

La explotación de corpora textuales en soporte electrónico

Xavier BLANCO
Universidad Autónoma de Barcelona

0. INTRODUCCIÓN

El presente artículo se propone mostrar cómo el empleo de corpora textuales puede resultar útil para la descripción lingüística, incluso cuando ésta presenta una naturaleza esencialmente hipotético-deductiva. Si bien, por una parte, todo parece indicar que el recurso a un corpus, por extenso que éste sea, nunca bastará para la modelización de ciertos fenómenos lingüísticos, nada lleva a pensar, por otra parte, que la verificación de las hipótesis del investigador no exija una confrontación empírica. En otras palabras, parece posible convenir en que el uso de un corpus no es suficiente para la descripción lingüística, no así en que no es necesario.

Con tal objeto, procederemos a contrastar una descripción metalingüística explícita del lexema francés *invent-* con su empleo real en discurso tal y como se observa en un corpus periodístico formado por 21.804.745 ocurrencias (241.893 formas distintas), el cual reúne todos los números del periódico francés *Le Monde* correspondientes a 1992.

Del mismo modo, emplearemos, a título de ilustración, un pequeño corpus de español periodístico (*El País*), a fin de mostrar algunas de las posibilidades del uso de corpora con fines traductológicos.

Llegado el caso, haremos referencia también a diccionarios monolingües del francés (PR, Lexis) y del español (DRAE, DUE), así como a diccionarios bilingües francés-español, español-francés (LFE-LEF, BFE-BEF), para completar los aportes de la principal fuente metalingüística empleada (cf. 4.1).

1. LA DESCRIPCIÓN METALINGÜÍSTICA EXPLÍCITA

Dada la riqueza del paisaje diccionario en lengua francesa, no resulta difícil disponer de descripciones lexicográficas muy completas del léxico francés, algunas de ellas basadas en grandes corpora textuales (v.g. *Trésor de la langue française*). No obstante, las finalidades pedagógicas y/o normativas de dichas obras, unidas a su condición de productos comerciales, hacen que sus microestructuras rara vez ofrezcan una información lo suficientemente completa, explícita y formalizada para servir de base a un estudio lingüístico en profundidad de una forma determinada.

Los trabajos llevados a cabo en el LADL (Paris VII) y en el LLI (Paris XIII), encaminados a la construcción de un léxico-gramática que supone un inventario sistemático del léxico francés y de sus principales propiedades sintáctico-semánticas, parecen ofrecer un punto de partida más adecuado desde el punto de vista de la investigación lingüística. Partiremos, pues, de las entradas correspondientes a la raíz *invent-* en el léxico-gramática (Gross, M. 1975: 293) (Gross, M. 1981: 377), a fin de comprobar en qué medida la confrontación de las mismas con un gran corpus textual válida o contradice la descripción propuesta.

Dentro del marco teórico citado, se contemplan cuatro entradas correspondientes a *inventer*. La primera de ellas (i), que podríamos asociar a la definición del Lexis «créer une chose, originale ou nouvelle, à laquelle personne n'avait encore pensé, dans tous les domaines de l'activité» (s.v. **inventer** 1 | 1), contempla las siguientes posibilidades:

- (A1) *Max a inventé (l'aube + la roue à aubes).*
- (B2) *Max invente (E + sans arrêt + des choses diverses).*
- (B3) *Max est un inventeur.*
- (B4) *Max fait des inventions.*
- (B5) *Max est inventif.*
- (B3f) *Ida est une inventrice.*
- (B5f) *Ida est inventive.*
- (B6) *Max a (de la + une certaine) inventivité.*
- (B6d) *Max est d'une certaine inventivité.*
- (B6n) *Max n'a (pas de + aucune) inventivité.*
- (B6s) *Max est sans inventivité.*
- (A3d) *Max est l'inventeur de l'aube.*
- (A7) *L'aube a un inventeur.*
- (A1p) *L'aube a été inventée par Max.*
- (A8) *L'aube est (la + une) invention de Max.*
- (A8d) *L'aube est (de l'invention de Max + son invention).*
- (A8j) *Il y a eu invention de l'aube par Max.*
- (A1pm) *L'aube pourrait être inventée par un non-spécialiste.*
- (A9) *L'aube est inventable par un non-spécialiste.*
- (A1pmn) *L'aube ne pourrait pas être inventée par un non-spécialiste.*
- (A9n) *L'aube n'est pas inventable par un non-spécialiste.*

(A11) *L'aube est ininventable par un non-spécialiste.*

(A10) *De tels procédés ont une certaine inventabilité.*

(A10n) *De tels procédés n'ont aucune inventabilité.*

(A12) *De tels procédés ont une ininventabilité totale.*

(A13) *Max a fait l'invention de l'aube.*

(B3f) *Léa est une inventrice.*

(A3df) *Léa est l'inventrice de l'aube.*

(A17f) *L'aube a une inventrice.*

Como puede observarse, una entrada del léxico-gramática constituye una clase de equivalencia que reúne un conjunto de frases simples vinculadas por relaciones transformacionales. Dicha clase puede representarse de manera sencilla mediante un grafo en que cada nodo corresponda a una frase simple y cada arista (dotada de una notación mnemotécnica, v.g. «p» para «pasiva», «m» para «inserción de modalidad», «n» para «negación», etc.) a una relación transformacional.

La segunda entrada (ii) correspondiente a *inventer*, puede asociarse a la definición «concevoir quelque chose qui, dans une occasion déterminée, serve à un usage particulier» (Lexis, s.v. **inventer** 1 | 2). Por razones de espacio, sólo transcribiremos la primera frase de las siguientes tres entradas.

(C1) *Max a inventé une (solution de rechange + astuce pour s'absenter).*

En cuanto a la tercera entrada (iii), podría definirse como «concevoir d'une manière arbitraire, concevoir quelque chose de fictif» (Lexis, s.v. **inventer** 1 | 3):

(D1) *Max a inventé cette histoire.*

Finalmente (iv), hallamos una acepción técnica de *inventer*, correspondiente a «trouver» (Lexis, s.v. **inventer** 2):

(E1) *Max a inventé ce trésor.*

Aún cabría añadir, a la descripción de la raíz *invent-*, el N *invention* en *Invention de Bach* (composición musical), los N compuestos *l'invention de la Sainte Croix* y *le Bureau des inventions* y expresiones como *Max n'a pas inventé (la poudre + l'eau (tiède + chaude) + le fil à couper le beurre)*; *Si Max n'existait pas, il faudrait l'inventer*; *Max ne sait (E + pas + plus) quoi inventer pour nous agacer* o *Max est allé inventer N1*.

2 LA CONSULTA DE CORPORA TEXTUALES

2.1 LA DESCRIPCIÓN MONOLINGÜE

Acto seguido, procederemos a contrastar las informaciones presentadas en el apartado anterior con el empleo en discurso del lexema *invent-*, sirviéndonos para ello del corpus arriba anunciado (cf. 0). Precisemos, no obstante, que no siendo posible un estudio en profundidad de las formas retenidas, nos limitaremos a resaltar algunos aspectos de interés que puedan resultar reveladores de la utilidad de un corpus como sistema de retroalimentación de una descripción de carácter metalingüístico.

El lexema *invent-* presenta en *Le Monde* 1992 un total de 2.041 ocurrencias, cuya repartición es la siguiente:

31 INVENTA; 6 INVENTAIENT, 26 INVENTAIT, 50 INVENTANT;
184 INVENTE; 51 INVENTENT; 373 INVENTER; 4 INVENTERA;
1 INVENTERAIENT; 2 INVENTERAIT; 1 INVENTERONT; 171 INVENTEUR;
53 INVENTEURS; 49 INVENTIF; 27 INVENTIFS; 385 INVENTION;
94 INVENTIONS; 35 INVENTIVE; 16 INVENTIVES; 31 INVENTIVITÉ;
1 INVENTRICE; 11 INVENTERENT; 294 INVENTÉ; 68 INVENTÉE;
30 INVENTÉES; 47 INVENTÉS.

Cabe señalar que ignoramos 316 ocurrencias de INVENTAIRE(s), 38 ocurrencias de las distintas formas flexivas de INVENTORIER y 1 ocurrencia de INVENTORIAGE ya que, si bien presentan la misma raíz de origen que las arriba retenidas (*inventire*), no parece que exista ninguna relación en sincronía entre dichas formas y las presentadas en las cuatro entradas del léxico-gramática citadas¹.

El análisis de las concordancias correspondientes a las ocurrencias de *invent*² revela que, como cabía esperar, predominan las ocurrencias del verbo, especialmente del propio infinitivo con 373 casos. Se trata, en su gran mayoría, del empleo correspondiente a C (v.g. «...personnage principal du livre, et inventer une façon de montrer les scènes d'...»), seguido a distancia por D (v.g. «...moments, j'avais des scrupules à inventer l'histoire de mes ancêtres...»). En cambio, los casos A y B apenas si están representados, pese a que las descripciones metalingüísticas parten siempre de dicho empleo para la presentación del lema *inventer*.

Señalemos que un listado de las formas que ocupan la posición N1 regida por *inventer* permite observar que, de hecho, el sentido del verbo viene dado por el tipo de complementos que éste adopta. Así, encontramos N1 como *engin, machine, satin loutre* correspondientes a A, otros como *un style, l'avenir, le programme «habitat et vie sociale», des esthétiques, les projets les plus fous, des spots kitsch* correspondientes a C y una tercera serie como *passé, histoire, mythe* vinculados a D. De manera general, podríamos caracterizar como N concretos los N1 asociados al sentido A. El sentido C requeriría N abstractos en dicha posición y el sentido D se caracterizaría por la presencia de complementos apropiados de verbos de dicción (*raconter, relater, narrer, rapporter...*)³. Una descripción basada en las clases de objetos (Gross, G. 1994: 15) adoptadas por cada uno de los sentidos de *inventer* parecería, pues, idónea para la desambiguación de dicha forma⁴.

1. Precisemos que la presentación conjunta de dichas entradas tampoco implica que exista un vínculo en sincronía entre las mismas.

2. Por razones de espacio, no ofreceremos el total de las concordancias, sino que nos limitaremos a extraer de ellas los ejemplos necesarios para nuestra exposición.

3. Incluyendo las completivas. Nótese también que esta hipótesis se ve reforzada por el carácter estilísticamente marcado de un caso como «...j'invente une histoire que je ne raconte pas...»

4. Con la salvedad de que, el hecho de tener como N1 de *inventer* nombres predicativos, nos llevaría a definir no sólo clases de objetos (empleo A), sino también clases de predicados (empleos C y D).

Respecto al infinitivo, destacaríamos por último la frecuencia en discurso de giros como *être à inventer* (v.g. «...d'un marché nouveau, tout était à inventer. Il fallait montrer que c'est une...») y *rester à inventer* (v.g. «...négociation européenne restent à inventer. En fait, constate Jacques Ion...») que, sin embargo, no quedan recogidos en ninguno de los diccionarios monolingües consultados.

Por lo que respecta a las formas flexivas del verbo considerado, encontramos ocurrencias pertenecientes tanto a A (v.g. «...Conrad Johannes Van Houten invente le chocolat en poudre. En 1847...») y B (v.g. «...la vieillesse, alors qu'un savant invente jeune. Vous venez de souligner...»), como a C (v.g. «...par l'objet d'amour. Elle invente la poésie de cette passion et...») y D (v.g. «...osera d'un rosé Trinquevedel (je n'invente rien !). Décidément, la bourride...»), si bien, de nuevo, el sentido C es el más usual.

Resulta interesante observar, respecto a D, que dicho sentido aparece a menudo en un empleo intransitivo vinculado con la negación. Así, si el único empleo intransitivo de *inventer*, según la descripción léxico-gramatical del mismo, correspondía a B, la consulta de un corpus permite constatar la existencia de un buen número de casos como «...ne connaît et qui s'appelle, je n'invente pas, M. Laperruque. Dans tous ses...». Es curioso comprobar que, sin embargo, dista mucho de constituer un caso excepcional. Notemos igualmente que, si bien suele tratarse de una forma de primera persona del singular, hallamos también la tercera persona: «...M. Deniau a raison de dire qu'il n'invente pas. Comme un bon conteur, il se...».

Respecto al N *invention*, una vez más observamos que el sentido más frecuente es C, que parece derivado de A mediante un fenómeno de transición semántica. Precisemos, no obstante, que *invention* C confiere a menudo un matiz peyorativo a su N1. Dicho extremo se ve, en ciertos casos, reforzado por un Adj (v.g. «...fait de ruse et d'inventions drolatiques pour déjouer des...»). En ocasiones, el carácter diaconnotativamente marcado queda reflejado, desde el punto de vista de la tipografía, por el empleo de las comillas (v.g. «...qu'une de ses propres «grandes inventions» par rapport au communisme...»). Podemos añadir que ninguno de los diccionarios consultados ofrece indicación alguna respecto al posible matiz peyorativo de *invention*.

Destaquemos, por otra parte, que en la descripción de *invent-* propuesta en el apartado 1 no se halla el empleo de *invention* correspondiente a «*faculté d'inventer, don d'imagination*» (Lexis, s.v. **invention** 2), posibilidad que sí contemplan los diccionarios y que se encuentra ampliamente representada en el corpus (v.g. «...lui sont réclamés, ensemble, invention et rigueur, hauteur de vue et souci...»).

En cuanto a *inventeur*, cabe destacar que sólo los sentidos A y B⁵ son recogidos en las descripciones metalingüísticas (léxico-gramática y diccionarios). Sin embargo, el empleo más frecuente en el corpus parece estar vinculado a C (v.g. «...ce régime un tout organique. Inventeur du «harcelement démocratique» du...» o «...États indépendants ! Pour l'inventeur de ce projet un peu fou,

5. También encontramos E, si bien se trata de una acepción técnica muy poco frecuente en discurso.

André...»). Como puede observarse, *inventeur* C suele implicar cierto matiz peyorativo de su N1, como sucedía con *invention* (cf. tb. «...constructeur du labyrinthe et l'inventeur de quelques autres tours pendables...»).

En lo tocante a *inventif*, la descripción metalingüística contempla los sentidos A-B y C. El corpus ofrece sólo casos de C. Subrayemos que este adjetivo rara vez aparece en solitario, sino que lo hallamos casi siempre acompañado de otros (v.g. «...cet État-capitale de l'Europe, inventif, souple, dynamique...»; «...et de la mémoire humaine. Inventif et enthousiaste, Jacques Ninio comme...»).

El N *inventivité* (B y C según la descripción metalingüística, sólo C en el corpus explotado) presenta la particularidad de encontrarse muy a menudo combinado con extensiones aspectuales de su verbo soporte *avoir*. Así, tenemos «...routine. La province déborde d'inventivité. Décentraliser, c'est tout...»; «...de suite qu'elle perd de son inventivité. La domination est toujours suivie...». Sería, por tanto, interesante que los diccionarios indicasen con qué tipo de extensiones aspectuales se combinan este y otros N predicativos, ya que se trata de una información léxicamente condicionada.

El análisis de corpora puede aportar aún muchos otros datos de interés, especialmente en lo que se refiere a la confirmación o refutación de hipótesis experimentales. Así, M. Gross avanza la hipótesis de que el prefijo *ré-* puede aplicarse a todas las formas A de *inventer* (v.g. *Max a réinventé l'aube*), pero no a las formas B, lo cual podría explicarse por el aspecto iterativo de *ré-*, que entraría en contradicción con el aspecto durativo de B. Sin embargo, entre las 160 ocurrencias correspondientes a las distintas formas del paradigma de *RÉINVENTER*, tenemos en el corpus casos como «...ils ne reproduisent jamais : ils réinventent. Ils ont étudié les tissus, les...», ocurrencia aparentada con B2 *Max invente (E + sans arrêt + des choses diverses)*.

Hemos visto también que algunas observaciones relativas a la frecuencia de ciertos empleos en discurso podrían igualmente tener consecuencias para la descripción metalingüística e incluso para la enseñanza/aprendizaje de FLE. Así, es importante hacer notar que entre las 21.804.745 ocurrencias de *Le Monde* 92 no hallamos ninguna atestación de *ne pas avoir inventé (la poudre + le fil à couper le beurre)* ni de *ne pas avoir inventé l'eau (chaude + tiède)*. Tampoco encontramos *Bureau des inventions* ni *si NO n'existait pas, il faudrait l'inventer*. Añadamos que tampoco tenemos ninguna ocurrencia de *inventer* D ni de *invention* en su acepción técnica musical. Dichas secuencias suelen quedar recogidas, sin embargo, por las descripciones metalingüísticas explícitas que, por otra parte, ignoran giros muy frecuentes en discurso como pueden ser *je n'invente pas* o *NO (est + reste) à inventer*⁶

2.2 LA DIMENSIÓN BILINGÜE

Si la confrontación de la descripción metalingüística monolingüe de una forma dada con el uso de dicha forma en un corpus parece presentar ya un interés

6. Señalemos, no obstante, que el tipo de corpus empleado (exclusivamente periodístico y de un nivel de lengua cuidado) puede desempeñar un papel importante en los juicios de disponibilidad de ciertas unidades. Sería, por tanto, conveniente consultar textos de procedencia más variada (literarios, técnicos, orales...).

innegable, la posibilidad de emplear corpora bilingües con objeto de completar las descripciones metalingüísticas contrastivas puede resultar aún más beneficiosa.

Señalemos que entendemos aquí por corpus bilingüe el formado por textos de similar naturaleza en dos lenguas distintas (en este caso, textos periodísticos). Excluimos, pues, el corpus formado por un texto y su traducción, ya que, si bien se trata de una herramienta preciosa para estudios de tipo traductológico, su empleo no está exento de ciertos inconvenientes⁷.

Debido a la necesaria brevedad de este trabajo, nos limitaremos a considerar sólo algunos aspectos relativos al funcionamiento, como equivalentes de traducción, de las formas verbales *inventer* e *inventar*.

Empecemos diciendo que el artículo *inventer* de los principales diccionarios bilingües francés-español suele limitarse a poner en relación dicha forma con el verbo *inventar*. Se trata, pues, de microestructuras muy pobres que se limitan a la ecuación **inventer** v. tr. = *inventar*, sin establecer distinción de ningún tipo entre las distintas acepciones de *inventer*. Esta descripción, que podría tolerarse si las formas francesa y española tuviesen un funcionamiento idéntico, se ve contrariada por el hecho de que *inventar* no presenta, según los diccionarios monolingües españoles, ni el sentido E ni el equivalente exacto de C. Así, frente a «concevoir quelque chose qui, dans une occasion déterminée, serve à un usage particulier», definición que parece convenir a *inventer* C, tenemos, en español, una acepción más limitada: «hallar, imaginar, crear su obra el poeta o el artista» (DRAE, **inventar** 2). Por lo tanto, y remitiéndonos estrictamente a las descripciones metalingüísticas, no podemos traducir todas las ocurrencias de *inventer* C por *inventar*. Ello supone una limitación importante, ya que, recordemos, se trata precisamente de la acepción más frecuente en el corpus empleado.

Si, a su vez, observamos el comportamiento de *inventar* en un corpus periodístico español⁸ tenemos numerosas ocurrencias del tipo «...capacidad de Johan Cruyff para inventar alternativas de juego a partir de...» que parecen encajar perfectamente en la definición de C a juzgar por el tipo de N1 (*alternativas de juego, una necesidad, la perestroika, la filosofía trascendental, etc.*). No parece, por tanto, que la limitación sugerida por los diccionarios monolingües españoles se verifique en la práctica.

3. CONCLUSIONES

Digamos, a modo de conclusión, que un corpus textual ofrece una serie de prestaciones de las que no parece justificado prescindir, máxime cuando los medios técnicos necesarios para su gestión resultan cada vez más accesibles. Si el recurso a la introspección sigue siendo, a nuestro entender, un medio

7. En efecto, desde el punto de vista técnico, se requiere un sistema capaz de alinear texto de partida y texto de llegada (Isabelle, 1993: 288). Por otra parte, en este tipo de corpus, una de las lenguas queda representada por traducciones, que tienden a asimilar el funcionamiento de un lexema dado en L2 al de su equivalente más cercano en L1 y, por tanto, conllevan el riesgo de neutralizar ciertas divergencias.

8. Nos servimos de un pequeño corpus constituido por 6 Mb de texto de *El País*.

privilegiado para el establecimiento de descripciones metalingüísticas de una forma dada, la consulta de un corpus constituye poco menos que un paso obligado para el lingüista, ya sea para llevar a cabo a un primer balizaje de su campo de estudio, ya sea para proceder a una verificación minuciosa de sus propuestas.

Sería preciso subrayar, sin embargo, que la interacción entre descripción metalingüística y corpus no se limita a labores de verificación de los empleos de una forma determinada, sino que un corpus puede utilizarse igualmente para la estimación de frecuencias, la actualización del léxico repertoriado, la creación de un diccionario de nombres propios, la extracción de información colocacional, la estimación de valores diasistemáticos, la desambiguación o la extracción de conocimientos. Del mismo modo, una descripción metalingüística lo suficientemente completa y explícita puede emplearse ventajosamente para el análisis de nuevos corpora que, a su vez, permitirán seguir perfeccionándola.

4. REFERENCIAS

4.1 DICCIONARIOS

ALVAR EZQUERRA, M. (dir.) (1990): *Diccionario actual de la lengua española*, Barcelona: Vox Biblograf (DALE).

DUBOIS, J. (dir.) (1993): *Dictionnaire de la langue française lexis*, Paris: Larousse (Lexis).

GARCÍA-PELAYO; GROSS, R. et J. TESTAS (1992) : *Grand Dictionnaire Français-Espagnol, Espagnol-Français*, Paris: Larousse (LFE-LEF).

RAE (1992) : *Diccionario de la Lengua Española*, 21ª edición, Madrid: Real Academia Española / Espasa Calpe S.A (DRAE).

REY-DEBOVE, J. et A. REY (dir.) (1993): *Le Nouveau Petit Robert*, Paris: Dictionnaires Le Robert (PR).

VIDAL, J.-P. (1989) : *Dictionnaire Français-Espagnol / Espagnol-Français Hispano Bordas*, Paris: Bordas (BFE-BEF).

4.2 ESTUDIOS

BLANCO, X. (1995): «Description et traduction d'opérateurs dans la lexicographie bilingue français-espagnol, espagnol-français» en LABELLE, J. (ed.): *Lexiques-Grammaires comparés et traitements automatiques*, Montréal: Université du Québec à Montréal, 299-310.

GROSS, G. (1994): «Classes d'objets et description des verbes», *Langages* 115, Paris: Larousse, 15-30.

GROSS, M. (1975): *Méthodes en syntaxe*, Paris: Hermann.

GROSS, M. (1981): «Sur la structure des articles d'un lexique-grammaire» en CIGNON, L.; C. PETERS (eds): *Linguistica Computazionale VI; Computational Lexicology and Lexicography. Special Issue Dedicated to Bernard Quemada I*, Pisa: Giardini Editori e Stampatori. Promoted by A. Zampolli, 365-394.

ISABELLE, P. (1993): «Les corpus bilingues : une nouvelle ressource pour le traducteur» en CLAS, A.; P. BOUILLON (eds): *La traductique*, Montréal: Les Presses de l'Université de Montréal / AUPELF-UREF, 288-306.