui la responsabilitat de marcar el valor de referència. En l'anàlisi de contingut es pot seguir una tàctica semblant: quan calen analistes per a mesurar i buidar un corpus abundant, un analista amb experiència pot fixar els valors de referència (manualment o amb l'ajuda d'un programa d'anàlisi), de manera que els analistes més novells poden practicar després el buidat i el recompte del mateix text fins a coincidir amb aquells valors. Un analista

novell podria trobar 46 vegades una categoria en una secció de text, mentre que l'expert que l'entrena podria haver-la trobat 48 vegades, diferència que marcaria el nivell de precisió del primer.

En qualsevol cas, és obvi que l'exactitud correspon a la precisió, si les entenem totes dues a partir de la possibilitat d'atansar la mesura a uns valors de referència estimats convencionalment com els vertaders.

Al marge d'aquesta filosofia de la cerca de la mesura vertadera, l'exactitud o la precisió basen les seves comparacions en els resultats obtinguts per dos o més observadors o aparells de mesura, com quan es tracta de fiabilitat interobservadors o reproduïbilitat. Per tant, els comentaris que hem fet en el subapartat 3.2.1 també són aplicables aquí.

Els conceptes de fiabilitat, com *estabilitat*, *reproduïbilitat* i *exactitud*, propis de l'anàlisi de contingut, serveixen una correspondència quasi total amb els tres conceptes de fiabilitat previstos per la metodologia qualitativa.

### 3.3. L'aproximació correlacional a la fiabilitat

L'estabilitat o el grau de coincidència entre dos registres es poden estimar a partir de la seva correlació lineal.

Tal com confiem haver aclarit, la cerca de fiabilitat implica, en la pràctica, la comparació de dues o més sèries de dades per a avaluar el seu grau de coincidència. En la perspectiva de l'estabilitat, l'horitzó de l'investigador és aconseguir que successius registres siguin o acabin essent iguals o quasi iguals, de manera que demostrï així la consistència dels seus criteris de categorització i el rigor dels seus recomptes o càlculs. En la perspectiva de la reproduïbilitat i, per tant, de la fiabilitat interobservadors, l'objectiu apunta igualment a la convergència de les mesures dels diferents analistes. En la de l'exactitud o precisió, el que interessa particularment és l'atansament dels resultats d'un o més analistes fallibles o inexperts als d'un analista expert.

En totes aquestes perspectives, doncs, podem necessitar una eina per a la valoració quantitativa d'aquesta coincidència, equivalència o convergència de registres o mesures. Diem *podem* perquè també cal puntualitzar que un tarannà científic però plenament qualitatiu podria prescindir, en la seva rutina de treball, d'aquesta quantificació. L'assoliment de l'estabilitat com a igualtat en dues o tres sèries de mesures consecutives pot revelar-se amb una simple inspecció de les mesures "a ull" (tot i que preferentment sobre la seva representació gràfica). I l'atansament i coincidència final de dues o més sèries de dades,

generades pels corresponents analistes, pot resoldre's per mitjà de la negociació i el consens entre ells, d'un examen i discussió conjunts dels criteris de categorització i de buidat que duguin a més sinergia en el procés de mesura.

Tanmateix, si considerem que una referència quantitativa pot ser un suport important a l'hora de calibrar el grau de fiabilitat aconseguit, disposem de diferents estris estadístics o probabilístics en forma de coeficients o d'índexs. Ens aturarem en els dos més utilitzats i més justificats des d'un punt de vista tècnic: la correlació i l'índex probabilístic  $\kappa$ .

La correlació s'ha utilitzat de manera típica i habitual en l'àmbit de la fiabilitat intraobservadors aplicada a tests. Com ja hem indicat i és ben conegut, l'estabilitat dels resultats de l'aplicació d'un test se sol valorar mitjançant la correlació entre els resultats de dues administracions consecutives de l'instrument, separades per un cert interval de temps, segons l'esquema test-retest. En aquesta modalitat (i ens referim a la més simple) es busca concretament el coeficient  $r$  de Pearson entre les puntuacions obtingudes en les dues administracions, a partir de la fórmula producte – moment.

Si un analista de text pretén conèixer el grau d'estabilitat o consistència de les seves mesures, podrà buscar també la correlació  $r$  entre dues o més sèries de dades obtingudes, sobre el mateix text, en dos moments diferents.

### Exemple d'aproximació correlacional a la fiabilitat

Passem a exposar un exemple, sota el supòsit que les mesures són de freqüència. Un analista ha fet dos recomptes d'unes categories  $A, B, C, D, E, F, G$  amb una setmana de separació, amb l'objectiu fàcilment deduïble de comprovar que els dos recomptes coincideixen, o de saber en quin grau coincideixen.

Els dos registres donen els valors següents:

Categories:	A	B	C	D	E	F	G
1r. registre	13	9	21	34	5	17	12
2n. registre	15	10	19	29	5	15	13

La correlació és  $r = 0,987$ .

Aquesta és una correlació positiva molt alta. Tot i així, és obvi que la coincidència entre els dos registres no és gran. Aquest tipus d'incongruències deriva de les característiques de la correlació, que és un estri útil però al qual no se li pot demanar allò que no pot donar. La correlació ens informa sobre la intensitat i la direcció de la relació lineal matemàtica entre les dues sèries, però no és capaç de dir-nos si realment hi ha coincidència.

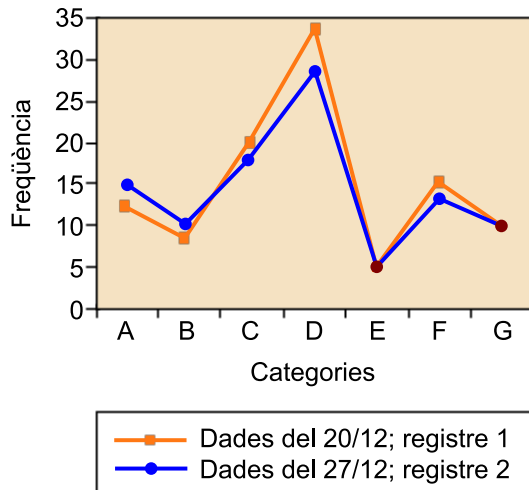
Si una de les sèries hagués estat obtinguda abans de Nadal, en unes bones condicions mentals, i la segona just després, encara amb ressaca de les festes, i en aquesta segona les freqüències haguessin estat sistemàticament molt augmentades o disminuïdes respecte a les del primer registre, la correlació també ens hauria donat alta, malgrat que la coincidència hauria estat més escassa encara. Si, per exemple, en el segon registre totes les freqüències estiguessin 5 unitats per sobre de les del primer,  $r$  donaria valor 1, de manera força enganyosa. Existeixen procediments per a esbrinar si la correlació indica coincidència o, simplement, correspondència matemàtica entre sèries de dades, però aquí ens limitarem a recomanar la inspecció gràfica dels resultats, per a apreciar si aquests tenen un recorregut paral·lel o hi ha una autèntica aproximació o coincidència.

#### Vegeu també

Recordeu que el coeficient  $r$  de Pearson es va estudiar en l'assignatura *Anàlisi de dades*.

En l'exemple exposat una breu inspecció ens indica, en la taula i en la següent figura, que l'analista tendeix clarament en el segon registre a augmentar els registres baixos i a reduir els alts. Tanmateix, el registre amb freqüència més baixa (el de la categoria *E*) roman igual, probablement perquè l'escassa magnitud de la xifra tolera menys marge d'error.

### Correlació entre dos buidatges d'un text



Les dificultats que, com les que tot just hem presentat, acompanyen la interpretació d'un coeficient de correlació en el context de la fiabilitat qualitativa, aconsellen l'ús d'índexs amb una base més probabilística i que subministrin una informació més pertinent en relació amb la coincidència de mesures.

Ultra això, cal adonar-se que el càlcul que acabem de fer es basava en les freqüències globals de cada categoria respecte al total del text analitzat. Ara bé, pot interessar-nos una estimació més per menut de la coincidència basada, no en el còmput global, sinó en cada registre local, en cada unitat o segment i segons una determinada regla d'enumeració. Això significaria que els registres es compararien de línia en línia, de paràgraf en paràgraf o de pàgina en pàgina.

El coeficient de correlació entre dos registres com a estimador de la fiabilitat planteja algunes dificultats en la interpretació i es limita a la comparació de les mesures globals sobre un text.

### 3.4. L'aproximació probabilística i "punt a punt" a la fiabilitat

L'índex més acceptat actualment per a estimar el grau d'acord entre dos analistes és la kappa de Cohen.

### 3.4.1. Acord o coincidència sobre la totalitat del sistema de categories

En comparar dos registres amb les seves mesures corresponents disposarem de més informació i, per tant, podrem extreure conclusions més sòlides sobre fiabilitat, si allò comparat no són les mesures globals preses per dos analistes diferents, o les mesures globals diferents preses per un mateix analista, sinó cada un dels registres fets en cada segment o context de mesura del material examinat.

Aleshores, posem per cas, no contrastarem dues freqüències totals d'una categoria X, obtingudes en dues computacions diferents del mateix corpus examinat (diguem, per exemple, dues freqüències com 34 i 29 en l'exemple de l'apartat anterior); allò que contrastarem serà cada una de les assignacions de categoria realitzades: controlarem si en la línia 79, en el paràgraf 12 o en la pàgina 3 ha estat assignada la mateixa categoria, dins de cada un de dos registres consecutius d'un mateix analista o en dos registres de dos analistes diferents; i així en totes les unitats del text. Tot aquest raonament i el que segueix podria ser igualment aplicable a qualsevol nombre de registres (tres, quatre, etc.), però en les exemplificacions que hi ha a continuació ens cenyirem a dos, que tanmateix és el que es fa servir més.

#### Exemple d'acord i coincidència sobre la totalitat del sistema de categories

Fixem-nos en les taules següents. Cada una representa un dels dos registres comparats, posem que de dos analistes diferents. En les files figuren les categories d'anàlisi, A, B, C; en les columnes, els segments o contextos de mesura, que, en aquest exemple, hem decidit que fossin els paràgrafs.

Paràgraf:		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	T	
Analista 1	A	X				X	X			X			X	X			X	X					8
	B		X					X							X								3
	C			X	X				X		X	X				X			X	X	X		9

Paràgraf:		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	T	
Analista 2	A	X				X	X	X		X		X	X	X			X						9
	B		X	X											X				X				4
	C				X				X		X					X		X		X	X		7

Amb ànim de simplificar i destacar els aspectes més nuclears de la tècnica hem suposat, de moment, que el sistema de categories és exhaustiu i exclusiu i que, per consegüent, les categories no es poden encavalcar dins d'un paràgraf; cada paràgraf només pot contenir una categoria i no hi pot haver paràgraf "buit", és a dir, sense categoria.

Amb aquest tipus d'informació no tenim excusa per a limitar-nos a comparar les tres freqüències globals de les tres categories, 8 i 9, 3 i 4, 9 i 7, respectivament. Ara, paràgraf a paràgraf, estem capacitats per a detectar les coincidències o discrepàncies entre tots dos analistes. Així, en el primer paràgraf tots dos analistes han assignat la categoria A; en el segon, la B, i en el tercer ja s'ha produït una discrepància, ja que el primer analista ha

registrat C mentre que el segon ha preferit B. I així podríem revisar els registres successius de l'exemple fins al paràgraf 20.

A partir de taules de registre com l'anterior es construeix una taula o matriu de confusió, on s'arrangen els registres de manera adequada per tal de tractar-los més endavant. Aquesta és una taula simètrica en què en files i en columnes apareixen els registres de cada analista, els quals, en creuar-se en les caselles, ens indiquen el nombre de coincidències i discrepàncies.

		Analista 2			
Analista 1		A	B	C	Total
	A	7	0	1	8
	B	1	2	0	3
	C	1	2	6	9
	Total	9	4	7	20

Així, respecte a la categoria A, els dos analistes haurien coincidit 7 vegades; respecte a la B, 2 vegades, i respecte a la C, 6 vegades. En canvi, en 1 ocasió l'analista 1 hauria vist A mentre que l'analista 2 hauria vist C; o 2 vegades l'analista 1 hauria consignat C mentre que l'analista 2 hauria consignat B; etc. Els totals de fila representen el nombre total de vegades que l'analista 1 ha vist cada una de les tres categories; els totals de columna són el nombre total de vegades que l'analista 2 ha vist cada una de les tres categories.

Amb aquest arranjament de coincidències i discrepàncies, acords i desacords, podem d'antuvi calcular una senzilla proporció d'acord respecte a la totalitat del sistema de categories:

$$\text{Proporció d'acord} = \text{nombre d'acords} / \text{nombre total d'ocurrències}$$

El nombre total d'ocurrències és, evidentment, el nombre d'acords més el nombre de desacords.

En una taula de confusió com l'anterior, aquesta proporció es deriva directament de la diagonal principal, ja que és on es localitzen els acords o coincidències. Per tant:

$$\text{Proporció d'acord respecte a A, B, C} = 7 + 2 + 6 / 20 = 0,75$$

Semblaria, doncs, legítim afirmar que els dos analistes han assolit un 0,75 d'acord o coincidència (o un 75%). I així es va entendre durant dècades en l'àmbit de la metodologia psicològica, fins que es van començar a alçar veus reclamant el paper de l'atzar en la magnitud de l'acord.

Efectivament, una part de la coincidència entre dos analistes, concebuts com dos sistemes independents de registre, es pot deure a l'atzar. Els dos analistes fan el registre per separat i, en conseqüència, el fet que extreguin d'un mateix segment de text una A, una B o una C pot ser degut a la casualitat. En l'exemple al qual ens estem referint, la probabilitat que l'analista 1 apunti A en qualsevol segment de text es podrà estimar amb  $8 / 20$ ; la que l'analista 2 apunti A en condicions idèntiques amb  $9 / 20$ . La probabilitat que tots dos apuntin A serà, doncs,  $8 / 20 \times 9 / 20 = 0,18$ . Podríem, és clar, repetir el mateix raonament probabilístic amb B i C. Sigui com sigui, s'infereix d'aquest raonament que aquesta part de l'acord deguda a l'atzar hauria de "restar-se" o "separar-se" de l'acord aparent abans calculat.

La història dels índexs que han tractat d'aconseguir aquesta separació d'una manera lògicament i matemàticament correcta és apassionant, però aquí l'ometrem. Hi ha una llarga llista d'índexs amb diferents patents, ideats per autors diferents i utilitzats en diferents àmbits científics i aplicats de la psicologia. No obstant, l'índex que s'ha imposat, per qualitat i facilitat de càlcul relativa, és la kappa de Cohen:

$$k = p_o - p_e / 1 - p_e$$

$p_o$  és la proporció d'acord observada o empírica a la qual ens acabem de referir;  $p_e$  és la proporció esperada d'acord, és a dir, la que pot ser deguda a l'atzar. Al numerador figura la resta entre aquests dos valors, que ens ha de donar el valor real de l'acord, un cop neutralitzat l'atzar. El denominador permet estandarditzar aquesta diferència.

Acabem de mostrar com es computa  $p_o$ . Pel que fa a  $p_e$ , el càlcul es duu a terme seguint el mateix raonament probabilístic que hem iniciat fa unes línies. Ja hem explicat que la probabilitat d'encertar-los tots dos per casualitat en registrar A en una mateixa localització del text seria:

$$p_{AA} = 8 / 20 \times 9 / 20 = 0,18$$

Llavors,

$$p_{BB} = 3 / 20 \times 4 / 20 = 0,03$$

$$p_{CC} = 9 / 20 \times 7 / 20 = 0,1575$$

La suma d'aquestes tres probabilitats ens permetran estimar la probabilitat global de coincidir tots dos analistes. Efectivament, ara considerarem com a esdeveniments dependents les tres coincidències AA, BB i CC, ja que la magnitud d'una afecta la de les altres. En conseqüència, les probabilitats poden ser sumades.

$$p_e = p_{AA} + p_{BB} + p_{CC} = 0,18 + 0,03 + 0,1575 = 0,368$$

Com que ja tenim els valors necessaris,  $p_o$  i  $p_e$ , només queda substituir-los en la fórmula de  $k$  que hem inserit més amunt:

$$k = 0,75 - 0,368 / 1 - 0,368 = 0,382 / 0,632 = 0,6044$$

S'aprecia, doncs, que la proporció d'acord calculada abans esdevé força corregida: hem baixat de 0,75 a 0,60; hem suprimit l'efecte de l'atzar en l'acord.

En la pràctica investigadora es considera que un valor de  $k$  superior a 0,7 garanteix força l'acord, tot i que hi ha opinions que es conformen amb valors inferiors.

#### Polèmica entorn de $k$

Cal dir que la popularitat de  $k$  molesta autors importants com Krippendorff per raons tècniques, en rigor justificades, en què no entrem aquí.

#### Una altra fórmula

Generalment, però, el càlcul no es presenta així en els manuals, sinó mitjançant la fórmula abreujada de "suma dels productes de files i columnes corresponents / total d'ocurrències al quadrat". En el nostre cas  $p_e = [(8 \times 9) + (3 \times 4) + (9 \times 7)] / 20^2$ . No cal afegir que és el mateix càlcul una mica més mecanitzat.

#### Nota

Hi ha maneres d'estimar la significació de  $k$  com si es tractés d'una puntuació  $z$ , però són considerablement complexes.

En qualsevol cas, la  $\kappa$  que hem obtingut nosaltres deixa molt a desitjar.

No obstant això, hi ha raons serioses per a no témer valors baixos de  $\kappa$ . La crítica principal al nostre parer es refereix al caràcter excessivament conservador d'aquest índex, que pot donar, en determinades circumstàncies, magnituds d'acord molt per sota de l'acord real. No obstant, cal tenir en compte que aquest problema es pot mitigar acceptant valors relativament baixos de  $\kappa$ , com acabem de suggerir.

D'altra banda, l'índex alternatiu que proposa Krippendorff és  $\alpha = 1 - d_o / d_e$ , en què  $d_o$  i  $d_e$  són el desacord observat i l'esperat. Aquest índex tanmateix obliga a fer un càlcul força tediós i complicat i segurament per això és poc utilitzat.

Un índex que ha estat força utilitzat en el context de l'anàlisi de contingut és el  $\pi$  de Scott, amb una fórmula conceptualment idèntica a la de  $\kappa$ . La diferència rau en la manera d'entendre el nombre total d'ocurrències. En el nostre exemple serien  $20 + 20 = 40$ , interpretant que cada analista ha fet, pel seu compte, 20 actes d'assignació categorial. En conseqüència, també serà diferent la manera de calcular  $p_e$  a partir de les probabilitats corresponents.

Des d'aquest punt de vista, les probabilitats d'acord s'entendrien sobre la base de la distribució conjunta d'assignacions categorials:

- L'analista 1 hauria detectat A amb  $p_A = 8 + 9 / 40$ ; B amb  $p_B = 3 + 4 / 40$ ; C amb  $p_C = 9 + 7 / 40$ .
- L'analista 2 hauria registrat idèntiques probabilitats que l'1. Aquestes probabilitats, fent les operacions, tornarien a ser 0,43; 0,18 i 0,40, respectivament.

Aleshores, multiplicant les probabilitats de l'un i l'altre com en el cas de  $\kappa$  ens quedarà una suma de quadrats (ja que les probabilitats són idèntiques):

$$p_e = 0,43^2 + 0,18^2 + 0,40^2 = 0,377$$

Amb aquest valor, i aplicant la mateixa fórmula que en el cas de  $\kappa$ :

$$\pi = 0,75 - 0,377 / 1 - 0,377 = 0,6051$$



Com podeu comprovar en un exemple senzill com el que hem utilitzat aquí, la diferència entre tots dos índexs és menyspreable.

### 3.4.2. Acord o coincidència sobre una categoria en concret

Fins aquí hem desenvolupat arguments i exemples relatius a la manera de valorar l'acord sobre l'ús d'un sistema de categories sencer,  $S = \{A, B, C\}$ . No ens ha de sorprendre tanmateix que l'investigador s'interessi per l'acord sobre una de les categories del sistema, i no sobre tot el sistema. El procediment d'estimació és molt semblant al que acabem d'exposar. La diferència principal descansa en la taula de confusió de partida, que en aquest cas contrasta els registres d'una categoria amb els de la resta. Una taula d'aquesta mena es pot derivar d'una taula general com la que hem utilitzat en l'apartat anterior. Imaginem que en comptes de buscar  $\kappa$  per a A, B i C la busquem per a A i prou. Llavors la taula adoptaria la forma següent:

		Analista 2		
		A	No A	Total
Analista 1	A	7	1	8
	No A	2	10	12
	Total	9	11	20

La casella ANoA amb freqüència 1 resulta, doncs, de la suma de les caselles AB i AC de la taula anterior; la casella NoAA amb freqüència 2, de la suma de les caselles BA i CA; finalment, la casella NoANoA resumeix amb freqüència 10 les caselles BB, BC, CB i CC de la primera taula. Sobre aquests quatre valors, 7, 1, 2, 10, i els seus totals de fila i columna, calculem  $\kappa$  de la manera que hem descrit.

$$p_o = 7 + 10 / 20 = 0,85$$

Ara, per tal d'anar pel dret, calcularem  $p_e$  mitjançant la recepta "suma dels productes de files i columnes corresponents / total d'ocurrències al quadrat".

$$p_e = [(8 \times 9) + (11 \times 12)] / 20^2 = 0,51$$

$$\kappa = 0,85 - 0,51 / 1 - 0,51 = 0,34 / 0,49 = 0,6939$$

Així, doncs, aquesta  $\kappa$  centrada en la categoria A té un valor superior a la  $\kappa$  general. Per si mateixa, aquesta categoria ha plantejat menys problemes d'acord que la totalitat del sistema.

#### Lectures recomanades

Amb relació als valors acceptables de  $\kappa$ , podeu consultar el text ja citat de Bakeman i Gottman (1989), p. 114.

Amb relació a la crítica de Krippendorff i al càlcul comparat de  $\kappa$ ,  $\pi$  i  $\alpha$ , vegeu:

K. A. Neuendorf (2002). *The Content Analysis Guidebook* (p. 150 i següents) Thousand Oaks (CA): Sage.

### 3.4.3. Acord quan no es compleix la condició d'exhaustivitat

En iniciar l'exposició de  $\kappa$ , hem simplificat les condicions d'aplicació suposant que es complien les condicions d'exclusivitat i exhaustivitat. De fet, que es compleixi la condició d'exclusivitat en categories d'anàlisi de contingut, no és una condició excessiva, ans al contrari. Si entenem correctament aquesta condició com l'exigència que cada segment o unitat d'anàlisi sigui atribuït a una sola categoria i només a una sola categoria, aquesta exigència no constitueix una simplificació abusiva ni s'ha de veure com una trampa. Al revés: és el criteri lògic nuclear que confereix rigor i qualitat a una definició o classificació categorial. Si cada segment fos assignable a més d'una categoria, el sistema resultant seria borrós, i la mesura, impossible.

#### Quan es dona la condició d'exhaustivitat

El compliment d'aquesta condició depèn sovint d'haver reeixit en la segmentació realitzada. Pot donar-se la circumstància, per exemple, que una frase complexa pugui ser assignada a dues categories, però si la subdividim en dues frases simples el problema desapareix i cada una d'aquestes frases correspon llavors nítidament a una categoria distinta.

Per altra banda, cal no confondre aquesta condició, tal com l'acabem de presentar, amb la condició que els indicadors o trets de cada categoria siguin exclusius. En aquest sentit ja sabem que una categoria pot tenir indicadors exclusius o bé tenir-ne una combinació particular. Una oració, presa com a unitat textual, ha de ser atribuïda a una sola categoria, però els elements que defineixen aquesta categoria poden trobar-se només en aquesta (i llavors l'oració només tindrà trets d'aquesta categoria) o bé trobar-se també en altres categories combinats amb altres diferents (i llavors l'oració tindrà elements d'aquesta categoria i d'altres, ja que alguns són compartits).

En canvi, la condició d'exhaustivitat no ha de ser sempre satisfeta. Per exemple, en l'anàlisi avaluativa que hem descrit més amunt, l'interès de l'investigador anava adreçat específicament a frases amb contingut avaluatiu relacionades amb objectes d'actitud, de manera que durant la progressiva revisió del text escollit podien trobar-se segments no categoritzables. La situació de dos analistes davant del text seria llavors diferent de la que hem imaginat fins ara. En les taules de confusió ja comentades, el fet que un analista no detectés una categoria comportava necessàriament que en detectava una altra (si no veia A, veia B o C). Però si el sistema no és exhaustiu respecte al text podrà donar-se el cas que un analista detecti una categoria i l'altre no en detecti cap. Això obliga a modificar la taula de confusió tal com es fa palès en l'arranjament de files i columnes següent:

Analista 2						
Analista 1		A	B	C	Cap	Total
	A					
	B					
	C					
	Cap					
	Total					

Tot comptat, la manca de categoria es tracta com una categoria més. Aquest truc lògic es pot mirar des d'un altre punt de vista: en realitat el que hem fet és definir una categoria zero o buida, com les que hem caracteritzat al final del capítol 2. Així, amb l'ajuda d'aquesta categoria, es compleix la condició d'exhaustivitat i la taula on és inclosa no presenta cap diferència amb les que hem vist fins ara.

Sobre aquest arranament, les proporcions d'acord i la  $\kappa$  es calculen segons els mateixos procediments que hem exposat fins aquí.

La bondat de kappa sobre altres índexs d'intenció semblant és que corregeix la part d'acord que pot ser deguda a l'atzar.

## 4. Síntesi de l'anàlisi de contingut

### 4.1. Un exemple integrador: la meditació transcendental mitjançant els seus documents

Una recerca d'anàlisi de contingut pot i sol combinar elements de metodologia qualitativa amb altres de metodologia quantitativa.

L'exemple que us oferim ara serà presentat a vista d'ocell però, en compensació, segurament oferirà més versemblança perquè es tracta del resum d'una investigació sencera, de la qual hem extret allò que era inseparable de les decisions metodològiques essencials. El punt de vista serà més general i això ens permetrà distingir amb més claredat les fases típiques d'una recerca d'anàlisi de contingut tal com han estat seqüenciades aquí.

1) **Objecte d'estudi i hipòtesis.** L'autor es proposa estudiar un moviment de caire religiós anomenat *meditació transcendental* mitjançant els seus documents i –també– mitjançant registres d'observació participant.

Més concretament, el focus de l'estudi apunta l'expressió de creences en els textos i l'exemplificació de rols socials i obligacions assumibles pels feligresos. Se sosté la hipòtesi següent: amb el temps, els continguts dels textos, tant pel que fa a les creences com a les obligacions, anirà derivant cap a una popularització o caràcter populista dels temes.

*Popularització* vol dir que els continguts religiosos esdevenen menys espirituals i més abocats a les coses terrenals i a la pràctica quotidiana. Aquest gir de l'espiritualitat cap a la vida quotidiana es capta gràcies a les referències als beneficis mundans de l'ortodòxia, al rendiment personal i social de rituals i cerimònies, a l'accessibilitat de Déu, a la inutilitat de certs conflictes i esforços i –per contra– a la utilitat de les actituds conservadores...

Els valors de la variable *temps*, essencials per tal d'avaluar aquests canvis temàtics, venen garantits per l'any de publicació dels llibres, revistes i documents examinats.

2) **Selecció i mostratge.** Pel que fa al mostratge intertextual, els textos són seleccionats dins d'un període de vint anys, des de la fundació del moviment (1961) fins als primers símptomes de crítica.

#### Lectura recomanada

L'article de referència és E. Woodrum (1984). "Mainstreaming" Content Analysis in Social Science: Methodological Advantages, Obstacles and Solutions. *Social Science Research*, 13, 1-19.

La part que resumim es troba entre les pàgines 8 i 18.

Són acceptats dins del corpus tots els materials que puguin considerar-se documents oficials de la doctrina i que descriguin creences, pràctiques i objectius socials. Més concretament, són oficials els documents que compleixen almenys una d'aquestes condicions:

- a) ser publicats per la premsa privada del moviment;
- b) ser escrits per un representant conegut i acceptat del moviment;
- c) ser recomanat en una de les llistes enviades als militants i simpatitzants del moviment des de la seva seu central.

Sobre aquesta base es recopilen 5.716 pàgines de literatura sobre meditació transcendental, però el període d'estudi queda reduït a setze anys, interval durant el qual es va publicar material que satisfia almenys una de les tres condicions a), b) o c).

El mostratge intratextual es resol per procediments aleatoris i quotes proporcionals a l'extensió de cada document del corpus. D'aquesta manera s'escullen 25 paràgrafs per any i, per tant, el material definitiu subjecte a anàlisi és de  $25 \times 16 = 400$  paràgrafs.

**3) Unitats d'anàlisi.** La unitat textual és el mot. L'anàlisi treballa sobre un inventari de 26 mots que, situats en el seu context serveixen d'indicadors de les categories que de seguida esmentarem.

La unitat de context (interpretatiu) és l'oració. Per tant, la unitat de registre text-context és el mot-oració.

**4) Categories.** L'autor troba 19 categories temàtiques construïdes a partir de diverses combinacions de les 26 unitats de registre un cop codificades, les quals funcionen com a indicadors semàntics.

#### **Exemple**

"l·luminació" i "esperit" són indicadors de categories de religió espiritual, mentre que "treball" o "salut" ho són de categories de religió popular o popularitzada, més adreçada als interessos seculars.

L'autor fonamenta el valor d'aquests indicadors en els seus estudis anteriors sobre la distinció entre religió popular o terrenal, d'una banda, i mística o espiritual, de l'altra.

El criteri d'exhaustivitat no es compleix totalment, ja que en un 3% del corpus textual (els 400 paràgrafs) no hi apareix cap de les 19 categories previstes.

**5) Mesures.** Es fan mesures de freqüència a dos nivells, comptant els indicadors o unitats de registre (mots-oració) i comptant també les categories temàtiques.

Com a regla d'enumeració o context de referència de la mesura es fa servir el paràgraf.

Tenim, doncs, mesures de freqüència de mot/paràgraf i de categoria/paràgraf en cada interval anual considerat. D'aquesta manera es pot validar la relació entre unitats de registre i categories a la vegada que s'estableixen dos nivells d'anàlisi de la popularització de temes religiosos.

**6) Fiabilitat.** La fiabilitat s'enfoca des del punt de vista de la consistència interna del repertori de mots i del sistema de categories. S'empra un coeficient correlacional diferent a l'exemplificat aquí, l'alfa de Cronbach.

No es calculen mesures d'intensitat o direcció.

**7) Resultats i anàlisi.** La hipòtesi de l'autor se sosté gràcies al següent:

- L'alta correlació entre la sèrie d'anys de publicació i les mesures corresponents de les categories temàtiques, amb tendència creixent a anar cap a la popularització de temes d'any en any. La correlació entre anys i mesures de mots no va ser tan significativa.
- L'anàlisi de la variància, amb nivells de confiança superiors també en el cas de l'efecte del temps sobre les categories temàtiques.

#### **Nota**

Cal tenir en compte que totes les inferències que han permès les mesures apunten una institució com a autora dels textos.

**8) Resultats i anàlisi.** Els textos oficials de la meditació transcendental han anat popularitzant els seus continguts al llarg del període d'estudi.

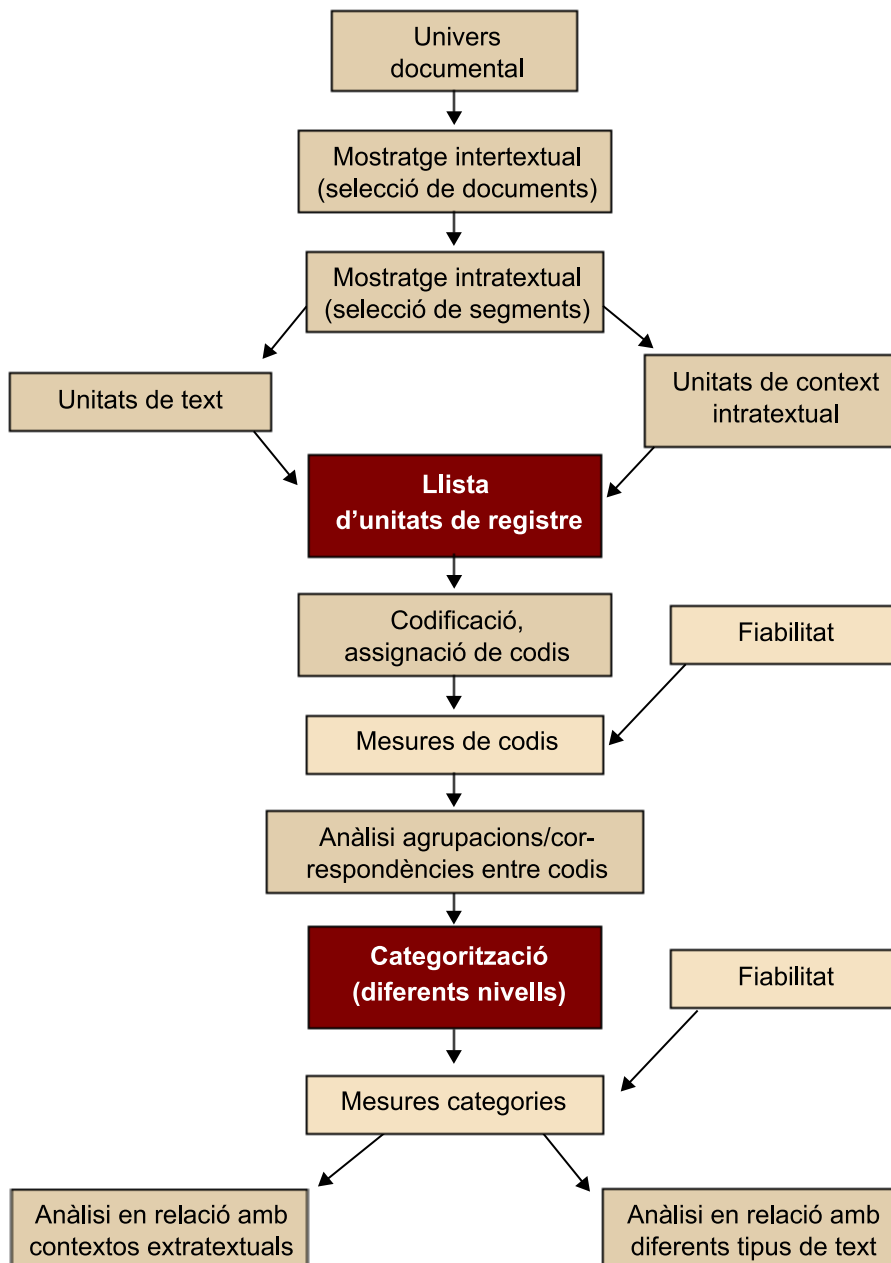
Una recerca d'anàlisi de contingut pot fer seguiments en perspectiva longitudinal per tal de saber la influència del temps en els textos.

## **4.2. Sinopsi final en forma de diagrama**

Una investigació d'anàlisi de contingut pot ser representada mitjançant una seqüència d'operacions semblant a les que exigeix l'execució d'un pla.

En l'esquema adjunt us oferim una possible columna vertebral de la seqüència d'operacions i accions necessàries per a dur a terme una anàlisi de contingut completa, des del punt de partida –l'establiment d'un univers documental– fins a les diferents direccions d'anàlisi que es poden seguir a propòsit de les dades obtingudes. Totes aquestes operacions han estat especificades i desenvolupades en els apartats previs.

Ens hem estalviat els retrocessos que tota investigació ha d'acceptar. En el dibuix hem distingit amb dues tonalitats de color les operacions de mesura –per definició més "quantitatives"– de les operacions purament qualitatives (selecció, codificació, categorització, etc.).



Les operacions necessàries per a dur a terme una anàlisi de contingut mostren un clar paral·lelisme amb les que exigeix qualsevol investigació observacional o qualitativa.



## Bibliografia

- Bardin, L. (1996). *El análisis de contenido*. Madrid: Akal.
- Duverger, M. (1975). *Métodos de las ciencias sociales* (8a. ed., pp. 182-183). Barcelona: Ariel.
- Grbich, C. (2007). *Qualitative data Analysis. An Introduction* (p. 116 i seg.). Los Angeles: Sage.
- Karsenti, T., Larose, F. i Núñez, M. (2002). La apertura universitaria a los espacios de formación virtual: Un reto a la autonomía estudiantil. *Revista Electrónica de Investigación Educativa*, 4 (1), 1-18.
- Kirk, J. M. i Miller, M. (1986). *Reliability and Validity in Qualitative research*. Beverly Hills: Sage.
- Kohlbacher, F. (2006, gener). The Use of Qualitative Content Analysis in Case Study Research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 7 (1), art. 21. [Data de consulta: 10 de gener de 2009]: <<http://www.qualitative-research.net/index.php/fqs/article/view/75/154>>
- Krippendorff, K. (1990). *Metodología del análisis de contenido*. Barcelona: Paidós
- López-Aranguren, E. (1989). El análisis de contenido. A M. García Ferrando, J. Ibáñez i E. Alvira (Comp.), *El análisis de la realidad social. Métodos y técnicas de investigación* (2a. ed., pp. 383-414). Madrid: Alianza.
- Mayntz, R., Holm, K. i Hübner, P. (1975). *Introducción a los métodos de la sociología empírica*. Madrid: Alianza.
- Mayring, P. (2000, juny). Qualitative content analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1 (2), art. 20. [Data de consulta: 5 d'octubre de 2004]: <<http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-e.htm>>
- Neuendorf, K. A. (2002). *The Content Analysis Guidebook* (pàg. 150 i següents). Thousand Oaks (CA): Sage.
- Penalva, C. i Mateo, M. A. (2006). *Tècniques qualitatives d'investigació*. Alacant: Universitat d'Alacant, Departament de Sociologia II, Psicologia, Comunicació i Didàctica / Universitat d'Alacant, Secretariat de Promoció del Valencià / Conselleria de Cultura, Educació i Ciència. [Data de consulta: 15 d'agost de 2008]: <<http://biblioteca.universia.net/ficha.do?id=35524203>>
- Pérez Serrano, G. (1994). *Investigación cualitativa: retos e interrogantes. II, Técnicas y análisis de datos*. Madrid: La Muralla.
- Quivy, R. i Van Campenhoudt, L. (1997). *Manual de recerca en ciències socials*. Barcelona: Herder.
- Weber, R. P. (1990). *Basic Content Analysis*. Newbury Park: Sage.
- Woodrum, E. (1984). "Mainstreaming" Content Analysis in Social Science: Methodological Advantages, Obstacles and Solutions. *Social Science Research*, 13, 1-19.

