

UN DICCIONARIO DE SINÓNIMOS INFORMATIZADO

Dr. Joaquim Rafel i Fontanals
Francesc Rosés i Albiol
Mercè Prat Sala
Auke Oosterhoff
Imma Fradera Castillo
Albert Llorens Muntada
Lídia Lluís Gavaldà
Montserrat Torras Conangla
Anna Moix Bergadà

Universitat de Barcelona

Con esta comunicación pretendemos dar a conocer el proyecto DICCSIN que tiene como objetivo principal la elaboración de un diccionario catalán de sinónimos. Este proyecto se realiza conjuntamente entre la Universitat de Barcelona y el Centro de Tecnologías de la Lengua de IBM.

Exponemos los criterios lingüísticos que seguimos así como las herramientas informáticas que han sido desarrolladas especialmente para este proyecto haciendo hincapié en la importancia que éstas tienen tanto en la productividad como en la detección y eliminación de errores. Dedicamos una especial atención a las herramientas diseñadas para la verificación de congruencia en series sinonímicas.

INTRODUCCIÓN

Nuestra comunicación, como su título indica, versa sobre nuestras experiencias en la elaboración de un diccionario catalán de sinónimos. Experiencias que han sido ricas, como veremos, tanto desde el punto de vista lingüístico como desde el punto de vista informático.

El proyecto de elaboración de un diccionario catalán de sinónimos (DICCSIN) arranca bastantes años atrás, hacia el año 1987, cuando se dio por acabado un proyecto destinado a elaborar una base de datos léxica que contuviera las palabras más frecuentes del catalán escrito actual. En este proyecto participaron tres partes: La *Generalitat de Catalunya*, el *Centro de Investigación UAM-IBM* y la *Universitat de Barcelona*. El proyecto dio como resultados una base de datos léxica con más de treinta mil entradas con información sobre su categoría gramatical y partición silábica. Las entradas, una vez expandidas, generaron un total de más de doscientas mil palabras. Esta base de datos, además de las palabras propias de un diccionario de la lengua, contiene listas cerradas con nombres de empresas, topónimos de la geografía catalana, valenciana y balear, antropónimos, topónimos del mundo entero, monedas, lenguas... Lo que convierte esa base de datos en un corpus idóneo para ser incluido en un corrector ortográfico. Esta fue, pues, la tarea de IBM. La base de datos resultante del proyecto se incluyó en el corrector ortográfico del procesador de textos de IBM *Display Write*.

Fue a partir de la finalización de ese proyecto que se empezó a hablar de empezar otro proyecto destinado a la elaboración de un diccionario catalán de sinónimos. El *Centro de Investigación UAM-IBM* había tenido ya una experiencia similar con el castellano. Nuestra estrategia, sin embargo, es claramente distinta a la seguida por IBM. Ellos adquirieron los derechos de un conocido diccionario de sinónimos castellano y lo entraron en soporte magnético para proceder después a una depuración de su contenido. La *Universitat de Barcelona* planteó desde el principio el proyecto como algo nuevo. Nos resultaba difícil plantear a nuestros estamentos internos un proyecto basado exclusivamente en una entrada de datos. Nuestra intención fue, pues, elaborar un diccionario de sinónimos con unos criterios lingüísticos modernos y empezar desde cero. La idea fue comprendida perfectamente por IBM que la

apoyó en todo momento. Finalmente, en 1990 firmamos el acuerdo con IBM y empezamos a trabajar en colaboración con el recién inaugurado *Centro de Tecnologías de la Lengua* en Sevilla.

El objetivo básico del proyecto es la elaboración de un diccionario catalán de sinónimos de entre ocho y diez mil entradas. El estado actual del proyecto es de más de ocho mil fichas acabadas. Creemos que podremos dar por finalizado el proyecto en Diciembre de este año.

En el proyecto DICCSIN trabajan dos coordinadores, un coordinador lingüístico y un coordinador informático, y un equipo de hasta seis lexicógrafos.

Como comentaremos más adelante, consideramos que actualmente es impensable un proyecto lexicográfico sin la ayuda de los recursos informáticos. Nuestro proyecto se beneficia del acceso al ordenador IBM 3090/600 del *Centre de Supercomputació de Catalunya (CESCA)* en el cual se almacenan las fichas lexicográficas y la base de datos *SQLDS* que contiene los diccionarios. Disponemos también de tres máquinas IBM PS/2 con un procesador INTEL 80286 en las que se elaboran las fichas lexicográficas que son transferidas al IBM 3090/600 para su proceso. Asimismo, disponemos de un IBM PS/2 modelo 70 con un procesador INTEL 80386 y 8MB de memoria RAM que utiliza el sistema operativo OS/2 v. 2.0 y se utiliza para mantener una copia de la base de datos de los diccionarios en el gestor de bases de datos relacionales *IBM Database Manager* destinada a elaborar programas prototipo que luego son pasados al IBM 3090/600 y a elaborar programas que permitan el acceso a la base de datos desde un entorno microinformático. La programación de aplicaciones se ha llevado a cabo en REXX, Turbo Pascal y KEXX.

EL CRITERIO GENERAL DE SINONIMIA

La simple observación de la mayor parte de diccionarios de sinónimos nos lleva a la conclusión de que éstos no aplican un criterio general de determinación sinonimia de manera rigurosa a todo el diccionario. Nuestra estrategia en el proyecto DICCSIN fue la de ceñirnos a un criterio general de determinación de sinonimia y elaborar una serie de subcriterios, siempre coherentes con el criterio general, que nos permitieran salir al paso de las incoherencias internas que con frecuencia nos encontramos en los diccionarios de la lengua.

Antes de exponer el criterio de determinación de sinonimia empleado en el proyecto creo que es conveniente entrar, aunque sea superficialmente, en el concepto de sinónimo.

Entendemos por sinónimo aquella palabra que presenta alguna acepción similar o idéntica a la de otra palabra. La relación de sinonimia es, pues, una relación entre significados. Dos palabras son sinónimas porque tienen uno o más significados comunes.

Ahora bien, ¿cómo determinamos si una palabra tiene una acepción similar o idéntica a la de otra palabra? Es aquí donde hay que marcar un criterio que nos permita decidir cuándo consideramos que dos palabras son sinónimas y cuándo no.

Entendemos que una palabra agrupa uno o más significados cada uno de los cuales se pone de manifiesto en contextos propios. Tomemos, por ejemplo, dos de los significados de la palabra *obligació* (obligación). Por una parte, tendríamos *aquello que uno tiene que hacer, deber impuesto por una ley, un contrato...* y por otra, *título que representa una parte de los capitales de una sociedad financiera, industrial, etc., que produce un interés fijo y es amortizable en un período limitado*.¹ Ciertamente, encontramos contextos que actualizan cada uno de estos significados. Ejemplos del primero de ellos los tendríamos en frases como: *Mi obligación es salvarle la vida. Cumplir con una obligación. Las obligaciones de un padre de familia.* Y para el segundo, hablaríamos de *Obligaciones del Estado*.

Siendo así, diremos que dos palabras son sinónimas para una acepción concreta cuando ambas son conmutables en una misma frase sin que se observe un cambio importante en el sentido de la frase.

¹ Fabra, Pompeu (1988).

Así, la palabra catalana *obligança* (obligación) es sinónima de *obligació* porque comparte con ésta el primero de sus significados. Podemos hablar de *Complir amb una obligança* (cumplir con una obligación), pero no de **Les obligances de l'Estat* (las obligaciones del Estado).

Nuestro criterio de sinonimia es muy similar al propuesto por D. A. Cruse:²

X is a **cognitive synonym** of Y if (i) X and Y are syntactically identical, and (ii) any grammatical declarative sentence S containing X has equivalent truth-conditions to another sentence S', which is identical to S except that X is replaced by Y.

De todos es sabido que son pocos los casos de sinonimia estricta e incluso en el caso de los llamados *geosinónimos* hay matices que hacen que la frase resultante de la substitución, a pesar de tener el mismo sentido o significado conceptual, proporcione información suplementaria. *Habemos tres lápices* y *Tenemos tres lápices* son expresiones sinónimas. Su significado conceptual es exactamente el mismo, pero la primera nos informa, además del hecho de que poseemos tres lápices, del posible origen del hablante. Lo mismo ocurriría con las frases *Pedro ha encontrado un trabajo* y *Pedro ha encontrado un curro*. El significado conceptual es el mismo, pero al contrastarlas, observamos un significado social distinto en ellas.³ En nuestro proyecto tenemos en cuenta los significados no conceptuales y los marcamos utilizando lo que llamamos *calificadores* (más adelante hablaremos de ellos).

Nuestro criterio de sinonimia es, sin duda, más restrictivo que el de la mayoría de diccionarios. Podría parecer, pues, que la cantidad de palabras consideradas como sinónimos de otra fuera menor en nuestro diccionario. La experiencia nos demuestra, sin embargo, que hay veces en que esto es así y otras en que no. Observemos las siguientes muestras de los diccionarios de Pey (1983) y Franquesa (1987):

PEY:

flauta: caramella, flautí, flabiol || flautista.

FRANQUESA:

flauta: xeremia, caramella, dolçaina, sonador, mirlitó, flabiol, flautí.

DICCSIN:

flauta: flaüta <U dialectal> || flautista.

Podemos ver que Pey y Franquesa incluyen como sinónimos palabras que lo son (*flautista*), hiperónimos (*sonador*) y cohipónimos (*caramella, dolçaina, flabiol, etc.*). En nuestro caso sólo figura un sinónimo (dialectal) *flaüta* para la primera acepción y un sinónimo para la segunda *flautista*.

Sin embargo, en otros casos nuestro criterio nos permite incluir más sinónimos que los otros diccionarios. Es el caso, por ejemplo de *golafre* (glotón):

PEY:

golafre: golut, fart, menjador, destraler, menjaire, insaciable, famolenc.

FRANQUESA:

golafre (nombre): fart, fartaner, llepaplats, escuracassoles.

²D.A. Cruse (1986), pp. 88-89.

³ Para una mayor profundización sobre los distintos tipos de significado, ver Leech, G. (1985) pp. 27 y ss.

DICCSÍN:

- golafre (nombre): gormand, golut, fartaner, glot, gorjut, fart, destraler, golós, llefre, glotó, escuracassoles, llepaplats, escuraplats, tragador <U antigament>, dragador || golut, glot, gorjut, golós, glotó, llefre, llepaplats, devorador.
- golafre (adjetivo): gormand, golut, glot, gorjut, pafart, destraler, golós, llefre, glotó, dragador, tragador <U antigament> || golut, glot, gorjut, golós, glotó, llefre, insaciable, insadollable, voraç, devorador.

El criterio de sinonimia que utilizamos es lo suficientemente restrictivo como para evitar el conocido *efecto de cadena* que se suele encontrar en la mayor parte de los diccionarios de sinónimos (es decir, *a* puede ser un sinónimo de *b*, *b* un sinónimo de *c*, *c* un sinónimo de *d*, siendo *d* a la vez el antónimo de *a*). Por ejemplo, en Pey (1983), es posible llegar al antónimo siguiendo la serie de sinónimos que figuran en una determinada entrada:

claror —> besllum —> ombra —> obscuritat (claridad, trasluz, sombra, oscuridad)

Hemos observado que el uso de nuestro criterio de sinonimia evita el efecto de cadena.

Para acabar, decir que al aplicar en la práctica nuestro criterio de sinonimia han aparecido, de forma persistente, dos obstáculos. La falta de normalización de los diccionarios de la lengua⁴ y la dificultad para encontrar ejemplos en los que aplicar nuestro criterio. No siempre los ejemplos que vienen en el diccionario de la lengua son adecuados y no disponemos de un corpus textual grande para poder extraer los diversos contextos de una palabra. En ambos casos hemos tenido que recurrir a la competencia lingüística de los lexicógrafos.

EL DICCIONARIO DE SINÓNIMOS

Nuestro criterio de sinonimia se basa, como hemos visto, en un contraste contextual de dos palabras. Por eso es natural que nuestro diccionario incluya los ejemplos y las definiciones que nos han permitido establecer un sinónimo. Este hecho no es frecuente en los diccionarios de sinónimos, que se limitan a marcar una separación entre acepciones sin que el lector pueda saber exactamente a qué corresponde cada acepción.

La relación de sinonimia se corresponde con una de las relaciones léxicas básicas: la de identidad. Por eso, la estructura que hemos diseñado para nuestro diccionario permite incluir las otras relaciones de una manera sencilla. Así, con la finalidad de poder hacer un mejor seguimiento de las series sinonímicas, decidimos incluir los hiperónimos de las palabras que los tuvieran. La implementación de nuestra estructura en un gestor de bases de datos relacionales nos permite obtener más información de los datos que tenemos almacenados. Así, podemos obtener fácilmente los cohipónimos y podemos observar la clasificación taxonómica resultante. Podemos, por ejemplo, ver qué instrumentos tenemos en el diccionario pidiendo *todas las palabras que tengan instrumento como hiperónimo*. Observemos, por ejemplo, la ficha de *flauta* ahora ya con hiperónimos:

⁴ Rafael i Fontanals, J. (1988) trata de manera sistemática esa falta de normalización ejemplificándola en los diccionarios catalanes. Apunta los defectos básicos de esos diccionarios (definiciones múltiples, definiciones repetidas y circularidad vacua) y propone un modelo normalizado de artículo.

flauta:f

1. <H> instrument;
<S> flaūta <U dialectal>;
2. <H> músic, tocadora, sonadora, music <U pejoratiu>,
instrumentista;
<S> flautista.

flauta:m

- <H> músic, tocador, sonador, music <U pejoratiu>, instrumentista;
<S> flautista.

Gracias a la inclusión del diccionario en un gestor de bases de datos relacionales podemos formular preguntas cuyas respuestas difícilmente podrían ser obtenidas de un diccionario impreso convencional. Así, porejemplo, podemos preguntar *qué palabras tienen como sinónimo 'flauta'* o obtener todas las palabras que tengan un significado especial en algún dialecto...

Nuestro diccionario se construye sobre fichas lexicográficas. Estas fichas recogen toda la información de una palabra. Sus principales elementos son:

Cabecera:	Información sobre el estado de elaboración de la ficha, fecha de inicio, autores, comentarios generales.
Entrada:	Incluye un código de categoría gramatical. No hace falta escribirla. Un programa lo hace al crear la ficha.
Definiciones y Ejemplos:	Sobre los que se aplica el criterio de determinación de sinonimia. Cada definición y cada ejemplo incluye una referencia a la fuente de donde proviene.
Sinónimos e Hiperónimos:	Lista de hiperónimos y sinónimos separados por acepciones.
Calificadores:	Marcan una restricción en el ámbito de una palabra. Hay tres niveles. El primero restringe una entrada a nivel morfológico. Por ejemplo, si un verbo es transitivo e intransitivo, separa los dos casos y distingue acepciones para cada uno de ellos. También se utiliza para marcar algún significado no conceptual; por ejemplo, <i>vulgar, familiar, infantil...</i> Los de segundo nivel tienen las mismas funciones que los de primer nivel pero distinguen acepciones que ya tienen un calificador de primer nivel. Así, un verbo con un calificador <i>transitivo</i> puede tener un calificador de segundo nivel <i>figurado</i> . El tercer nivel lo ocupan los que llamamos <i>calificadores individuales</i> . Éstos restringen el ámbito de un sinónimo. Así <i>bolquer</i> (pañal), tiene como sinónimo <i>drap</i> (trapo) pero esta palabra sólo funciona como sinónimo en el dialecto balear, por lo que hay que añadir un calificador individual a <i>drap</i> que indique que sólo en algún dialecto tiene el sentido de 'pañal'.

El siguiente ejemplo de ficha ilustra lo dicho hasta ahora sobre sus componentes:

```

*****
* MOT:      escandalitzar
* GRUP:     ALBERT, MONTSERRAT
* ESTAT:    A
* DATA:    21.10.91
* COMENTARI: Davant l'evidència, hem posat el qualificador absolut
*           a la segona accepció (vegeu [A]).
*****
escandalitzar:vtp
  <U transitiu>
<D> [E] ofendre els sentiments morals, la consciència (d'algú) amb una
      acció immoral blasmable
<E> [E] això que penseu fer escandalitzaria a tothom | [L] aquell
      strip-tease escandalitzà els més puritans
      <H> -
      <S> -

      <U2 absolut>
<D> [E] fer escàndol, xivarri
<E> [A] han passat una partida de jovenots escandalitzant pel carrer
      <H> cridar <U intransitiu>, bagolar <U intransitiu>;
      <S> tabalejar <U intransitiu>, esvalotar <U intransitiu>,
      avalotar <U intransitiu>;

      <U pronominal>
<D> [E] ésser escandalitzat
<E> [E] tota la gent de bé se n'escandalitzaria | escandalitzar-se a
      la vista d'una cosa | [L] us escandalitzeu per poca cosa
      <H> -
      <S> -.

```

En esta ficha podemos observar que tanto las definiciones como los ejemplos van precedidos de un código que indica su procedencia de manera que siempre podemos conocer la fuente del ejemplo o la definición en la que aplicamos el criterio de sinonimia. Así [E] se corresponde con Enciclopèdia Catalana (1990), [L] con Larousse (1990) etc.. La ficha ejemplifica también el uso de calificadores de primer nivel: <U transitiu>, <U pronominal>, de segundo nivel <U2 absolut> e individuales <U intransitiu>.

LAS HERRAMIENTAS INFORMÁTICAS

Como hemos avanzado más arriba, no entendemos que en la actualidad pueda realizarse un proyecto lexicográfico sin la ayuda de la informática. La cantidad de información manejada, la posibilidad de aumentar de la productividad y la eliminación de errores así lo aconsejan. Describiremos a continuación algunas de las herramientas desarrolladas específicamente para este proyecto agrupándolas según la necesidad que cubren.

El manejo de información es importante en cualquier proyecto lexicográfico. Nosotros hemos desarrollado herramientas que nos permiten acceder de manera instantánea presionando una combinación de dos teclas a la información del diccionario principal (con más de 93.000 entradas). Asimismo, podemos interrogar el conjunto total de las fichas lexicográficas haciendo preguntas del tipo *¿en qué fichas aparece la palabra 'casa' dentro de algún ejemplo o dentro de algún comentario o como hiperónimo?* o bien *¿en qué fichas aparece 'casa' pero no como hiperónimo?*.⁵ Preguntas que son extremadamente útiles durante el proceso de elaboración de las fichas. También el proceso de gestión de fichas está automatizado. No es posible que dos grupos trabajen al mismo tiempo con la misma ficha. Hay un programa de gestión de fichas que bloquea una ficha si alguien está trabajando con ella. Si otra persona quiere trabajar con ella, puede preguntar al ordenador quién la está utilizando y ponerse de acuerdo con esta persona. Las fichas sólo se bloquean, lógicamente, en modalidad de escritura. Para consultarlas no hay restricciones. Todas las operaciones de modificación de fichas actualizan un registro de fichas que nos permite saber en todo momento si una ficha está acabada o no, si alguien está trabajando con ella, en qué fecha se hizo la última modificación y quién la hizo etc. Por otra parte, las posibilidades de interrogación de la base de datos SQL son ilimitadas. Disponemos de menús de preguntas que permiten obtener los cohipónimos de una palabra o las palabras que tienen un mismo sinónimo o todas las combinaciones válidas de códigos de categorías gramaticales... Además, hay que tener en cuenta que el lenguaje de interrogación SQL es un lenguaje abierto y por tanto cualquiera puede hacer la pregunta que le interese de una manera sencilla aunque ésta no esté prevista en ningún menú de preguntas.

Hemos hablado también del incremento de la productividad que obtenemos empleando herramientas informáticas. Esta mayor productividad viene dada por una mayor velocidad en el acceso a la información, un menor número de errores y una facilidad en la localización de los mismos.

Sobre el arduo problema de la detección y eliminación de errores, hemos desarrollado herramientas en dos direcciones. Las primeras pretenden evitar los errores producidos por una acción mecánica como el copiado y las segundas se encaminan a la detección de errores de coherencia. En la primera línea está la aplicación de creación de fichas. El lexicógrafo sólo tiene que indicar al programa de qué palabra quiere abrir ficha. Automáticamente, el programa busca esa palabra en el diccionario principal y coloca en las posiciones convenientes toda la información que sobre esa palabra hay en el diccionario. Asimismo, confecciona la cabecera y rellena campos de fecha y autores. En todo ese proceso no hay posibilidad de error. Si el lexicógrafo tuviera que reproducir a mano la información del diccionario principal, muy posiblemente cometería algún error. Lo mismo ocurre con la colocación de campos. El lexicógrafo no tiene que introducirlos a mano. Simplemente pulsando una combinación de teclas el campo se coloca en la posición correspondiente. Además, si algún campo debe llenarse coloca el cursor en el punto adecuado para que sólo se tenga que empezar a escribir. Se verifica también el contenido de algunos campos. Así, en definiciones y ejemplos no permite poner un código de fuente incorrecto. Una vez la ficha ha sido elaborada, se pasa por un corrector ortográfico que detecta el 95% de los errores de escritura. Después de este proceso, creemos que el porcentaje de errores con origen en operaciones mecánicas queda muy reducido.

La determinación de errores de coherencia se desarrolla en dos puntos del proceso de producción. Una primera fase, durante el proceso de paso de las fichas a la base de datos SQL y una segunda fase sobre los datos ya cargados en SQL. A esta segunda fase, por su importancia, le dedicaremos un apartado especial.

El proceso de paso de fichas a SQL tiene un objetivo evidente que consiste en transformar fichas en formato textual a un formato alfanumérico y almacenarlo en la base de datos. El formato alfanumérico final es poco práctico desde el punto de vista humano pero conveniente desde el punto de vista informático ya que se ahorran recursos de almacenamiento. Veamos, por ejemplo el resultado de cargar la ficha del sustantivo femenino *flauta*:

⁵ Hay 14 tipos de preguntas distintas previstas por la aplicación.

Palabra	Sinónimo	Núm. Aceptión	Tipo	Calificador 1	Calificador 2	Calificador I
41834	50824	1	H	0	0	0
41834	41854	1	S	0	0	10
41834	50833	2	H	0	0	0
41834	60266	2	H	0	0	5
41834	60508	2	H	0	0	0
41834	81617	2	H	0	0	0
41834	85499	2	H	0	0	0
41834	41842	2	S	0	0	0

Los números que ocupan las columnas de *palabra* y *sinónimos* son identificadores exclusivos de una entrada del diccionario principal. Así, el número 41.834 se corresponde con la entrada *flauta f* del diccionario principal, el 41.854 con *flautista f* y el 41.842 con *flautista f*. El paso del formato de ficha a formato de tabla SQL exige que se hagan una serie de comprobaciones. La primera consiste en verificar que cada entrada exista en el diccionario principal. Si es así, se obtiene el número correspondiente. Lo mismo hay que hacer para cada uno de los sinónimos y de los hiperónimos. En este caso el proceso es un poco más complejo ya que en la ficha sólo las entradas tienen código de categoría gramatical y por tanto son identificables con facilidad en el diccionario principal. Con frecuencia un posible sinónimo tiene más de una entrada en el diccionario principal y hay que escoger la correcta para obtener el número que irá a la tabla SQL. El programa de transformación de fichas se encarga de eso. Obtiene los posibles candidatos y elige de entre ellos el correcto. Si no hay ningún candidato, avisa del error para que pueda ser enmendado. Uno de los aspectos que tiene en cuenta para aceptar un candidato es que los códigos de categoría gramatical sean compatibles. Entra dentro de lo posible que un error de escritura haya generado una palabra que existe en el diccionario pero que no tiene sentido en esta entrada. El resultado puede ser que un sustantivo tenga como sinónimo un adverbio. El programa de carga dispone de una tabla con todas las compatibilidades de códigos admitidas y evita totalmente situaciones como la que acabamos de exponer.

Llegados a este punto, consideramos que la mayor parte de errores mecánicos han sido evitados o detectados. La detección de errores de coherencia la hacemos sobre la base de datos SQL. Este va ser el tema del próximo apartado.

LA VERIFICACIÓN DE CONGRUENCIA EN SERIES SINONÍMICAS

Hemos visto más arriba que la relación de sinonimia es una relación entre significados. Hemos visto también que una palabra puede tener uno o más significados. Así, por ejemplo, la palabra *renillar* (relinchar) sólo tiene un significado, «gritar el caballo». Todos los sinónimos de *renillar* lo serán para ese único significado. En catalán hay tres palabras que pueden sustituir a *renillar* en una frase: *ahinar*, *arniar* y *eguinar*.

El cavall va **renillar** tota la nit (El caballo relinchó toda la noche)
 ahinar
 arniar
 eguinar

La palabra *renillar* tiene, pues, tres sinónimos para su único significado. Sin embargo, la mayor parte de palabras tienen más de un significado. Por ejemplo la palabra *parla* (habla) tiene, entre otros, los siguientes significados:

1. Facultad de hablar: *paraula* (palabra)
'perdre la parla' (perder el habla)
2. Manera como uno expresa su pensamiento por medio de la palabra: *llenguatge* (lenguaje)
'no entenc aquesta parla' (no entiendo este lenguaje)
'parla culta, grollera' (lenguaje culto, grosero)
3. Lengua de una comunidad o pequeño grupo social: *argot*
'la parla dels gitanos' (el hablar de los gitanos)
4. Lengua de una pequeña región: *subdialecte* (subdialecto), *parlar* (habla)
'la parla de la ciutat de Lleida' (el habla de la ciudad de Lérida)

Observamos, pues, que para cada significado encontramos una o más palabras que lo comparten. Cada uno de estos grupos de palabras con uno de sus significados en común es lo que llamamos una *serie sinonímica*. La siguiente figura ilustra la serie sencilla, con un solo significado, de *renillar*.

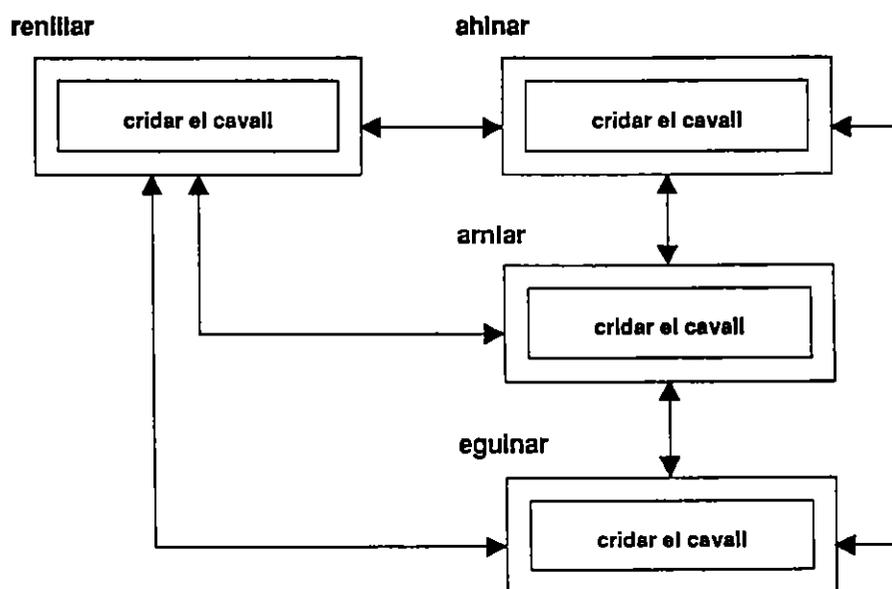


Figura 1: La serie sinonímica de 'renillar'

Observamos que las palabras de la serie sólo tienen un significado y que éste es el mismo para todas. Así, a partir de la palabra *renillar* podríamos deducir fácilmente los sinónimos de las otras palabras de la serie:

renillar:

ahinar, arniar, eguiñar

ahinar:

renillar, arniar, eguiñar

arniar:
 ahinar, renillar, eguinar
 eguinar:
 ahinar, arniar, renillar

Generalizando lo visto hasta ahora, podríamos establecer que si una palabra P_1 para una acepción A_n tiene como sinónimos S_1, S_2, \dots, S_n debe cumplirse que

S_1 tenga una acepción A_x con los sinónimos P_1, S_2, \dots, S_n
 S_2 tenga una acepción A_y con los sinónimos S_1, P_1, \dots, S_n
 [...]
 S_n tenga una acepción A_z con los sinónimos S_1, S_2, \dots, P_1

El principio aquí expuesto es lo que llamamos *principio de congruencia*. Hay que puntualizar, sin embargo, que no es lo mismo congruencia que corrección. Entendemos que si una serie es correcta es congruente, pero no al revés. Si decimos que *gato* tiene como sinónimos *perro* y *conejo* y queremos ser congruentes con esta afirmación, hemos de decir que *perro* tiene como sinónimos *gato* y *conejo* y que *conejo* tiene como sinónimos *perro* y *gato*. Lógicamente, nadie diría que esta serie sinonímica es correcta porque es obvio que la relación de sinonimia atribuida a *gato*, *perro* y *conejo* es falsa, sin embargo el planteamiento sería congruente.

Supondremos aquí que la relación de sinonimia establecida por los lexicógrafos es correcta y nos centraremos exclusivamente en la verificación de congruencia de la serie. Fácilmente nos podemos imaginar la inmensa complejidad que supone elaborar un diccionario de sinónimos siguiendo estrictamente el principio de congruencia entre las palabras de una serie. Un ejemplo. Una palabra con cuatro acepciones y tres sinónimos por acepción nos obligaría a comprobar un total de doce acepciones de tres palabras cada una. Consideramos que es del todo imposible realizar manualmente estas comprobaciones de una manera sistemática para todas y cada una de las palabras de un diccionario de sinónimos.

La imposibilidad material de hacer las verificaciones manualmente es lo que hace que con frecuencia encontremos una falta de congruencia en el tratamiento de las series sinonímicas. Que sepamos, hasta ahora no se ha elaborado ningún diccionario catalán de sinónimos utilizando medios informáticos y la verificación total y sistemática de la congruencia de las series sólo es posible si se utilizan esos medios. Un ejemplo de lo que suele pasar en los diccionarios tradicionales es el tratamiento de la serie de *renillar*. Una serie sencilla, con un sólo significado en cada palabra de la serie. A guisa de ejemplo, expondremos el tratamiento que hacen de la serie encabezada por la palabra *renillar* el diccionario Pey (1989).

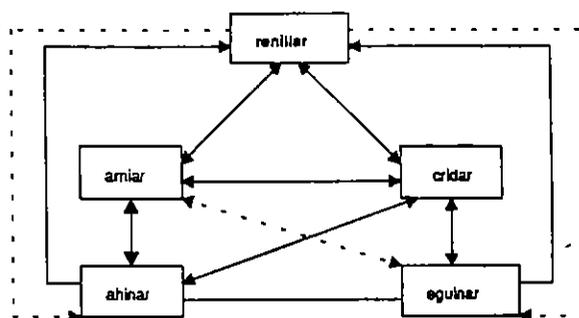


Figura 2: La serie de 'renillar' en Pey (1989)

En la figura, las líneas continuas indican las relaciones establecidas por los diccionarios y las discontinuas las que debieran haber establecido. Los recuadros en línea continua indican que la palabra tiene entrada en el diccionario y los dibujados en línea discontinua indican que no hay una entrada en el diccionario para la palabra del recuadro. Así, observamos en Pey (1989), todas las palabras de la serie tienen entrada, pero hay una asimetría doble entre *renillar* y *ahinar* y *eguinar*. Observamos que hay dos tipos de incongruencia posibles: la asimetría y la referencia a palabras que no tienen entrada en el diccionario.⁶ Estos son, pues, los defectos que detecta nuestro programa de verificación de congruencia.

El sistema utilizado por el programa es sencillo. A partir de una acepción de una palabra, crea una acepción teórica para cada una de las palabras que aparecen en la serie y comprueba la existencia de esa acepción teórica en cada una de las palabras. Si no hay ninguna acepción igual a la teórica, el programa nos informa del error; pero aún va un poco más lejos. Supongamos el caso de *arniar* en Pey (1989). El programa nos informa que no hay ninguna acepción igual a la teórica deducida de *renillar* y además nos dice si *renillar* aparece en alguna acepción de *arniar*. Es posible que el error consista en haber olvidado, por ejemplo, *eguinar*. Esta información facilita la localización del error por parte del lexicógrafo.

El programa de verificación de congruencias complementa la detección y eliminación de errores en el proceso de elaboración del diccionario. Creemos que los resultados finales pueden ser de interés tanto por la metodología empleada, basada en la aplicación uniforme de unos criterios lingüísticos y una sistemática de producción altamente informatizada, como por las posibilidades de consulta de la base de datos que configura el diccionario.

BIBLIOGRAFÍA

- ALCOVER, A. M. y MOLL, F. B., (1988) *Diccionari Català-Valencià-Balear* (10 volúmenes), Ed. MOLL, Palma de Mallorca.
- ARIMANY, M. (1979) *Diccionari pràctic de sinònims catalans, mots i frases* Ed. Miquel Arimany, S.A. Barcelona, 2ª edición.
- ARIMANY, M. (1982) *Diccionari usual de sinònims catalans, mots i frases* Ed. Miquel Arimany, S.A. Barcelona, 2ª edición.
- BIBLIOGRAF (1989)₁, *VOX. Diccionari manual de sinònims i antònims* Ed. BIBLOGRAF, S.A., Barcelona, 5ª edición (reimpresión)
- BIBLIOGRAF (1989)₂, *VOX. Diccionari manual Castellà-Català Català-Castellà* Ed. BIBLOGRAF, S.A., Barcelona, 9ª edición (reimpresión).
- COLLINS (1989)₁, *Collins Diccionario: Español/Inglés, Inglés/Español* Ed. GRIJALBO S.A., Barcelona.
- COLLINS (1989)₂, *Collins Dictionary of the English Language* W. Collins Sons & Co. Ltd., Glasgow.
- CRUSE, D. A. (1989) *Lexical Semantics* Cambridge University Press, Cambridge.
- EDHASA (1989) *Diccionari manual de la llengua catalana* Ed. EDHASA, Barcelona, 6ª reimpresión.
- ENCLICLOP_DIA CATALANA (1989) *Gran enciclopèdia catalana* (24 volúmenes), Ed. Enciclopèdia Catalana, S.A., Barcelona, 2ª reimpresión.

⁶ En Franquesa (1987), por ejemplo, *ahinar* y *eguinar* no tienen entrada en el diccionario pero son referenciadas por *renillar* y *criidar*.

- ENCICLOPEDIA CATALANA (1990) *Diccionari de la llengua catalana* Ed. Enciclopèdia Catalana, S.A., Barcelona, 10ª edición (reimpresión).
- FABRA, Pompeu (1988) *Diccionari general de la llengua catalana* Ed. EDHASA, Barcelona, 24ª edición.
- FRANQUESA, M. (1987) *Diccionari de sinònims* Ed. Pòrtic, S.A., Barcelona, 5ª edición.
- LAROUSSE (1990) *Gran Larousse Català* Ed. Edicions 62, Barcelona.
- LEECH, Geoffrey (1985) *Semántica* Ed. Alianza Editorial, Madrid.
- LONGMAN (1990) *Longman Dictionary of Contemporary English* Longman Group UK Limited, Essex.
- McARTHUR, Tom: *Longman Lexicon of Contemporary English* Longman Group UK Limited, Essex.
- MOLINER, María (1989) *Diccionario de uso del español* (2 volúmenes), Ed. GREDOS S.A., Madrid.
- PEY, Santiago (1989) *Diccionari de sinònims i antònims* Ed. TEIDE, Barcelona, 11ª edición.
- PITTÀNO, Giuseppe (1988) *Sinonimi e Contrari: Dizionario Fraseologico delle Parole Equivalenti, Analoghe e Contrarie* Ed. Zanichelli, Bologna.
- RAFEL i FONTANALS, Joaquim (1988) *Sobre la normalització dels diccionaris catalans: el tractament de les sèries sinonímiques* en *Miscel·lània Antoni M. Badia i Margarit* 8, Publicacions de l' Abadia de Montserrat, Barcelona.
- RASPALL i JUANOLA, J. y MARTÍ i CASTELL, J. (1989) *Diccionari de locucions i de frases fetes* Ed. Edicions 62, Barcelona, 4ª edición.
- REAL ACADEMIA ESPAÑOLA (1984) *Diccionario de la lengua española* (2 volúmenes), Ed. Espasa-Calpe, S.A., Madrid.
- REAL ACADEMIA ESPAÑOLA (1989) *Diccionario manual e ilustrado de la lengua española* Ed. Espasa-Calpe, S.A., Madrid, 4a. edición revisada.
- SALVAT (1989) *Salvat Català, diccionari enciclopèdic* (10 volúmenes), Salvat Editores, S.A., Barcelona.