# Utilidad de los corpus lingüísticos en el aula de español para extranjeros

AQUILINO SÁNCHEZ
Universidad de Murcia

## 1. ¿Qué es un corpus lingüístico?

Un corpus lingüístico es un conjunto de datos lingüísticos (pertenecientes al uso oral o escrito de la lengua, o a ambos), sistematizados según determinados criterios, suficientemente extensos en amplitud y variedad, de manera que sean representativos del uso lingüístico o de alguno de sus ámbitos y dispuestos de tal modo que puedan ser procesados mediante ordenador con el fin de obtener resultados varios y útiles tanto para el análisis como, en este caso, para el trabajo en el aula.

Los corpus lingüísticos tienen una historia muy corta. Sin embargo, los profesores de lengua están familiarizados con los **vocabularios fundamentales**, que se hicieron famosos en la época de la metodología audio-oral. Los profesores de inglés han utilizado con frecuencia los listados de voces más representativas de esta lengua, elaborados por Thorndike y Lorge en 1944 (*The Teacher's workbook of 30.000 words*) y por Michael West (*General Service List of English Words,* 1953). Ambas publicaciones han constituido punto de referencia obligada para muchos autores de manuales a lo largo de muchos años. Entre los profesores de francés, la norma del *Français Fondamental,* elaborada en la década de los sesenta, fue casi universalmente aceptada. En el caso del español también se han elaborado algunos vocabularios fundamentales o de frecuencia, pero sobre principios no siempre rigurosamente científicos. Entre ellos destacan los de García Hoz (1953), Márquez Villegas (1975) y recientemente el de Sánchez Lo-

bato y Aguirre (1992). Solamente los recopilados por Thorndike y Michael West se hicieron sobre un conjunto de textos equiparables a los corpus. Pero ha sido la aparición de los ordenadores la que ha acabado por dar un impulso decisivo a este tipo de estudios.

A finales del presente milenio, puede afirmarse ya sin lugar a duda que la *era de los corpus* no solamente se va consolidando, sino que empieza a ser tomada en serio con carácter general. El corpus, como conjunto de datos lingüísticos ordenados en múltiples modalidades, constituye ya un paradigma de la investigación y del trabajo lingüístico. Mas no solamente es útil para los lingüistas: es una necesidad para las empresas que se ocupan del procesamiento de los lenguajes naturales y constituye un utensilio de grandes posibilidades para otros colectivos, entre ellos el de los profesores de idiomas.

La disponibilidad de corpus del español útiles para la investigación o la docencia es aún muy escasa. Además del corpus ***Cumbre***, de propiedad privada y ya finalizado en su primera etapa (diez millones de palabras), que comenzará próximamente a comercializarse en alguna de sus partes y con distintas finalidades, puede accederse gratuitamente, mediante FTP (Universidad Autónoma de Madrid), al *Corpus del Español de la República Argentina* (dos millones de palabras de textos escritos de Argentina), al *Corpus Chileno de Referencia* (dos millones de palabras de textos escritos de Chile) y el *Corpus Oral de Referencia del Español Contemporáneo* (un millón de palabras). El *Informe sobre Recursos Lingüísticos para el Español*, elaborado por el Instituto Cervantes, menciona varios proyectos más relacionados con la recopilación de datos lingüísticos, pero la finalización o disponibilidad de casi todos ellos aún debe ser confirmada. En la actualidad, el proyecto más importante es el encomendado a la Real Academia de la Lengua Española (CREA), que aspira a recopilar cien millones de palabras del español actual.

## 2. El tratamiento informático del corpus

Un repertorio de textos se presenta como una secuencia de elementos lingüísticos, en la que las palabras se diferencian gracias al espacio en blanco que media entre una y otra. El corpus precisa, por lo tanto, de una adaptación o adecuación tanto en lo que se refiere al formato como a la disposición del contenido. La adaptación a que nos referimos aquí puede variar según la finalidad del corpus. En el caso que nos ocupa, deben tenerse en cuenta los siguientes aspectos:

1. Posibilidad de identificar los textos que constituyen el corpus: origen, modalidad (oral, escrito), ámbito de uso, tipo de lenguaje o tema.

2. Posibilidad de etiquetar las palabras del corpus, de manera que las búsquedas o ejercicios que interesen puedan contar, como mínimo, con información sobre las partes de la oración.
3. Posibilidad de elegir textos en tamaño y formato adecuados para trabajar con ellos en clase: los textos no deben ser ni demasiado extensos ni demasiado cortos.
4. Posibilidad de contar con un soporte físico adecuado (ordenador) y con suficiente capacidad de almacenamiento.
5. Acceso a un programa de gestión que permita llevar a cabo todas las funciones que requiera el trabajo en clase mediante textos.

Siempre que se trate de un corpus comercializado (como será el caso del sub-corpus de ***Cumbre***, de próxima publicación) estas tareas deben ser realizadas por la empresa responsable de su elaboración. Si se tratase de textos que el profesor añade a un corpus ya diseñado y estructurado, el corpus *maestro* debe incluir en su diseño la posibilidad de ampliarlo y, en consecuencia, permitir llevar a cabo esta tarea sin romper la armonía ya existente en los textos recopilados.

## 3. Listados de palabras

### 3.1. Consideraciones previas

Como ya anoté anteriormente, la utilidad de los corpus lingüísticos se puso de manifiesto en el ámbito de la enseñanza de idiomas, primeramente a través de los listados de palabras más usadas, que sirvieron a muchos autores de manuales para seleccionar el vocabulario de los textos introducidos en cada lección.

Los listados de frecuencia no eran desconocidos para los autores de materiales docentes, tal cual acreditan los nombres y estudios mencionados anteriormente. La historia de la enseñanza de idiomas cuenta con algunos ejemplos más sobre diversos intentos realizados para determinar y fijar el número de palabras «básicas» o más usadas, con la declarada finalidad de utilizar esos resultados en la enseñanza.

Para que tenga un grado razonable de fiabilidad, el listado debe reunir, al menos, dos requisitos, referidos a:

— La amplitud o extensión, que debe estar basada en muestras representativas, de acuerdo con los objetivos o finalidades propuestas.
— El tamaño, que ha de sustentarse sobre una cantidad de palabras tal, que garantice la validez estadística.

En el ámbito del español, el vocabulario fundamental de mayor relieve ha sido hasta ahora el de Vicente García Hoz (1953). Este autor hace referencia a tres tipos de niveles léxicos:

usual (12.402 palabras),
común (1.971 palabras) y
fundamental (208 palabras).

Los criterios para elaborar tales listados fueron, al menos en parte, «peculiares», aunque no carecían de aspectos valiosos. García Hoz fundamentó la elección de 400.000 palabras sobre textos tomados de cuatro «aspectos de la vida»:

1. la vida familiar (concretada en cartas privadas: 620 cartas),
2. la vida social indiferenciada (concretada en una variedad de periódicos),
3. la vida social regulada (concretada en documentos oficiales: políticos, religiosos y sindicales)
4. la vida cultural (concretada en textos tomados de los libros más vendidos en aquellos años, sean del género que fueren).

De cada uno de esos cuatro sectores tomó 100.000 palabras. La suma de los cuatro bloques dio origen a un total de 12.913 palabras diferentes, con una frecuencia media de 100 (las denominadas «usuales»).

Naturalmente, el método utilizado para seleccionar las voces condiciona sustancialmente los resultados. Si comparamos éstos con las palabras más frecuentes obtenidas de un sub-corpus de ***Cumbre*** (1995), con un total de cuatro millones de palabras (diez veces más extenso que el de García Hoz y diseñado sobre bases más representativas) referidas a la letra **A** (que equivale, aproximadamente, a un 10% del total del léxico), comprobamos que en la lista de García Hoz sólo aparece un 58,6% del total del vocabulario fundamental equivalente del mencionado sub-corpus. La desviación es, por tanto, notable. Muchas de las voces que contiene el listado de García Hoz y que no aparecen en el corpus, se justifican por el contexto histórico del momento y los valores propios de la sociedad en aquel entonces. Así se explica que el vocabulario fundamental de García Hoz contenga, entre las 2.000 voces más frecuentes, términos como:

***adorar, adornar, agrupación, alabar, altar, amado, apóstol, arca, artillería, ascender...,***

mientras que no figuran voces como:

***aborto, accidente, acceso, actor, adiós, aeropuerto, agrícola, alarma, amenaza, antena, asamblea, asesinato, asesino, atentado, autonomía, avión...***

Pero también puede apreciarse que entre las 2.000 palabras más frecuentes que aparecen en nuestro corpus, probablemente algunas no habrían estado ausentes del vocabulario de García Hoz si este autor hubiese realizado la recopilación con criterios más representativos. Se trata de palabras como:

> ***adentro, adiós, afuera, agradecer, aguantar, ajeno, alimentación, amable, amenaza, amplio, anual, aprobado, área, arreglar, aunque...***

El ejemplo ilustra, por tanto, la importancia del diseño de un corpus, ya que ello incide en los resultados. También se pone de manifiesto el grado de relatividad de cualquier listado de frecuencia, puesto que aquél siempre estará condicionado por el contexto que define cada periodo histórico.

Admitidos estos condicionantes, parece razonable concluir que, si una voz aparece muy a menudo en el corpus, su incidencia e importancia en el proceso comunicativo son muy elevadas y, por consiguiente, es más «rentable», en términos comunicativos, que un alumno extranjero aprenda esta palabra antes que otras de menor frecuencia y uso. En consonancia con esta norma, la solución más rápida y expeditiva es elaborar listados de frecuencias por tramos (las 800, 1.500, 2.000, ... palabras más frecuentes) y valerse de ellas para la redacción de textos, diálogos o ejercicios, según preconice la metodología de base. Así se hacía en los años sesenta con los métodos audio-orales. En este caso es patente la utilidad de los listados de frecuencia.

### 3.2. Aplicaciones prácticas de los listados de palabras

No es habitual que el profesor de idiomas lleve a clase listas de palabras, sin más, para trabajar con ellas o memorizarlas. Pero aunque los listados de palabras en cuanto tales no sean recomendables como instrumento didáctico, sí tienen otras funciones positivas.

En primer lugar, los listados de palabras que pueden obtenerse de un repertorio de textos no se limitan al bien conocido *listado alfabético* de voces. Pueden aportar otro tipo de información útil,

**a. Sobre la frecuencia de cada palabra en los textos analizados**, o en el texto en el que se trabaja. La frecuencia léxica tampoco es de interés por sí misma, pero el profesor sí que puede aprovecharse de esa información para tomar determinadas decisiones, como sería, por ejemplo, insistir más o menos en el aprendizaje de una voz según el grado de importancia que ésta tenga en el texto (hecho que deriva de su índice de frecuencia).

Este hecho incide directamente en la eficacia docente y discente. Y no debemos olvidar que el aprendizaje de una lengua busca precisamente la rentabilidad comunicativa y ésta depende de la eficacia en el aprendizaje.

**b. Sobre aspectos diversos de las palabras o léxico** con que se trabaja, como pueden ser:

**b.1. El listado de voces con una determinada letra o secuencia intermedia, prefijo, sufijo o similar**. En muchos momentos a lo largo de la docencia se hace necesario disponer de palabras que empiecen por **retro-, hiper-,** ... o que acaben en **-dad** ... o que contengan **-rr-, -qu-, -sp-**, etcétera. En estos casos el corpus es capaz de suministrarnos tal información en cuestión de segundos, tal cual sugieren los siguientes ejemplos:

**-ZC-**
aborrezca
aborrezco
amanezca
aparezcan
convalezco

**-RR-**
arranque
arranqué
arranques
cerrarse
cerrazón
cerré
cerril
cerro
cerró
cerrojo
cierra
cierran

**-QU**
alambique
alfeñique
almanaque
aquí
Arequipa
arlequín
arlequines
arquean
arqueo
arqueología
arqueológica

**-SP-**
asperezas
asperjada
asperjados
áspero
aspiraciones
aspiramos
aspiran

**-AD**
prioridad
probabilidad
productividad
profesionalidad
profundidad
progresividad
propiedad
proporcionalidad
prosperidad
proximidad

**HIPER-**
hiperactividad
hiperinflación
hipertensos
hipertrofia

**RETRO-**
retroceder
retrocedía
retrocedido
retrocediendo

**b.2. De no menor utilidad es poder disponer de un listado de palabras pertenecientes a la parte de la oración deseada**: nombres, adjetivos, etcétera. Tal listado es viable solamente si el corpus está etiquetado a nivel básico (como se verá más adelante).

**b.3. El listado útil puede cifrarse en la obtención de palabras de una determinada extensión**: de 10, 9, 5, ... letras, al estilo de los ejemplos siguientes:

| *Palabras de 12 letras* | *Palabras de 3 letras* |
|---|---|
| alternativas | con |
| complementen | del |
| comprensible | las |
| comunicación | les |
| constitución | los |
| consumidores | por |
| continuación | que |
| culpabilidad | son |
| demográficas | ten |
| especulación | una |

### b.4. Listados inversos

La utilidad de los listados puede residir también en proporcionar información que se base no en la ordenación por el principio de cada voz, sino por las letras finales:

| | |
|---|---|
| a | biblioteca |
| acaba | dedica |
| acercaba | indica |
| señalaba | Bélgica |
| hablaba | tónica |
| brillaba | Guernica |
| pasaba | música |
| pensaba | informática |
| necesitaba | practica |
| calentaba | práctica |
| estaba | blanca |
| Córdoba | bizca |

Conviene advertir que no se trata de juegos más o menos interesantes o divertidos, sino de instrumentos que pueden resultar útiles en un determinado momento, atendiendo a las necesidades y finalidades de cada clase en cada momento. Quien acostumbra a valerse de juegos y rompecabezas para el aprendizaje del léxico encontrará en este tipo de utensilios informáticos una ayuda tan valiosa como imprescindible. Sin perder de vista que el factor motivación también puede resultar beneficiado por actividades de este tipo.

### b.5. Listados de sintagmas o palabras que concuerden en género o número.

Este tipo de listados amplía considerablemente el ámbito de utilidad de los corpus, ya que permite al profesor disponer de mate-

riales sobre los cuales ilustrar determinados comportamientos sintácticos del español, así como determinadas reglas gramaticales o usos fundamentales que afectan a la gramática. Estos listados pueden obtenerse o bien como listados simples o bien como listados con su contexto (*concordancias*). Piénsese en la utilidad de elementos como los siguientes:

## Secuencia de palabras con terminación -as -as ...

**algunas semillas** 1

He111 He FIC; 6569 sólo apretarlas soltaban la cáscara. Yo traje aquí **algunas semillas**, pocas, apenas una bolsita... después pen

**algunas sorpresas** 1

He031 MEX PRN; 1957 ha construido un relato en el que hay apenas **algunas sorpresas**. Al alcance de un cuerpo perfecto Hoy c

**algunas tiendas** 1

HE31 BOL PRN; 614 participación de parlamentarios que responden a **algunas tiendas** políticas y no se tomó en cuenta a la

**armas negras** 1

He12 HIS POE; 1329 con dientes y cuchillos, con diccionarios y otras **armas negras** con citas respetables, se lanzaron a disputar

**armas químicas** 1

HE31 BOL PRN; 4610 El interés del espionaje israelí en el experto en **armas químicas** chileno obedece a que desean saber la

**arquitecturas novelísticas** 1

HE006 MEX ANT;A 3687 en la obra de Carlos Fuentes con las grandes **arquitecturas novelísticas**. Las cinco piezas que reúne rep

## Secuencia de palabras con terminación -iz -a...

**actriz australiana** 1

He031 CHI PRN; 3956 a enviudado hace dos años de su última esposa, la **actriz australiana** Coral Browne. Tuvo dos hijos: Vincent d

**actriz italiana** 1

He031 CHI PRN; 2466 ile. De refinamiento, nada. Basta imaginarse a la **actriz italiana** con esos escotes y vestidos rasgados, pala

**matriz convulsa** 1

HE006 MEX ANT;A 3567 trilingüe, lépera y barroca aparece como una gran **matriz convulsa** dentro de la cual se gesta el monstruo de l

**perdiz olorosa** 1

He12 HIS POE; 1102 austral, nave de plata y enlutado terciopelo, la **perdiz olorosa** y el relámpago de los fosforescentes picafl

**raíz ésa** 1

He12 HIS POE; 2920 ha dado a luz, la madre del cordero, la causa, la **raíz ésa** no habrá ahora para mí? ¡Siquiera aquella otra que

## Secuencia de palabras con terminación -iz -o...

**infeliz Ireneo** 1

HE013 HE CUE;C 274 igable como la que día y noche convergía sobre el **infeliz Ireneo**, en su pobre arrabal sudamericano. Le era m

## Secuencia de palabras con terminación -os -os...

**bajos fondos** 3

HE031 VEN PRN;P 8182 iudad. Felipe Antonio Montilla es conocido en los **bajos fondos** con el remoquete de El Pingüino y estaba sien

**Bajos instintos** 1

He031 MEX PRN; 1875 r Sidney Lumet dirige un guión del mismo autor de **Bajos instintos**. El engaño, The Guardian, La mano que mece

**baños colectivos** 1

He113 He NBI;N 9561 ampaña, cruzarían el estadio para ducharse en los **baños colectivos**, volverían apresurados a ponerse el unifo

**barcos sardineros** 1

He031 MEX PRN; 4252 a alerta continua sobre las operaciones de los 30 **barcos sardineros** y de sus tripulaciones, que enfrentan pr

## Secuencia de palabras con terminación -e -e...

**corriente favorable** 1

He031 MEX PRN; 4123 ristianismo. El Copei se encuentra dividido en la **corriente favorable** a Caldera y la del ex secretario gener

**creciente porcentaje** 1

HE031 URU PRN;P 1597 scar mano de obra más joven y más maleable, con un **creciente porcentaje** de mujeres y miembros de minorías étn

**creyente practicante** 1

He031 CHI PRN; 3701 Piñera. —Cómo es su relación con Dios? —Soy muy **creyente, practicante**, de misa dominical y converso con Di

## Secuencia de palabras con terminación -a -a...

**Dama Loca** 2

HE006 MEX ANT;A 2723 Pedro el hombre libre, fray Julián el pintor, la **Dama Loca**, el cronista y la Reina Isabel son otros tantos

HE006 MEX ANT;A 2622o se hundirá inexorablemente en su melancolía, la **Dama Loca** lo estará más cada día. Imaginación cronófaga,

**damisela suburbana** 2

HE031 VEN PRN;P 7717 miserias y grandezas que le hacen una subyugante **damisela suburbana**. Tiene 464 años y su nombre, Maracaibo,

HE031 VEN PRN;P 1927 miserias y grandezas que le hacen una subyugante **damisela suburbana**. Tiene 464 años y su nombre, Maracaibo,

**danza aérea** 1

HE006 MEX ANT;A 1265 or un cuerpo de doctrina y por una geografía, una **danza aérea** y acuática alrededor de la atracción universal

## Secuencia de palabras con terminación -o -o...

**maestro consejero** 1

HE031 PER PRN;P 9055 habilidad para la escritura y podrían ser un buen **maestro consejero**. Sus poderes de observación son agudos y

**maestro enfilado** 1

He014 He RCO;R 199 io donde ya nadie se decidiría a penetrar: el caño **maestro enfilado** en dirección al río, la reunión tormentos

**mandado completo** 1

He111 He FIC; 5140 chas. —Sí, abuela. —Y de paso, para que hagas el **mandado completo**, dile que nos empreste un cernidor y una

**museo imaginario** 1

HE006 MEX ANT;A 216 hace de este objeto, tan parecido a un libro, un **museo imaginario**, un espacio al que sólo le faltarían las

**musgo tierno** 1

He111 He FIC; 3286 pio de los vegetarianos La piel se le cubrió de un **musgo tierno**, semejante al que prosperaba en el chaleco an

**c. Sobre las palabras que suelen acompañar a determinadas voces** o que pueden aparecer en determinadas situaciones o contextos comunicativos. Los libros de texto ya suelen tener en cuenta estos hechos, aunque ello se haga de manera más intuitiva que experimental. Un corpus permite que el profesor compruebe por sí mismo qué palabras, por ejemplo, suelen acompañar con un mayor índice de frecuencia a la palabra ***casa***. Este sería un ejemplo real, basado en un corpus de un millón de palabras, pero que podría hacerse sin ninguna dificultad en cualquier otra muestra:

| | | |
|---|---|---|
| mi | decir | problemas |
| bueno | familia | amas |
| hacer | hijo | marido |
| cosas | padre | hija |
| trabajo | mujer | gente |
| fuera | hermano | persona |
| mañana | vez | papá |
| momento | mamá | madre |
| días | teléfono | vivir |
| años | acá | cosa |
| tiempo | puerta | hijos |
| tener | señora | chicos |

Con esta información el profesor tiene la posibilidad de orientar el aprendizaje del léxico en la dirección que más interese, tomando siempre como guía la rentabilidad comunicativa de lo que aprende el alumno.

## 4. Las concordancias

Otro de los grandes pilares sobre los que se ha justificado la utilidad de un corpus es el de las concordancias léxicas. Por *concordancia* se entiende el conjunto de ocurrencias de una forma, con referencia explícita al contexto en que dicha forma aparece. El contexto aporta la información necesaria para descubrir cómo «actúa» y cómo se «comporta» cada palabra cuando es utilizada por los hablantes nativos de una lengua. La gran ventaja de las concordancias estriba en que cada elemento lingüístico se presenta arropado en su entorno tantas veces como se encuentre en un texto, de manera unificada, sin que haya necesidad de recurrir a la lectura secuencial de ese texto para recabar la información contextual. Su presentación suele revestir la siguiente forma:

económica 21

Ho111 VAF-viii Eco 126 mercado es, hoy por hoy, la única concepción **económica** en el mundo no sólo en Venezuela ni en América
Ho111 VAF-viii Eco 127 rica Latina, en el mundo la única concepción **económica** que está impulsando las economías en el mund
Ho111 VAF-viii Eco 130 país comunista que ha emprendido una reforma **económica** a fondo y que está dentro de una estructura
Ho111 VAF-viii Eco 227 la gestión gubernamental que con la posición **económica** propiamente dicha. Entonces, vamos a retomar
Ho111 VAF-viii Eco 411 conomía de mercado es simplemente la sanidad **económica**, por qué?, porque economía de mercado signif
Ho111 VAF-viii Eco 523 gran problema que confrontaría una política **económica** centrada en la revaluación del bolívar. Una
HO112 VABF-viii Eco 533 ados del Caribe para fomentar la integración **económica** de la región. Esta asociación tendría como o
HO112 VABF-viii Eco 534 dría como objetivo avanzar en la integración **económica** y la cooperación entre los países de la cuen
HO112 VABF-viii Eco 550 re comercio de América del Norte. En la parte **económica** en esta reunión del grupo de los tres CONCRIC
Ho122 VAF-viii Eco 343 reocupados con los subsidios de la Comunidad **Económica** Europea y particularmente los quesos holande
Ho122 VAF-viii Eco 641 española para otros capitales de la Comunidad **Económica** Europea. O sea que esto pudiera inclusive pe
Ho122 VAF-viii Eco 932 Venezuela tenemos una economía y una cultura **económica** que por años ha estado acostumbrado por un M
Ho122 VAF-viii Eco 1199 Caldera concibe así la economía, la política **económica** y Fernández tampoco. Aquí no se ha planteado
Ho122 VAF-viii Eco 1228 más inteligentes como economía, la política **económica**. Hoy se manejan en los países de forma mucho

Ho122 VAF-viii Eco 1363 ue por favor los responsables de la política **económica** por la vía fiscal, por la vía comercial, por
Ho122 VAF-viii Eco 1426 xxx el gobierno ha fracasado en su política **económica** y el mismo tiene que rectificar. Ahora, sí ha
Ho123 VABF-viii Eco 251 is. La crisis está originada por la política **económica** que ha llevado este gobierno y se ve pues el

Como ayuda complementaria, es posible ordenar las ocurrencias de cada forma, no sólo por su orden de aparición, sino atendiendo a la primera letra de la palabra que sigue inmediatamente a la forma en cuestión; u ordenar las ocurrencias según el último signo de la palabra que precede a la forma alineada (orden alfabético reverso).

**económica** 21

Ho111 VAF-viii Eco 130 país comunista que ha emprendido una reforma **económica** a fondo y que está dentro de una estructura
Ho111 VAF-viii Eco 523 gran problema que confrontaría una política **económica** centrada en la revaluación del bolívar. Una
Ho123 VABF-viii Eco 266 isis. La crisis está generada por la política **económica** de este gobierno. - El paro, por ejemplo, d
HO112 VABF-viii Eco 533 ados del Caribe para fomentar la integración **económica** de la región. Esta asociación tendría como o
Ho111 VAF-viii Eco 126 mercado es, hoy por hoy, la única concepción **económica** en el mundo no sólo en Venezuela ni en mérica
HO112 VABF-viii Eco 550 re comercio de América del Norte. En la parte **económica** en esta reunión del grupo de los tres CONCRIC
Ho123 VABF-viii Eco 1049 nsideran que su salario, en la circunstancia **económica** en que estamos viviendo de una inflación gal
Ho122 VAF-viii Eco 641 española para otros capitales de la Comunidad **Económica** Europea. O sea que esto pudiera inclusive pe
Ho122 VAF-viii Eco 343 reocupados con los subsidios de la Comunidad **Económica** Europea y particularmente los quesos holande
Ho122 VAF-viii Eco 1228 más inteligentes como economía, la política **económica**. Hoy se manejan en los países de forma mucho
Ho123 VABF-viii Eco 482 es el pueblo, no es la gente, es la política **económica** neoliberal. Enfrentémosla juntos. Vamos a en
Ho122 VAF-viii Eco 1363 ue por favor los responsables de la política **económica** por la vía fiscal, por la vía comercial, por
Ho111 VAF-viii Eco 411 conomía de mercado es simplemente la sanidad **económica**, por qué?, porque economía de mercado signif
Ho111 VAF-viii Eco 227 la gestión gubernamental que con la posición **económica** propiamente dicha. Entonces, vamos a retomar
Ho111 VAF-viii Eco 127 rica Latina, en el mundo la única concepción **económica** que está impulsando las economías en el mund

Una rápida ojeada a estos datos revela el alto índice de preposiciones *(a, de, en, por)*, de la conjunción copulativa *y* y del nexo *que* que acompañan a la forma *económica*, entre otras posibles observaciones.

Es importante resaltar la claridad y sencillez con que esta organización permite reconocer los compuestos o locuciones más frecuentes de una palabra. En este caso observamos el alto índice de ocurrencias de la secuencia *política económica* (nueve veces), seguida de otras como: *Comunidad Económica Europea, integración económica* y *concepción económica*, entre otras.

Todas las ejemplificaciones presentadas hasta ahora son exclusivas de formas o palabras aisladas, es decir, fruto de la selección y búsqueda de una única secuencia específica de letras. Sin embargo, también es posible obtener resultados y datos sobre determinados grupos de palabras: sintagmas, locuciones, perífrasis verbales, giros, correlaciones, etc. He aquí un ejemplo:

**tener que** 4

Eo113 EABF-viii Eco 61 ha crecido, por fin, la condición de que van a **tener que** vender alguno de sus compañeros de viaje para
Eo113 EABF-viii Eco 451 s no lo ha dicho así, dice, poco menos vamos a **tener que** reunirnos un día, los que estamos en comisión
Eo113 EABF-viii Eco 454 dos a comisiones e investigaciones, un día van a **tener que** acordar, celebrar un pleno del Congreso para
Eo124 EAF-vii Rel 317 o voces, por aquí, a ver. (C) Bueno, yo lamento **tener que** quebrar este diálogo entre pastores y ovejas,

Los resultados se pueden extender no solamente a palabras consecutivas, sino también a partículas correlativas e interdependientes separadas por otras palabras. A menudo la distancia que aleja a dos elementos lin-

güísticos no es predecible y puede variar en cada caso (por ejemplo, *por un lado ... por otro lado, tanto ... como*). El ejemplo que sigue corresponde a la correlación *no ... ni*, para los casos en que la separación máxima entre ambas partículas no excede de cuatro palabras.

no ni 8

Eo113 EABF-viiiEco 101 les servidores del amo y no de quien los elige, no tendremos aquí, ni democracia, ni justicia, ni una eco
Eo113 EABF-viiiEco 578 ez años, no sólo no tenemos pensiones sino, que no tendremos, ni siquiera, sueldo. (D) Fernando, no seas
Eo113 EABF-viiiEco 958 yentes que lo han hecho posible porque, si esto no hubiera funcionado, ni antena ni nada. O sea, hay tres
Eo124 EAF-vii Rel 215 los sitios de apariciones, que en aquel momento no había videntes, ni había, digamos, apariciones, sino q
Eo124 EAF-vii Rel 319 o un diálogo muy interesante, sobre todo, cuando no has sido ni pastor, ni oveja. Porque, uno ve el mundo
Eo124 EAF-vii Rel 628 ro, que el espectador no quede desorientado. Eso no es, ni como ha dicho el señor Sánchez Dragó, ni como h
Eo124 EAF-vii Rel 1004 talmente inválida y unos dolores espantosos que no se pueden ni contar, ¿no? Al momento yo xxx las aparic
Eo124 EAF-vii Rel 1009 aba curada, y hasta hoy. Yo llevaba plantillas, no podía andar ni un paso, incluso me querían poner en un

## 5. Otras implicaciones del corpus en la enseñanza de E/LE

Anteriormente se hizo referencia a dos de los capítulos en que se ha centrado la utilidad de los corpus hasta el momento: los listados de palabras y las palabras en contexto. En ambos casos se ilustró el tema con ejemplos.

**En lo relativo a los listados**, es conveniente añadir a lo dicho anteriormente que la prueba más fehaciente de la adecuación entre la realidad del uso y lo que ofrece un libro de texto reside en una comparación que establezca los niveles de correlación entre las palabras más frecuentes de un corpus y las que ofrece un manual en concreto, tomando en consideración las formas no flexionadas y las flexionadas. La diferenciación entre formas flexionadas y no flexionadas no es inocua, ya que si una forma concreta aparece cien veces en un método, es más probable que se aprenda antes o mejor que otra que sólo aparece una o dos veces. Ejemplificando el tema en el verbo *tener*, se llega a las siguientes conclusiones:

La frecuencia de *tener*, incluyendo todas las formas flexionadas, en un corpus oral de más de un millón de palabras, es del 0,96 %. En relación con ese total de apariciones, *tiene* ostenta el récord de frecuencia con un 30 %, *tengo* sigue muy de lejos con un 8,5 %. Todas las formas de indicativo agrupadas ocupan el 66 % de frecuencia.

Esa misma frecuencia, en ***Cumbre 1***, manual de español para extranjeros que aquí tomamos como término de comparación, es del 0,77 % para el total de *tener*. En relación con el total de este verbo, *tiene* ocupa también un 30 %, *tengo* un 14 % y todas las formas flexionadas del presente de indicativo equivalen a un 66 %. Pueden apreciarse los datos con más claridad en el siguiente cuadro:

## Verbo *tener* y sus formas flexionadas en el corpus *Cumbre* y en *Cumbre 1*:

| **Sub-corpus de *Cumbre*** | | ***Cumbre-1*** (método de español para extranjeros) | |
|---|---|---|---|
| tendrá | 95 | tendrá | 4 |
| tendrán | 32 | tendrán | 1 |
| tendrás | 9 | tendrás | 6 |
| tendré | 7 | tendré | 2 |
| tendréis | 1 | tendremos | 2 |
| tendremos | 50 | tenéis | 2 |
| tendría | 118 | tenemos | 13 |
| tendríais | 2 | TENER | 13 |
| tendríamos | 47 | tengo | 24 |
| tendrían | 28 | tenía | 11 |
| tendrías | 2 | tenían | 2 |
| tenéis | 32 | tenido | 6 |
| tenemos | 1377 | tiene | 50 |
| tener | 1009 | tienen | 8 |
| tenga | 260 | tienes | 10 |
| tengáis | 1 | tuve | 1 |
| tengamos | 66 | tuvieron | 2 |
| tengan | 140 | tuvimos | 1 |
| tengas | 24 | tuviste | 1 |
| tengo | 900 | tuvisteis | 1 |
| tenía | 522 | tuvo | 2 |
| teníamos | 95 | | |
| tenían | 141 | | |
| tenías | 27 | (TOTAL de palabras del ***Cumbre-1***: 20.976) | |
| tenido | 420 | | |
| teniendo | 99 | | |
| tiene | 3197 | | |
| tienen | 1119 | | |
| tienes | 397 | | |
| tuve | 83 | | |
| tuviera | 39 | | |
| tuvierais | 1 | | |
| tuviéramos | 12 | | |
| tuvieran | 16 | | |
| tuvieras | 5 | | |
| tuvieron | 51 | | |
| tuvimos | 67 | | |
| tuviste | 2 | | |
| tuvo | 124 | | |
| (TOTAL de palabras del sub-corpus: 1.102.643) | | | |
| **PORCENTAJES DE FRECUENCIA de *tener*:** | | **PORCENTAJES DE FRECUENCIA de *tener*:** | |
| **Sobre total: 0,96 %** | | **Sobre total: 0,77 %** | |
| **tiene: 30 %** | | **tiene: 30 %** | |
| **tengo: 8,5 %** | | **tengo: 14 %** | |
| **Formas de presente de indicativo: 66 %** | | **Formas de presente de indicativo: 66 %** | |

La correspondencia no es exacta en todos los casos, pero sí razonablemente aproximada y en tal sentido puede afirmarse que este manual se acerca a la realidad comunicativa en lo que se refiere a la presencia del verbo *tener* y sus formas.

Los resultados anteriores no deberían inducir a conclusiones sesgadas. El análisis del tema obliga a concluir que, si bien las distancias deberían ajustarse a ciertos límites, no es posible reducirlas a cero, eliminando todas las diferencias léxicas entre un corpus representativo de la totalidad de un idioma y un manual para la enseñanza de la lengua a extranjeros. Las razones son de dos tipos: de orden metodológico y de orden contextual (limitaciones derivadas del contexto).

La metodología sobre la que se basa un manual condiciona de antemano el contenido que debe seleccionarse para la docencia, al menos en parte. Una metodología «tradicional» da prioridad a los aspectos gramaticales, y el léxico, por ejemplo, no tiene por qué sujetarse a contextos o situaciones. En cambio una metodología comunicativa exige la selección de situaciones comunicativas y contextos que conllevan el uso de vocabulario relacionado con tales parámetros. Si no existiese limitación en el número de situaciones seleccionables y en la amplitud con que cada una de ellas debe tratarse en el nivel elemental, las coincidencias con un corpus como ***Cumbre*** podrían aproximarse al 100 %. Pero dado que los libros de texto deben limitar el número de situaciones y la extensión de su tratamiento por razones externas al proceso comunicativo en cuanto tal, es imposible alcanzar el grado de exhaustividad que caracteriza a un corpus. Reflexiones que no deben utilizarse como disculpa, sino como justificantes de una realidad que se impone por necesidad y que deriva de las premisas específicas sobre las cuales se asientan tanto un corpus como un manual para la enseñanza a extranjeros, cada uno por su lado.

El contexto impone también sus exigencias. Tomar un corpus como punto de referencia para los manuales no ha de entenderse como si se tratase de un corsé impuesto desde fuera y ajeno a la dinámica derivada de la docencia, del qué y cómo se enseña. El corpus debe ser un punto de referencia básico (e incluso ideal), pero no es el único eje en torno al cual se elabora un manual o materiales para la clase. En efecto, el corpus, de entrada, no puede recoger **toda la lengua** usada. En relación con las situaciones de la vida diaria, es también imposible encontrar en él todas las concreciones que pueden darse en la vida real y en el ámbito de las relaciones sociales (por ejemplo, «para pedir un favor»). Esta limitación repercute en el vocabulario recopilado: si la situación concreta de «sacar un billete de avión» no figura en el corpus, el término *avión* tendrá un menor índice de frecuencia. Circunstancias repetidas o similares de esta índole pueden hacer que algunas palabras no aparezcan (caso extremo) o no alcancen la frecuencia necesaria para figurar en los tramos

de frecuencias altas (en las cuales aparecen las 1.000, 2.000, etcétera palabras más frecuentes).

El caso puede darse también en sentido contrario: si por circunstancias inevitables, derivadas de la misma selección, un corpus contiene varios textos o conversaciones en los que se habla del «aborto» o de la «planificación familiar» (porque esos son temas candentes en un determinado momento y, por ende, tema frecuente de comunicación), es probable que las voces relacionadas con este tema (*aborto, abortar*, etc.) aumenten considerablemente su índice de frecuencia y, como consecuencia de ello, aparezcan en tramos de frecuencia alta. Es evidente que si el tema del aborto no se toca en un manual para extranjeros, la palabra *aborto* no figurará en el listado de voces del libro. ¿Se da entonces alguna deficiencia en el manual? Afirmar que sí equivaldría a situarnos fuera de una visión equilibrada y realista. Aparte del hecho de que términos de esta índole pueden ser considerados «circunstanciales», desde el punto de vista de su incidencia en la comunicación: su frecuencia en un corpus realizado diez años más tarde, si esta temática ha decrecido en importancia, sería, con toda seguridad, más baja.

Se impone, por tanto, admitir que los listados de frecuencia derivados de un corpus son puntos de referencia fiables, eficaces y útiles, pero no fuentes **exclusivas** y «**dogmáticas**». Los vocabularios básicos deben utilizarse por manualistas y profesores con gran flexibilidad y teniendo en cuenta la globalidad del uso y las restricciones y facilidades que derivan de la selección de un contexto determinado. Probablemente lo más útil que puede derivarse del listado de frecuencias de un corpus es que sirva de guía para deducir las contextualizaciones y situaciones que exijan tal vocabulario.

## 6. El corpus y la enseñanza de la gramática

En lo que respecta a la **gramática** (morfología y sintaxis principalmente), la ayuda del corpus se centrará en dos ámbitos:

- Ofrecer información respecto a la conveniencia o no de incluir sintagmas o estructuras determinadas que ofrezcan un nivel de uso suficiente para justificar su enseñanza en los diferentes niveles de que suele constar un manual (1.º, 2.º, 3.º). Los índices de frecuencia de tales estructuras constituyen la referencia obligada.
- Ofrecer los contextos en que tales estructuras gramaticales suelen darse en el uso lingüístico. He aquí algunos ejemplos, centrados en sintagmas o formas usuales, ***como esta mañana, este año, este mes, les gusta, me gusta, me gustaría, nos gustaría***:

**esta mañana** 66

EO111 E-v Ci Con uno de ellos, Luis Miguel Gil, dialogábamos **esta mañana** en la SER. D Las únicas recomendaciones que hemo
EO111 E-i Me A Emitimos una noticia que leemos **esta mañana** del diario El mundo. Azcoitia, el histórico milit
EO111 E-i Me **Esta mañana**, nuestro invitado se llama Segundo Andrés, Javier
EO111 E-i Me A Además de lo dicho como acabamos de leer **esta mañana** en El país, van a construir plantas lecheras en
EO111 E-i Me Visto así, tú que vas a trabajar **esta mañana** por un salario, ¿qué harías? Pues yo igual. Por
EO111 E-i Me **Esta mañana** los mandos de un Xxx en Sarajevo comprobarán la
EO111 E-i Me Fernando Isbar, cómo lo ve, qué nos propone **esta mañana** para la reflexión, una consideración. E Vamos,
EO111 E-i Me ntario o en pregunta a algunos de nuestros invitados **esta mañana** desde Sarajevo. F Hacer una gran pregunta que tod
EO111 E-i Me Durante **esta mañana**, eso sí, los aviones continúan patrullando por la

**este año** 66

HO121 MAF Ot antada y podría concretarse para la segunda mitad de **este año**. Sobre mediodía, el presidente Salinas fue el invita
HO121 MAF Ot cción Nacional dará la sorpresa en las elecciones de **este año** ya que obtendrá la victoria. Así lo afirmó el candid
HO121 MAF Po A xxx nación al panorama político de nuestro país en **este año** de mil novecientos noventa y cuatro en que los mejic
HO121 MAF Po cusiones y comentarios sobre la política nacional en **este año** de gran cambio político. música B El veintiuno de
HO121 MAF Po dor de cómo se desarrollarían los comicios federales **este año**, pero cómo se desarrollarían en términos de la orien
HO121 MAF Po peradores, agentes de viaje, líneas aéreas europeas, **este año** va a tener lugar en Querétaro, precisamente para
HO122 MBF Cu ncia de ciento veinticinco años, exactamente durante **este año**, desde el dos de diciembre de noventa y dos hasta
HO122 MBF Po s Unidos es indiscutiblemente el centro de su agenda **este año** pero también está el combatir el crimen, reformar
HO124 MC- Cu Y ahora, dice, ¿ganaré el óscar **este año** o se lo darán a otro? Qué tan grave es haberle evita
HO124 MC- Cu su esquema de libre comercio el primero de enero de **este año**. En la zona fronteriza del país el problema fue más
etc.

**este mes** 9

EO111 E-i Cu rsión de esta ópera, que se va a ver los días, y de **este mes**. Plaza y Martín han hecho una escenografía que tiene
HO005 MJ Te de acuerdo, Félix, para ver, digo en esta semana, en **este mes**, para ver qué día concretamos la hora, ¿sí, Félix?
HO111 MAF Ot los grandes temas del foro que se celebrará el y de **este mes** de enero, en la Universidad Iberoamericana, convocad
HO114 MC- Po tarle una audiencia, la cual, nos fue prometida para **este mes** de enero, sin que hasta ahora tengamos una respuesta
HO124 MC- Cu la colonia Roma, estarán abiertas al público durante **este mes**, hay para todos los públicos. Todo tiene un principi
HO112 V-i Ci , definitivamente, el proyecto va antes de finalizar **este mes** a consejo de ministros de manera que sea aprobado
Ho123 V-v Ec ordar los conflictos que se presentaron durante todo **este mes** de abril y que dieron pie a este programa especial
etc.

¿Qué profesor dudaría de la utilidad de esta información en el aula para decidir, por ejemplo, si determinadas secuencias deben mantener la prioridad de que gozan o no en la praxis docente? ¿O para extraer ejemplos de contextos de uso? De igual utilidad podrían ser los ejemplos referidos a «ser» y «estar» (eterno problema en la enseñanza del español a extranjeros), o a las formas verbales del pretérito indefinido frente al imperfecto de indicativo o frente al pretérito perfecto (otro de los temas siempre presentes en los libros de texto para extranjeros) y extraer de ellos las consecuencias adecuadas.

El uso de los pronombres personales precediendo al verbo, frente a su ausencia y en qué contextos, es objeto de prácticas frecuentes y reglas más o menos afortunadas. Ambos extremos pueden sistematizarse para la clase sobre los materiales que aporta el corpus:

**les gusta** 12

EO111 E-i Me la carpa es más bien un pez que casi deportivo, que **les gusta** a los pescadores para ejercer el deporte de la pesc
EO111 E-i Me rreno de su complicidad, para unos la complicidad que **les gusta** es la de la denuncia sistemática, que no aporta nad
EO111 E-i Me B Y luego **les gusta**. A Es que a los niños, los niños no son tontos, los
EO112 E-i So s autonómicos no quieren comparecer en el senado, no **les gusta**?. B No, no creo que sea así. En realidad es que has
EO113 E-i So ellos, para asociarse y decir que bueno, realmente no **les gusta** el mundo en el que viven, que realmente piensan que
EO114 E-x Ot C ¡Qué bueno! A **Les gusta** la sociedad y eso. En un sitio limpio, como es debi
EO114 E-x Ot iciendo, bueno, que el régimen, la forma de estado no **les gusta**... A En definitiva, señora. B En definitiva, en d

**les gustan** 2

EO114 E-i Cu componentes del gabinete, porque no crean que a ellos **les gustan** todos los temas que proponemos desde la redacción
E0114 E-x Po a dar su opinión de porqué les votan y qué es lo que **les gustan** de ellos, porque a mí me interesaría saberlos. Pero

**les gustaría** 2

EO113 E-i So que la mujer esté ganando terreno sobre ellos y que **les gustaría** vivir en un mundo en el que hubiera más acuerdos
EO114 E-i Cu l, usted de qué tema de los que aún no hemos hablado **les gustaría** hablar. E Yo pienso lo mismo que Jorge, que este

**me gusta** 83

EO111 E-i Cu na orquesta, pero que ha oído muchos discos dice, no **me gusta**, yo tengo la grabación de Xxx y no sé qué, que es un
EO111 E-i Me no sé muy bien, tampoco, qué quiere decir eso, pero **me gusta**. C Bueno, cada proyecto es personalizado, cada proye
EO111 E-i Me ue parecía, vamos, estoy viendo que no, por eso a mí **me gusta** mucho y creo que sí, que sería un buen alcalde, adem
EO111 E-i Me B Bueno, yo lo lamento pero no estoy en absoluto, no **me gusta** el señor Matanzos. No me gusta nada y quería pues, m
EO111 E-i Me No **me gusta** nada y quería pues, muy bien, que se presentara como
etc.

**me gustan** 10

EO111 E-i Cu A mí **me gustan** los argumentos, yo creo que la historia está llena
EO111 E-i Me dieciséis años, ¿eh?, soy una chica joven pero a mí **me gustan** esos celos, esos pequeños celos que demuestran bast
EO112 E-i So fans que le siguen año tras año, pero ya te digo, no **me gustan** esos alardes de voz. Primero, ese romperse la garga
etc.

**me gustaría** 119

EO111 E-v Ci A Hay una última pregunta que **me gustaría** hacerte, Olga, a tu consulta, a la consulta de lo
EO111 E-i Cu **Me gustaría** saber, dice que tú has sido amigo de Xxx , ¿es
EO111 E-i Ed A Sí, **me gustaría** hacer algunos comentarios más referentes a ese ti
EO111 E-i Me l, ya lo decía hace un momento el señor Xxx, pero sí **me gustaría** invitar al señor Coma, Fernando Xxx, Antonio Garc
etc.

**me gustó** 8

EO111 E-i Me Yo no sé nada de ajedrez y **me gustó** la película y está muy bien realizada. Tan bien real
EO121 E-x Po india, de una manera que, bueno, a mí desde luego no **me gustó** mucho, ¿no?, y es el que dijo no, no, no puede haber
EO122 E-i So H Bueno, yo a Sofía la conocí y **me gustó** muchísimo y, bueno, pues, cosas que se hacen en la v
EO122 E-i So ás o menos, como dos semanas y me lo pensé muy bien, **me gustó** las sinceridad de él sobre todo, porque xxx que siga

**nos gusta** 10

EO111 E-v Ec pañeros de viaje en esta aventura, que a lo mejor no **nos gusta** que se sepan. A Perdona Luisa, en ese cena que acud
EO113 E-x Po Es decir, no **nos gusta** que, que los partidos políticos se peguen ... B Pe
EO113 E-v Re hecho Anastasio fijándose en los aspectos positivos, **nos gusta** que se les diga que la vida es lo mejor que tenemos
EO113 E-i So ensar a dónde queremos llegar o tenemos claro que no **nos gusta** el mundo como está o que no nos tratan, pero, no sé
EO114 E-v Ci A Y porque **nos gusta** enmarcarnos de vez en cuando, pese a que es un tema
etc.

**nos gustaría** 5

EO113 E-i Cu inserto todo el Madrid de los Austrias y bueno, pues **nos gustaría**, nos gustaría ya que ha tenido usted la gentilez
EO113 E-i Cu l Madrid de los Austrias y bueno, pues nos gustaría, **nos gustaría** ya que ha tenido usted la gentileza de comparece

EO114 E-i Cu **Nos gustaría** que nos dijeran de qué temas no hemos hablado, q
EO121 E-x Po A Nos hacíamos una pregunta general y de eso **nos gustaría** que fuera el gran titular de futuro para esta no
Ho113 MC- So os, se utilice realmente para algo, y no digo que no **nos gustaría** sus felicitaciones, pero, creo, que nos van a gu

**te gusta** 12

EO111 E-i Me A ¿Qué **te gusta** más, eh, Cristina Almeida, Ministra de Asuntos Socia
EO111 E-i Me de esculpir en hierro y en frío, ¿qué es lo que más **te gusta** a ti en la vida? ¿Cuáles son tus aficiones?, y apart
EO114 E-v Ci Es un juego de palabras que **te gusta** mucho hacer. D No, el problema está en que todo tie
EO114 E-i Cu en primer lugar, piensa la gente que si eres, que si **te gusta** alguien como Sánchez Mazas, inmediatamente piensan q
etc.

**te gustan** 2

EO114 E-i So A Que son los que más **te gustan** a ti. D Los ojitos de tu cara. A Los ojos derechos
HO003 MJ C- pero, cuando **te gustan** las cosas, te metes tanto, que puedes brillar, si te

**te gustaría** 1

HO114 MC- So B Qué gentil, pues, qué **te gustaría** a ti desayunar. A Chile con huevo. B Chile con h

La enseñanza de la lengua **para fines específicos** ha acaparado la atención en los últimos años. Y con toda razón, ya que muchos discentes de un idioma extranjero lo son precisamente porque lo necesitan para comunicarse mejor en el ámbito de una determinada profesión, como la medicina, la electrónica, la informática, los negocios, etcétera. En estos casos es todavía más necesario contar con dos ingredientes, tanto en los materiales de clase como en el enfoque docente: los listados de léxico y estructuras específicas y los contextos en los que ambos elementos se dan. Para el logro de tales objetivos es inestimable la ayuda que puede ofrecer un corpus: fiabilidad de los recuentos léxicos por sectores y contextos (oracionales y textuales). El utensilio ha nacido y empieza a ser conocido. Sólo resta ampliar su uso y perfeccionar su accesibilidad al usuario; en este caso, el profesor.

## 7. Los textos de un corpus como fuente de ejercicios

Los repertorios textuales no constituyen por sí mismos un conjunto de ejercicios o actividades que el profesor puede aplicar directamente a la clase. Sería equivocado aproximarse a ellos desde esta perspectiva. Los corpus constituyen, sin embargo, un conjunto de materiales lingüísticos susceptibles de convertirse en una fuente inagotable de referencia e inspiración para autores y profesores.

Uno de los inconvenientes habituales que acechan a autores y a profesores es la carencia de **variedad en el uso de formas y palabras o en el uso de textos contextualizados en situaciones determinadas**. El corpus puede paliar esta dificultad propiciando que los ejercicios no se refieran siempre a lo que hacen los alumnos cada día o en cada clase. Con la ayuda

del corpus es fácil ampliar el horizonte situacional y formal. Además, la tipología de ejercicios habituales en el aula es susceptible de ampliarse a materiales del corpus. Piénsese, por ejemplo, en algunos modelos, como los siguientes:

**a.** Presentar un texto del cual se han extraído determinadas palabras
   — al azar
   — buscando formas determinadas (concordancias, adjetivos, verbos...)

**b.** Presentar un texto con lagunas, listando las palabras extractadas.

**c.** Presentar un texto con lagunas y opciones de sinonimia para las voces extractadas.
   Este tipo de ejercicio puede ser más sofisticado si lo hacemos con ordenador y éste cuenta con un diccionario de sinónimos y antónimos.

**d.** Reconstruir palabras en oraciones tomadas del corpus.

**e.** Reconstruir oraciones del corpus a las que se les ha extractado la parte final.

**f.** Elaborar una gran variedad de ejercicios sobre preposiciones, adverbios y similares, mediante el simple recurso a la eliminación y reconstrucción de estos elementos.

**g.** Organizar un texto que ha sido desorganizado. También esto puede hacerse de manera automática con la ayuda del ordenador, tanto para reconstruir como para deshacer.

Además, el profesor encontrará ayuda para:

— elaborar listados de palabras con terminaciones iguales o parecidas en los aspectos definidos,
— listar palabras acabadas en **-as -as, -iz/a, e/e, ...**
— listar palabras de igual longitud,
— listar palabras acabadas en una determinada secuencia o que empiezan por una secuencia especial,
— identificar palabras masculinas, femeninas, ...
— aislar palabras que contengan **h, g, gue, güe...**
— seleccionar textos en los que aparezca un determinado número de palabras: **economía, casa, medicina, automóvil**, ...

Parte de los ejercicios reseñados se generarán de manera automática mediante el correspondiente programa, el cual incluirá las pautas para eliminar una de cada *x* palabras, que el alumno debe reconstruir; o para eliminar ciertas terminaciones.

De igual manera puede ser manipulado con vistas a la sustitución de plurales por singulares o viceversa, hacia el cambio de un tipo de oración

en otra, hacia la búsqueda de sinónimos o antónimos (mediante la ayuda de un diccionario interno), etcétera. Se trata, en todos los casos, de ejercicios estructurados que pueden revestir una gran variedad, dentro de la rígida lógica a la que están condicionados por el programa informático.

## 8. Los corpus etiquetados

Actualmente los corpus pueden ofrecer mayores posibilidades mediante la ayuda de la etiquetación. Etiquetar un corpus consiste en añadir a cada palabra una etiqueta con determinada información sobre ella. De esta manera el ordenador es capaz de identificar, mediante las órdenes oportunas, si lo que en el texto es una secuencia de caracteres separada de las demás secuencias por un espacio en blanco es, además, un nombre, un adjetivo, un tiempo de pasado del verbo «ser», etc. Las etiquetas pueden revestir una gran complejidad: todo depende de las decisiones que se tomen a la hora de anotar la cantidad de información que deseamos consignar. El etiquetado para la clase de E/LE empieza a ser *rentable* con información que contenga las partes de la oración. A partir de ahí un etiquetado será tanto más aprovechable en el aula cuanta mayor información pueda recabarse de cada palabra. He aquí un ejemplo de un etiquetado básico en el cual los elementos lingüísticos aparecen unidos a su categoría gramatical mediante un guión:

El_ART Consejo_NP de_PREP Universidades_NP ha_VHABER dejado_PART claro_ADV que_CONJ los_ART préstamos_N avalados_ADJ por_PREP el_ART Estado_NP para que_CONJ los_ART estudiantes_N puedan_V cursar_VINF carreras_N universitarias_ADJ ,_PUNT son_VSER incompatibles_ADJ con_PREP el_ART disfrute_N de_PREP cualquier_QUANT beca_N ._PUNT Según_PREP este_DEM organismo_N ,_PUNT los_ART créditos_N se_PRON conciben_V como_CONJ una_ART retribución_N a_PREP los_ART estudiantes_N con_PREP buenas_ADJ calificaciones_N académicas_ADJ ,_PUNT pero_CONJ con_PREP un_ART nivel_N de_PREP renta_N familiar_ADJ que_REL les_PRON impide_V acceder_VINF a_PREP una_ART beca_N ._PUNT
Por lo tanto_CONJ ,_PUNT se_PRON estima_V que_CONJ en_PREP el_ART periodo_N comprendido_ADJ entre_PREP los_ART años_N 1995_NUM y_CONJ 2004_NUM ,_PUNT unos_ART 250000_NUM estudiantes_N pueden_V recibir_VINF un_ART crédito_N avalado_ADJ por_PREP el_ART sector_N público_ADJ ._PUNT Con_PREP el_ART préstamo_N concedido_ADJ ,_PUNT estas_DEM personas_N podrían_V disponer_VINF anualmente_ADV de_PREP 500000_NUM pesetas_N para_PREP sufragar_VINF los_ART gastos_N que_REL ocasiona_V estudiar_VINF una_ART carrera_N universitaria_ADJ ._PUNT Los_ART mismos_ADJ prestatarios_N dispondrán_V de_PREP un_ART periodo_N de_PREP carencia_N para_PREP devolver_VINF el_ART crédito_N ,_PUNT asociado_ADJ al_CNTR momento_N en_PREP que_REL logren_V encontrar_VINF empleo_N ;_PUNT en_PREP este_DEM sentido_N ,_PUNT las_ART administraciones_N públicas_ADJ destinarán_V casi_ADV 43_NUM millones_N de_PREP pesetas_N para_PREP la_ART subvención_N de_PREP intereses_N y_CONJ avales_N ._PUNT

Con textos anotados de esta manera, el profesor ya puede pedir al ordenador que substituya o elimine de un texto los nombres, las formas verbales en infinitivo, los adjetivos, etcétera. Sobre esta base se incrementa notoriamente la calidad y variedad de los ejercicios. El texto anterior podría dar origen a ejercicios como: **Sustituye los nombres/preposiciones/artículos... que faltan en el texto**, **Reconstruye los nombres/verbos... con sus correspondientes terminaciones**, etcétera.

Los corpus no tienen por qué ser considerados como algo cerrado y finalizado. El profesor tiene la posibilidad de añadir sus propios textos a un corpus que ya posea. Sobre esos textos añadidos le será posible elaborar también ejercicios, aumentando las posibilidades de trabajo y, sobre todo, ajustándose mejor a los fines especiales de cada clase o temática.

Una última ventaja que debe atribuirse a los corpus reside en el hecho de que reflejan **de manera más sobresaliente el uso lingüístico**, mientras que los manuales suelen poner el énfasis en lo que **es gramaticalmente correcto**. Esta realidad presenta también sus peligros: los corpus, en cuanto que son reflejo del lenguaje usado, pueden contener usos poco rigurosos en relación con lo que las gramáticas y manuales catalogan como 'correcto'. En tal caso el profesor tendría que ser capaz de tomar las decisiones oportunas para no introducir en el aprendizaje elementos que lo distorsionasen o entorpeciesen. No obstante, tal problema suele eliminarse de antemano si la empresa que comercializa un corpus restringe el ámbito de uso a lo que es gramaticalmente correcto y aceptado, objetivo con el que hasta el momento se han hermanado las muestras publicadas.

## Bibliografía


García Hoz, V. (1953), *Vocabulario usual, común y fundamental*, Madrid, CSIC.

Instituto Cervantes (1995), *Informe sobre Recursos Lingüísticos para el Español*, Alcalá de Henares, Instituto Cervantes.

Instituto Cervantes (1996), *Informe sobre Recursos Lingüísticos para el Español (II): Corpus escritos y orales disponibles y en desarrollo en España*, Alcalá de Henares, Observatorio Español de Industrias de la Lengua, Instituto Cervantes.

Márquez Villegas, L. (1975), *Vocabulario del español hablado*, Madrid, SGEL.

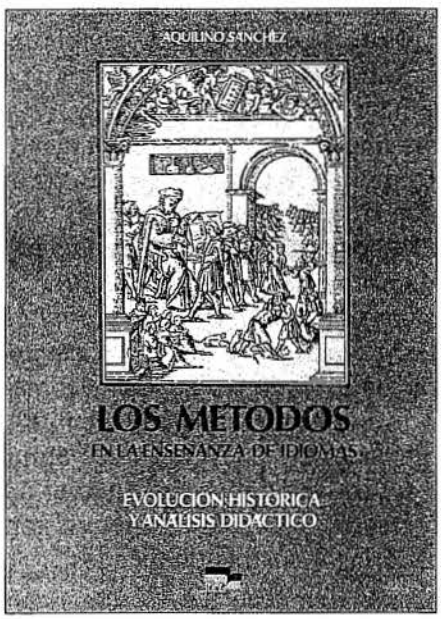
Sánchez, A., y otros (1995), *Cumbre. Corpus lingüístico del español contemporáneo. Fundamentos, metodología y aplicaciones*, Madrid, SGEL.

Sánchez, A., y Cantos, P. (1997), *Sub-corpus 'Cumbre' etiquetado para la investigación y práctica lingüística*, Madrid, SGEL (de próxima aparición).

Sánchez Lobato, J., y Aguirre, B. (1992), *Léxico fundamental del español*, Madrid, SGEL.

# LOS MÉTODOS

## EN LA ENSEÑANZA DE IDIOMAS




**AQUILINO SÁNCHEZ**

Descripción detallada y analítica de los
principales métodos que, a lo largo de la historia,
se han utilizado
en la enseñanza de lenguas extranjeras:

MÉTODO INTERLINEAL, VÍA MEDIA,
MÉTODO DE LA DOBLE TRADUCCIÓN, MÉTODO DIRECTO,
MÉTODO DEL SILENCIO, MÉTODO SUGESTOPÉDICO,
MÉTODO AUDIO-ORAL...

SGEL

SOCIEDAD GENERAL ESPAÑOLA DE LIBRERÍA, S.A.