

NOTAS

VERBOS MÁS COMUNES DELANTE DE LOS PRONOMBRES
INGLESES *I* Y *WE*. ANÁLISIS CONTRASTIVO BASADO
EN DOS TIPOS DE CORPUS

WILLELMIRA CASTILLEJOS LÓPEZ*

Dirección de Centros Regionales, Universidad Autónoma Chapingo

Resumen: Este es un estudio contrastivo basado en el análisis de dos corpus lingüísticos automatizados, un corpus de inglés nativo y uno de aprendientes cuya lengua materna es el español. Los dos tipos de corpus se almacenaron en el programa *Wordsmith* con los nombres de LOCNESS (Corpus de Inglés Nativo) e ICLESP (Corpus de Aprendientes-Español). El objetivo es comparar los resultados obtenidos por ambos corpus a través del manejo de la herramienta de concordancias (*Concord*), con el fin de conocer cuáles son los verbos que comúnmente siguen a los pronombres *I* y *We*. Ambos corpus son parte del acervo de datos del Centro de Lingüística de Corpus en Inglés de la Universidad Católica de Lovaina, Bélgica.

PALABRAS CLAVE: CORPUS, CORPUS NATIVO, CORPUS DE APRENDIENTE, HERRAMIENTA DE CONCORDANCIA, PRONOMBRE

Abstract: *This is a contrastive study based on the analysis of two types of corpora: a native English corpus and a learner corpus produced by native Spanish speakers. Both corpora were stored in the Wordsmith program under the labels LOCNESS (native corpus) and ICLESP (Spanish-speaking learner corpus). The objective of this work is to compare the results produced by both corpora through the use of the*

* williecastillejos@hotmail.com

concord tool, so as to know what verbs typically follow the pronouns I and We. The corpora make part of the computerized databank of the Louvain Center for English Corpus Linguistics at the Catholic University of Leuven.

KEY WORDS: CORPUS, NATIVE CORPUS, LEARNER CORPUS, CONCORD TOOL, PRONOUN

INTRODUCCIÓN

El método basado en la lingüística de corpus ofrece la ventaja de aislar una muestra de datos lingüísticos a partir de los cuales se observa un fenómeno particular, por ejemplo, la compañía preferida de unas palabras con otras (colocación). En este trabajo, se observarán las expresiones verbales comúnmente utilizadas por hablantes nativos y por aprendientes de la lengua inglesa después de los pronombres de primera persona en singular y plural: *I* y *We*. Esta observación es interesante porque refleja en qué medida se aproximan o separan los usos que nativos y aprendientes dan a una lengua.

Si se parte de la idea de que el hablante nativo es quien utiliza su propia lengua, por que es quien mejor conoce las funciones y formas contextuales de su idioma, se podría suponer que el aprendiente de la lengua en cuestión, en este caso del inglés, tendría que aproximarse a esos usos. La lingüística de corpus permite probar, con datos lingüísticos reales, en qué medida los aprendientes se aproximan a esos usos, o en todo caso, permite observar las diferentes estrategias utilizadas por ellos al expresarse en una lengua extranjera, estrategia normalmente caracterizada por lo que Selinker (1972) llama *interlengua*. Una vez realizado el análisis, se tendrá la posibilidad de afirmar cuáles son las expresiones verbales comúnmente utilizadas después de los pronombres *I* y *We* en el inglés escrito por hablantes nativos y por aprendientes, además se observará si las expresiones difieren dependiendo de la lengua materna de los hablantes.

DESCRIPCIÓN DEL CORPUS

Los dos corpus utilizados en este trabajo son parte de los archivos almacenados en *Wordsmith tools*, un programa informático diseñado para observar el *comportamiento* de las palabras en inglés. El corpus nativo, LOCNESS que significa

Louvain Corpus of Native English Essays, contiene textos argumentativos en forma de ensayos escritos por estudiantes hablantes nativos del inglés estadounidense. Para cumplir con el propósito de esta investigación utilicé una muestra de 149 841 palabras o *tokens*.

El otro corpus es ICLE (*International Corpus of Learner English*), este fue el resultado de un trabajo en el que colaboraron universidades de diferentes países. El corpus completo consta de más de dos millones de palabras y se elaboró a partir de ensayos escritos por estudiantes de inglés con lenguas maternas diferentes. De acuerdo con estas lenguas se clasificaron los textos en subcorpus. Actualmente existen 19 subcorpus, el español es uno de ellos (ICLESP). El nivel de los estudiantes de inglés como lengua extranjera es avanzado. La muestra del subcorpus de español constó de 197 435 palabras o *tokens*.

Aunque difieren en extensión, los corpus descritos contienen información que puede analizarse de forma paralela, ya que la redacción de los ensayos es homogénea en el estilo y los temas, por este motivo la comparación es apropiada.

MÉTODO

La información sobre el tamaño de la muestra se obtuvo a través de la herramienta *Wordlist*, en *Wordsmith*. Esta herramienta genera datos específicos de los archivos almacenados en el programa y proporciona información sobre el total de palabras, la longitud de las oraciones y la variedad léxica, entre otros aspectos. Por otro lado, como el tamaño de los corpus era diferente, se utilizaron frecuencias relativas con el fin de interpretar la frecuencia de uso de los pronombres *I* y *We* en ambos corpus. Aunque el objetivo principal de este trabajo no consiste en saber las frecuencias de uso de los pronombres *I* y *We*, sino las expresiones verbales que introducen, se registra esa información debido a que los porcentajes calculados se basan en el total de entradas que corresponden a la concordancia.

La principal herramienta de *Wordsmith* que se utilizó fue *Concord* y los parámetros para extraer la información se establecieron de la siguiente manera:

- La palabra de búsqueda *I* y *We* se escribieron en el apartado *concordance settings*.

- No se establecieron palabras específicas de contexto, sólo se determinaron los horizontes de búsqueda, que fueron 5L y 5R.
- La clasificación (*sort*) de la concordancia se fijó en *Main sort 1R, then by 2R*, y distribución ascendente.

Para una mejor observación de las regularidades, el total de listados de las concordancias se imprimieron. De este modo se pudo apreciar con mayor facilidad cuáles fueron los verbos más utilizados, tanto por los nativos como por los extranjeros. Para confirmar el total de ocurrencias relacionadas con el verbo, se recurrió nuevamente al programa, sin embargo, esta vez se incluyó una palabra de contexto en *concordance settings*. También se mantuvieron los mismos parámetros que para el primer caso y se verificó el total de ocurrencias con el conteo manual, además se determinó con criterio de *frecuente* a todos los verbos que hubieran ocurrido al menos dos por ciento del listado total de la concordancia. Aunque arbitrario, el criterio de este porcentaje fue importante para establecer una comparación confiable entre ambos corpus. Recuérdese que el objetivo de este trabajo es identificar lo que se encuentra después de *I* y *We*, más que exponer la frecuencia de estos pronombres en los corpus.

Los resultados se expresan en forma de tablas. La determinación de las diferencias significativas entre las distribuciones de palabras en uno y otro corpus se basó en la prueba *Chi* cuadrado. A través de ella, se pudo observar que los hallazgos fueron resultado de diferencias auténticas entre dos palabras o no.

RESULTADOS Y ANÁLISIS

El primer aspecto que se consideró en el estudio fue el tamaño de los corpus. Como se mencionó anteriormente, esto se realizó con la herramienta *Wordlist*, y los resultados fueron:

- LOCNESS: 149 841 palabras o *tokens*
- ICLESP: 194 435 palabras o *tokens*

Después se establecieron las concordancias con las palabras de búsqueda *I* y *We*, utilizando como parámetros 1R y 2R. Esta operación me permitió calcular el total de ocurrencias de *I* y *We*, y, por consiguiente, observar de los verbos más frecuentes en los distintos corpus. El resultado se muestra en la tabla 1.

Corpus	Total de ocurrencias	
	<i>I</i>	<i>We</i>
LOCNESS	679	432
ICLEPS	712	1595

TABLA 1: FRECUENCIAS ABSOLUTAS DE LOS PRONOMBRES *I* Y *We* EN AMBOS CORPUS.

Debido a que ambos corpus no comprenden la misma extensión no es posible comparar la frecuencia directamente; por eso sólo se calculó la frecuencia relativa.

Corpus	Total de ocurrencias	
	<i>I</i>	<i>We</i>
LOCNESS	453	288
ICLEPS	360	870

TABLA 2: FRECUENCIAS RELATIVAS DE *I* Y *We*.

ANÁLISIS DE LAS FRECUENCIAS *I* VS. *We*.

Las tablas anteriores muestran un índice distinto en el uso de los dos pronombres. En el corpus nativo (LOCNESS) se observa un mayor uso de *I*, en relación con *We*, mientras que en el corpus de los aprendientes (ICLEPS) ocurre lo contrario. En términos de discurso, este resultado aporta datos sobre la forma en que unos y otros hablantes utilizan la lengua. El pronombre *I* determina a la persona que emite una idea atribuyéndose la responsabilidad directa de la misma; en cambio en el caso de *We*, la implicación involucra una especie de invitación a compartir esa responsabilidad. Según Émile Benveniste (citado en Meunier y Peraya, 2004: 86), *We* no es la pluralidad de *I* sino una conjunción de *I* y *We*, conjunción que no tiene el mismo estatus debido a que el pronombre *I* desempeña el papel principal. *We* existe porque *I* existe primero y a esto se agrega el elemento que implica también *not I*. *We* es una especie de *I* amplificado y puede tomar dos formas: *We* inclusivo (*I + You*), cuando un profesor dice en clase: *We must work hard this semester*. La otra forma es *We* exclusivo (*I +*

They), conservando el ejemplo de ese mismo profesor, diría: *We are developing new forms to teach languages in the campus.*

ANÁLISIS DE LOS ELEMENTOS DELANTE DE *I*

La tabla 3 muestra los verbos y auxiliares modales introducidos por el pronombre *I* en ambos corpus. Recuérdese que se expone únicamente 2 por ciento de los resultados relacionados con el total de ocurrencias de *I* y *We*. Los verbos que comparten ambos corpus son *be*, *believe*, *have*, *think* y *know*. Como *be* y *have* funcionan además como verbos auxiliares, presentan un índice mayor de ocurrencia, en particular en el corpus nativo. Debido a que el corpus contiene ensayos escritos, la tendencia de los hablantes en ambos corpus hacia verbos performativos es rara; los verbos de mayor uso son aquellos que describen, no tanto que transforman —hablando en términos austinianos—. Los verbos utilizados comprometen la responsabilidad del hablante hacia opiniones (*think*), creencias (*believe*), o bien, dan cuenta de cualidades específicas (*be*, *have*). Es curioso observar que en el corpus nativo la línea semántica de los verbos utilizados se mantiene, mientras que en el corpus hispanohablante se observan otras líneas, una categoría asertiva, como en los verbos: *say*, *agree*, *mean* e incluso una directiva (*want*). Los verbos *like* y *hope* mantienen la misma fuerza ilocutoria de los verbos no performativos, es decir, dan cabida a una opinión libre que no implica “transformar el mundo”, simplemente emiten una evaluación sobre algo, al describirlo. Algunos ejemplos ilustran estas afirmaciones:

- 1) *I would like to make a distinction between [...]*, (ICLESP 687)
- 2) *I like to think that the European Union is the [...]*, (ICLESP 349)

LOCNESS	Número de ocurrencias	%	ICLESP	Número de ocurrencias	%
<i>Be</i>	113	17	<i>Agree</i>	14	2
<i>Believe</i>	30	4	<i>Be</i>	82	11
<i>Feel</i>	55	8	<i>Believe</i>	28	4
<i>Have</i>	96	14	<i>Have</i>	54	8
<i>Know</i>	39	6	<i>Like</i>	33	5
<i>Think</i>	57	8	<i>Know</i>	28	4
			<i>Mean</i>	24	3
			<i>Say</i>	59	8
			<i>Think</i>	125	18
			<i>Want</i>	15	2
<i>Can</i>	19	3	<i>Can</i>	18	3
<i>Could</i>	14	2	<i>Will</i>	20	3
<i>Would</i>	32	5	<i>Would</i>	38	5

TABLA 3: VERBOS Y AUXILIARES MODALES FRECUENTES DELANTE DE I.

En lo que se refiere a los auxiliares modales, *can* y *would* se manifiestan en ambos corpus mientras que *could* sólo se usa en LOCNESS y *will* en ICLESP. La idea de posibilidad predomina entonces en el corpus nativo. En cuanto a *would*, es interesante observar que en el corpus nativo se usa con una variedad de verbos, mientras que en el no nativo más de 50 por ciento de las ocurrencias utilizaron sólo dos verbos: *like* y *say*. Una posible interpretación de este hecho se debe a que los hablantes no nativos disminuyen así la fuerza de sus aserciones.

La comparación entre el corpus nativo y el corpus aprendiente registró algunas diferencias en relación con el verbo *believe* y los modales *can* y *would*. Aunque la distribución no fue significativa, la observación de los usos verbales demuestra grandes diferencias: con el verbo *believe*, en 46 por ciento de los casos, los aprendientes utilizaron la conjunción *that*, lo que no ocurrió en el corpus nativo. *Can* también se utilizó en la forma afirmativa, la mayoría de los casos en el corpus aprendiente, y *would*, como se mencionó antes, prefirió los verbos *like* y *say*. En cuanto al verbo *to be*, hubo mayor variación gramatical en el corpus nativo mientras que en el aprendiente el uso principal fue el presente. El verbo *have* se utilizó como auxiliar del presente perfecto en la misma proporción en ambos corpus, y resulta

interesante notar que los principales verbos utilizados en el corpus aprendiente fueron *say* y *mention*, que se refieren a ideas previas en el texto. Por su parte, el verbo *think* ofrece otro dato importante en este, ya que 30 por ciento de los casos presentaron la conjunción *that* —igual que con el verbo *believe*— delante del verbo. El uso de *that* después de *think* y *believe* refleja una preferencia hacia esquemas subordinados por parte de los aprendientes.

En esta sección me referiré al análisis con el pronombre *We*. La tabla 4 expone los verbos más frecuentes y el porcentaje en relación con el total de ocurrencias con *We*.

LOCNESS	Número de ocurrencias	%	ICLESP	Número de ocurrencias	%
<i>Be</i>	78	18	<i>Be</i>	206	13
<i>Have</i>	78	18	<i>Consider</i>	35	2
<i>Live</i>	13	3	<i>Find</i>	91	6
<i>Know</i>	20	5	<i>Get</i>	34	2
<i>Need</i>	19	4	<i>Have</i>	258	16
<i>See</i>	21	5	<i>Know</i>	44	3
			<i>Live</i>	28	2
			<i>Say</i>	53	3
			<i>See</i>	143	9
			<i>Think</i>	45	3
<i>Can</i>	23	5	<i>Can</i>	236	15
<i>Must</i>	24	6	<i>Must</i>	61	4
<i>Should</i>	9	2	<i>Should</i>	43	3
<i>Will</i>	14	3	<i>Will</i>	43	3

TABLA 4: VERBOS Y AUXILIARES MODALES FRECUENTES DELANTE DE *WE*.

ANÁLISIS DE ELEMENTOS INTRODUCIDOS POR *WE*

Una vez más, los verbos *be* y *have* fueron los más frecuentes, sin embargo, a diferencia de *I* presentaron una mayor variedad de verbos compartidos en ambos corpus, éstos fueron *live*, *know* y *see*. Asimismo, destacó la tendencia sobre el uso de los verbos *consider*, *find* y *get* en el corpus aprendiente. Estos hallazgos son interesantes porque reflejan una posición diferente del hablante cuando

utiliza *We*, en relación con *I*. Como se ha mencionado, *We* implica una conjunción de *I* y *not I*, una suerte de invitación a que otros participen en alguna declaración hecha por *I*. Tal vez por eso se explica la frecuencia de uso de *see*, *know* y *live* en ambos corpus, pues se refieren a la existencia de una experiencia común, al menos así lo perciben los hablantes. Igual que en el caso de *I*, los verbos *be* y *have* son los de mayor uso debido a su función como auxiliares. Todos los verbos utilizados con *We* aluden a hechos que se comparten, además declaran o describen, pero no son ejecutivos. Contémplese la naturaleza de los verbos *know*, *consider*, *find*, *think*, *see*, etcétera. En realidad dan a entender significados semejantes: el hecho de concebir, visualizar o considerar un hecho en conjunto. Por supuesto, los verbos *find* y *see* son utilizados en este artículo en el sentido de “percibir una idea o una situación mentalmente, en relación con el entendimiento”. Aquellos verbos cuya naturaleza tiende hacia los performativos austinianos se presentaron en muy baja escala (*speak*: asertivo, *want*: directivo). Incluso el verbo *live* fue utilizado en más ocasiones con el sentido de “experimentar o conocer”.

Los auxiliares modales utilizados en ambos corpus fueron principalmente *can*, *could* y *will*, pero hubo también modales de obligación como *must* y *should*, hecho que resultó diferente en *I*, donde prácticamente no aparecieron. Los modales con *We* dan a entender la idea de un compromiso generalizado, la responsabilidad de lo que se dice no es individual, sino compartida con otros. Un hecho interesante en los resultados fue que, a diferencia del pronombre *I*, donde *would* tuvo alta frecuencia, en el caso de *We* ni siquiera alcanzó dos por ciento en ninguno de los dos corpus. Desde una perspectiva semántica, esto explica que los hablantes prefieran una expresión que implica posibilidad cuando utilizan *I*; ya que el nivel de compromiso es mayor; por ello el hablante busca una especie de *halo protector* en sus aserciones con *I*. En cambio con *We*, al compartir ese compromiso, la tendencia de los modales se orienta más hacia la obligación y menos a la posibilidad.

En cuanto a la prueba *Chi* cuadrado, cuatro elementos presentaron una distribución poco significativa: *have*, *must*, *should* y *will*; mientras que otros cuatro se registraron con diferencias muy notorias: *be*, *known*, *see* y *can*. Los elementos con distribución poco significativa presentaron un uso distinto: en el corpus nativo se utilizó el verbo *have* principalmente en presente simple y presente perfecto; en cambio, en el corpus aprendiente, también se empleó el verbo *have* en alta escala, hubo una preferencia importante (28 por ciento)

por el uso de *have + to*, lo que no ocurrió en el corpus nativo. En los casos de *must* y *should* no se observaron rasgos significativos, ambos corpus utilizaron verbos variados. Sólo con *will*, el corpus nativo indicó una tendencia predominante hacia los verbos *see* y *have* (23 y 14 por ciento, respectivamente).

Aquellos verbos cuya distribución fue notable reflejan un sobre uso por parte de un corpus, como se aprecia en la tabla 4. El verbo *know*, que se utilizó considerablemente en el corpus nativo, ofrece un detalle curioso en el corpus aprendiente: en 18 por ciento de los casos, se anotó el verbo precedido por el pronombre *all* y seguido de la conjunción *that* (en 23 por ciento de los casos). El verbo *see* se combinó muchas veces con el modal *can* en el corpus aprendiente (52 por ciento), y otras expresiones frecuentes fueron *see that* (ocho por ciento) y *will see* (ocho por ciento).

Con excepción de *can*, los tres modales compartidos en ambos corpus: *must*, *should* y *will*, no tuvieron una distribución significativa. El corpus nativo ofreció una variedad de verbos en las ocurrencias; por el contrario, el corpus aprendiente prefirió los verbos *be*, *remember* y *consider* acompañados con *must*. *See* y *have* se combinaron con *will*. En el caso de *can*, nueve por ciento de las veces se utilizó con el adverbio *also*.

CONCLUSIONES

En el corpus nativo, el índice de uso del pronombre *I* es más que en el del aprendiente. Los hablantes que no son nativos utilizan *We* en una proporción casi doble. La explicación de esto es de orden psicosociolingüístico. El carácter asertivo de una persona se ve modificado cuando se usa una lengua extranjera. Los hablantes de lengua materna parecen comprometer el contenido de sus expresiones de modo individual, mientras que los aprendientes tienden a generalizar y hacer participar un colectivo en sus opiniones y aserciones.

Karin Aijmer (1980: 15) señala que las personas expresan las ideas de manera distinta dependiendo de “la fuerza del compromiso que sienten hacia la verdad”. Este estudio evidencia esa afirmación. Además de los verbos *be* y *have*, cuya frecuencia se deriva de su función como auxiliares, aquellos verbos con mayor “compromiso hacia la verdad” fueron empleados en gran medida por los nativos. Según Aijmer (1980: 39), el verbo *know* “implies that the subject is satisfied that his evidence is adequate (*complete*) and is prepared to disregard any further

information". De acuerdo con este autor, *know* posee un carácter diferente al de *believe*, puesto que *I know* declara algo, mientras que *I believe* se utiliza para admitir que existen declaraciones alternativas que podrían ser verdaderas. En términos pragmáticos, *I believe* da vueltas alrededor de una idea en una situación en la que el hablante pretende evitar una responsabilidad personal. Lo mismo ocurre con el verbo *think*. Cuando los aprendientes querían ser asertivos utilizaron *mean* y *say*.

Los fenómenos muestran interesantes resultados en el uso excesivo, de *would* como papel de la modalidad hipotética en el discurso. Mediante el uso (frecuente) de *would* con los verbos *see* y *say* en el corpus aprendiente, se posiciona a estos hablantes en el mismo nivel de compromiso hacia la verdad que se supone con *think* y *believe*. También significa que existe una actitud de cortesía por parte de estos hablantes y que se respeta el rostro de los otros, pues no es tan directo. Estas ideas de los actos indirectos para expresar cortesía pueden encontrarse en Clark y Lucy (1975) y Clark y Schunk (1980).

En cuanto al uso de *We*, los aprendientes parecen más asertivos que cuando utilizan *I*. Además se notó que hubo elección considerable de verbos cuyo significado implica entendimiento, concepción de algo y observación desde una determinada perspectiva, como en el caso de *think*, *see* y *know*. El verbo *believe* no fue tan preferido esta vez. Si para *I*, la línea semántica que se mantuvo fue la de verbos que expresaban compromiso directo o responsabilidad de lo que se dice, en el caso de *We* la línea semántica tendió más a experimentar o expresar una vivencia compartida. Ejemplos de esto, son los verbos *live*, *need*, *find*, *get* y *want*. Asimismo, el sentido de obligación moral se acentuó en los modales introducidos por *We*.

El tiempo gramatical predominante en ambos corpus fue el presente en forma afirmativa. Debido a la naturaleza escrita de los ensayos que alimentan estos corpus, los hablantes optan en primer lugar, por la línea argumentativa del presente afirmativo.

BIBLIOGRAFÍA

Aijmer, Karin (1980), *Evidence and the Declarative Sentence*, Estocolmo, Suecia, Almqvist y Wiksell International.

- Ballmer, Thomas y Waltraud Brennenstuhl (1981), *Speech Act Classification. A Study in the Lexical Analysis of English Speech Activity Verbs*, vol. 8, Berlín/Heidelberg/ Nueva York, Alemania/Estados Unidos, Springer Verlag, Springer Series in Language and Communication.
- Clark, Herbert y Dale H. Schunk (1980), "Polite responses to polite requests", *Cognition*, vol. 8, pp. 111-143.
- Clark, Herbert y Peter Lucy (1975), "Understandign what is meant from what is said: A study in conversationally conveyed requests", *Journal of Verbal Learning and Verbal Behavior*, vol. 14, pp. 56-72.
- Degand, Liesbeth (2005-2006), *Notas del curso Introduction to Semantics*, Lovaina, Bélgica, Université Catholique de Louvain.
- Ehrman, Madeline (1966), "The meaning of the modals in present-day American English", *Linguistics*, vol. 4, núm. 28, publicado en línea: 19/11/2009, pp. 46-58.
- Granger, Sylviane (1996), "From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora", en Karin Aijmer, Bengt Altenberg y Mats Johansson (eds.), *Languages in Contrast. Text-based Cross-linguistic Studies*, Lund, Suecia, Lund University Press, pp. 37-51.
- Meunier, Jean-Pierre y Daniel Peraya (2004), *Introduction aux théories de la communication*, Bruselas, Bélgica, De Boeck.
- Westney, Paul (1995), *Modals and Periphrastics in English, an Investigation into the Semantic Correspondence between Certain English Modal Verbs and their Periphrastic Equivalents*, Tubinga, Alemania, Niemeyer.
- Selinker, Larry (1972), "Interlanguage", *International Review of Applied Linguistics*, vol. 10, pp. 209-230.

Páginas web

- [http://www.georgetown.edu/faculty/ballc/webtools/web_chi.html] consultado el 2 de marzo de 2007.
- [<http://www.unc.edu/~preacher/chisq/chisq.htm>] consultado el 14 de junio de 2007.
- [http://www.uqtr.ca/~vandervk/english_performative_verbs_ch6.pdf] consultado el 26 de julio de 2007.