

INTERNET COMO CORPUS: EL CASO DE *BIBIDÍ*

MILKA VILLAYANDRE LLAMAZARES

Departamento de Filología Hispánica y Clásica

Universidad de León

Campus Universitario de Vegazana

24071 León

milka.villayandre@unileon.es

Si bien los corpus lingüísticos, tal y como se conciben en la actualidad, aportan a los estudios sobre el lenguaje una serie de ventajas indudables, hoy en día no hay que desdeñar las posibilidades que Internet brinda a los investigadores. Mediante el presente artículo, pretendemos mostrar, a través del análisis de un caso concreto (“bibidí”), las ventajas e inconvenientes que ofrecen tanto los corpus como Internet.

Palabras clave: corpus, Internet, bibliotecas digitales, buscadores web

During the last decades, corpora have proved to be a very useful tool for studying language thanks to the careful design criteria considered to select the texts that form part of a corpus. Nevertheless, Internet appears as a powerful resource due to the huge amount of data that provides every day. This paper aims at examining the pros and cons of both corpora and Internet through a particular case study.

Keywords: corpora, Internet, digital libraries, web search engines

0. Introducción

Si un corpus se caracteriza por haber sido recopilado siguiendo unos criterios de diseño acordes con su finalidad prevista de explotación, no por ello existe la garantía de que el corpus nos vaya a proporcionar toda la información que necesitamos. Es aquí donde Internet entra en juego, al aportar a diario un número ingente de muestras de lengua de las que, por desgracia, no es posible tener el debido control (cualquier persona se puede convertir en autor gracias a la Red; los textos no siempre están documentados, por lo que a veces es difícil determinar su procedencia y autoría; su vida suele ser efímera, lo que implica que en ocasiones no es

Contextos XXI-XXII/41-44, 2003-2004 (págs. 205-231). ISSN: 0212.6192

posible recuperar la fuente de la que hemos tomado los datos en un momento determinado, etc.). Sin embargo, con los debidos filtros y un poco de paciencia, Internet se convierte en una fuente de información muy útil para los estudios lingüísticos.

En las páginas que siguen, contrastaremos los pros y los contras del uso de corpus y de Internet en la investigación lingüística a través del análisis de un caso concreto, el de la palabra *bibidí*.

1. El concepto de corpus

En la actualidad el concepto de corpus ha cambiado mucho con respecto al que manejaban los primeros lingüistas que lo empleaban como recurso para sus investigaciones¹. Hoy en día se considera que los corpus deben cumplir los siguientes requisitos (*cf.* McEnery y Wilson 2001:21 y ss.):

1) *Textos en formato electrónico*: un corpus, para ser una herramienta útil al lingüista, debe estar informatizado, es decir, los textos de que consta tienen que estar en formato electrónico (corpus informatizado o automatizado). El empleo del ordenador permite automatizar tareas tales como:

- *Búsqueda de información*: un corpus informatizado permite localizar de forma rápida una palabra, una secuencia de palabras o incluso una categoría gramatical en décimas de segundo.
- *Recuperación de información*: un corpus informatizado permite obtener todos los casos de una palabra, secuencia de palabras, etc. registrados en el corpus, normalmente con su contexto inmediato anterior y posterior (*concordancia*).
- *Cómputo de la frecuencia* de aparición de una palabra, secuencia de palabras, etc.
- *Clasificación de los datos* contenidos en el corpus según diferentes criterios: orden alfabético, frecuencia de aparición, autor, procedencia geográfica, tema, medio de publicación, etc.

¹ Conjunto de textos, no necesariamente informatizado, que se consideraba imprescindible para la descripción de determinados aspectos lingüísticos (en especial, los fonético-fonológicos) en el contexto particular del estudio de lenguas amerindias por parte de los lingüistas estructurales americanos.

2) *Autenticidad de los datos*: los textos recogidos en el corpus deben ser muestras reales de uso de la lengua objeto de estudio. A partir de ellas se construyen (o verifican) de forma empírica las teorías que tratan de explicar el funcionamiento de la lengua o las aplicaciones computacionales.

3) *Criterios de selección*: los textos que forman parte del corpus deben haber sido elegidos de acuerdo con unos determinados criterios –lingüísticos y/o extralingüísticos– para la finalidad concreta que persiga el corpus².

4) *Representatividad*: la selección de los textos, además de a unos criterios adecuados, debe responder a parámetros estadísticos que garanticen que los textos “representan” la variedad de lengua objeto de estudio (*muestra representativa*). Esta variedad puede referirse a la obra de un autor determinado, a un período de tiempo, a un género, etc. Cuando lo que nos interesa es la lengua en su conjunto, la opción de reunir en un corpus todas las muestras de esta se hace impracticable, a diferencia, p. ej., de lo que ocurre si queremos recoger todas las obras de Lope de Vega, que son un universo cerrado. La única solución posible, entonces, es tomar una muestra más pequeña de esa lengua, que refleje, a pequeña escala, el funcionamiento del todo que es la lengua. Chomsky criticó con acierto que los corpus corren el riesgo de ser sesgados. Precisamente para subsanar este problema se recurre a la selección, según criterios estadísticos, de textos de diversos géneros, tipologías, temas, medios de publicación, etc.

5) *Tamaño*: por lo general, los corpus constan de un tamaño finito, que se suele medir en millones de palabras (o formas) y que se fija antes de empezar la recogida de los textos (p. ej. un millón de palabras); una vez alcanzado ese número, se da por terminada la recopilación del corpus, que no es más que el primer paso de todo el proceso³. Sin embargo, también existen corpus abiertos o monitor, como el del proyecto COBUILD, de especial interés para la lexicografía. En el pasado se pensaba que el tamaño era muy importante: mientras mayor fuera el corpus, más posibilidades

² Precisamente el uso de unos criterios previos y explícitos diferencia un corpus de otras recopilaciones de textos tales como archivos, colecciones o bibliotecas electrónicas.

³ Una vez seleccionados los textos, estos se codifican, anotan y explotan de múltiples formas, por lo que la recopilación en sí no es más que una primera fase necesaria en todo proyecto de corpus.

tenía de reflejar el funcionamiento real de la lengua en todas sus variedades, pero en la actualidad priman los criterios de diseño, es decir, el tamaño solo es importante en la medida en que así lo exija la finalidad del corpus⁴.

A continuación se recogen algunas definiciones de corpus que ilustran estas características⁵:

A collection of texts assumed to be representative of a given language, dialect, or other subset of a language, to be used for linguistic analysis (Francis 1982:7 apud Francis 1992:17).

A collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language (Sinclair 1994:14).

A finite-sized body of machine-readable texts sampled in order to be maximally representative of the language variety under consideration (McEnery y Wilson 1996:24).

Un corpus és una mostra d'una llengua que habitualment s'ha construït a partir d'una selecció de textos realitzada segons uns determinats criteris i amb un determinat objectiu (Martí y Castellón 2000:151).

The term corpus should properly only be applied to a well-organized collection of data, collected within the boundaries of a sampling frame designed to allow the exploration of a certain linguistic feature (or set of features) via the collected data (McEnery 2003:449).

Estos criterios y definiciones permiten discriminar los corpus, en el sentido que se maneja en la lingüística de corpus, de otras colecciones de textos electrónicos⁶ (cf. Sinclair 1996; Torruella y Llisterrí 1999:51-52):

⁴ Lógicamente, un corpus que pretenda ser representativo de una lengua en toda su variedad (español, inglés, francés...) no podrá conformarse con unos pocos millones de palabras, mientras que un corpus cuyo objetivo sea describir un sublenguaje (jurídico, informático...) puede permitirse un tamaño más reducido. Además, la disponibilidad de los textos es otro factor que puede influir en el tamaño.

⁵ Se han destacado tipográficamente las características relevantes.

⁶ En lingüística de corpus, el término *texto* se refiere tanto a una muestra de lengua escrita como a una de lengua oral (cf. SINCLAIR 1996).

- *Archivo* (o colección) *informatizado*: se trata de un simple conjunto de textos electrónicos sin estructurar. El único criterio que prevalece a la hora de conformarlo es la disponibilidad de los textos.
- *Biblioteca digital*: se trata de un conjunto de textos electrónicos recogidos sin seguir criterios lingüísticos, pero guardados en un formato estándar⁷.

Algunos ejemplos de archivos y bibliotecas que reúnen textos en soporte electrónico, pero que no pueden calificarse como corpus en sentido estricto, son:

- *Proyecto Gutenberg*⁸: proyecto pionero en lo que se refiere a la recopilación de libros o textos electrónicos. Data de 1971, momento en que fue fundado por Michael Hart y, desde entonces, recoge textos clásicos, textos de literatura ligera y obras de referencia anteriores a 1923.
- *Búsqueda de libros de Google*⁹, antes *Google Print*: proyecto para digitalizar y poner a disposición de los usuarios de Internet los fondos bibliográficos de las universidades de Stanford, Harvard, Princeton, Oxford, Michigan, Complutense de Madrid y de la Biblioteca Pública de Nueva York, entre otros centros que se suman al proyecto cada año.

⁷ A diferencia de archivos y bibliotecas, en los corpus informatizados los textos que los componen se han seleccionado según unos criterios lingüísticos y/o extralingüísticos y se han codificado de acuerdo con un estándar, con la finalidad de proporcionar una imagen real de la lengua a partir de la cual se pueden extraer generalizaciones para su estudio.

⁸ URL: http://www.gutenberg.org/wiki/Main_Page

⁹ URL: <http://books.google.com/>



Ilustración 1. Google libros.

- ❑ *The Universal Digital Library. Million Book Collection*¹⁰, un proyecto de la Universidad de Carnegie Mellon que nace con el ambicioso objetivo de poner en Internet todos los libros publicados en la historia de la humanidad y que ha empezado por la digitalización de un millón de ellos.
- ❑ *The Oxford Text Archive*¹¹: ubicado en la Universidad de Oxford, desde 1976 recoge textos en veinticinco lenguas de autores individuales, obras de referencia y corpus.
- ❑ *Electronic Text Center*¹², Biblioteca de la Universidad de Virginia. Desde 1992 recoge textos en quince lenguas.
- ❑ *Electronic Text Collections in Western European Literature*¹³: reúne textos literarios en diferentes lenguas europeas distintas del inglés. Es mantenida por James Campbell, *Western European Studies Section*,

¹⁰ URL: <http://www.ulib.org/>

¹¹ URL: <http://ota.ahds.ac.uk/>

¹² URL: http://lib.virginia.edu/digital/collections/finding_digital.html

¹³ URL: <http://www.lib.virginia.edu/wess/etexts.html>

Association of College and Research Libraries, American Library Association.

- *Biblioteca Virtual Miguel de Cervantes*¹⁴: proyecto que nace en 1999 por iniciativa de la Universidad de Alicante y el Banco Santander con el objetivo de recopilar las principales obras de literatura en español, así como otros recursos relacionados con la historia de España e Hispanoamérica.

2. Internet como corpus: *bibidí*

Por último, hay que hacer mención de Internet como un corpus, no en el sentido estricto del término que hemos expuesto (*vid. supra*), porque no sigue unos criterios de diseño y en muchos casos falta información sobre el número y procedencia de los textos. Sin embargo, hay que reconocer su utilidad, aunque sea con los debidos filtros, como fuente de información para los estudios lingüísticos¹⁵.

La web ofrece la posibilidad de acceder a un conjunto de textos, en formato electrónico, que son muestras reales de uso de la lengua de todo tipo y materia; que constituyen un proyecto abierto en cambio continuo; y que pueden ser recuperados mediante las diferentes funciones de búsqueda de cualquier navegador.


Sirva de ilustración de este uso de la web como corpus y su contraste con un corpus propiamente dicho, el siguiente caso, que planteamos como parte de una de las prácticas de la asignatura Lingüística Computacional, que tenía por objeto el uso de recursos electrónicos para obtener información lingüística¹⁶. El punto de partida fue la localización de una forma, *bibidí*, en las respuestas a una encuesta del proyecto VARILEX¹⁷:

¹⁴ URL: <http://www.cervantesvirtual.com/>

¹⁵ *Vid.* el caso que comenta MORALA (2002) para la palabra *fuereño*.

¹⁶ En el momento de la redacción del presente artículo, hemos completado y actualizado los datos con nuevas búsquedas.

¹⁷ URL: <http://gamp.c.u-tokyo.ac.jp/~ueda/varilex/cues/cues2000.pdf>. Para más información sobre el proyecto VARILEX, *Variación léxica del español en el mundo*, remitimos a la URL: <http://gamp.c.u-tokyo.ac.jp/~ueda/varilex/index.html>



B092 [ATHLETIC SHIRT] Camiseta sin mangas.

1) **bibidí**; 2) Bid; 3) camiseta; 4) camiseta de hombreras; 5) camiseta de resaque; 6) camiseta de sport; 7) camiseta de tirantes; 8) camiseta de tiras; 9) camiseta imperio; 10) camiseta malla; 11) camiseta manga sisa; 12) camiseta sin mangas; 13) camisilla; 14) esqueleto; 15) esqueletos; 16) franela; 17) franelilla; 18) malla; 19) musculosa; 20) playera; 21) camiseta de tirillas; 22) camisola; 23) playera sin mangas.

&) Otros: _____; #) No se me ocurre.

S) Comentario:

Ilustración 2. Localización de la forma “bibidí” en VARILEX.

Sin facilitar esta información a los estudiantes, se les pidió que realizaran una investigación, utilizando todos los recursos electrónicos a su alcance (en especial, diccionarios y corpus), para determinar: el significado de la palabra, su origen, su ámbito geográfico, sus variantes formales, etc.

La experiencia resultó muy interesante y, gracias a ella, pudimos comprobar cómo en el *DRAE*¹⁸ esta palabra aparece recogida en su variante *bivirí*:

bivirí.

(De BVD, marca reg.).

1. m. *Perú*. Camiseta interior masculina.

Real Academia Española © Todos los derechos reservados

Ilustración 3. Localización de la forma “bivirí” en el DRAE.

Sin embargo, esta información fue la última a la que llegaron, pues, al buscar *bibidí* inicialmente en el *DRAE*¹⁹, este nos devuelve el mensaje de que la palabra no está en el Diccionario²⁰.

¹⁸ Los resultados de las consultas al *DRAE* son válidos hasta la fecha de redacción de este artículo.

¹⁹ Por defecto, el *DRAE* selecciona la opción de “Búsqueda por aproximación”, es decir, no solo va a intentar recuperar la forma propuesta, sino otras que estén en el diccionario y tengan una

Continuando las indagaciones, procedemos a realizar la misma consulta en los bancos de datos de la Real Academia, con la misma respuesta negativa tanto en el CREA²¹ como en el CORDE: “No existen casos para esta consulta”. Sin embargo, gracias al empleo de comodines (*?i?idí*) –que los bancos de datos académicos sí permiten–, obtenemos los primeros frutos de nuestra búsqueda en el CREA: 16 casos en 13 documentos.

Real Academia Española - Corpus de Referencia del Español Actual (CREA)

Consulta:

Criterios de selección:

Autor:

Obra:

Cronológico:

Medio: (Todos) Libros Periódicos Revistas Miscelánea Oral

Geográfico: (Todos) Argentina Bolivia Chile Colombia Costa Rica

Tema: (Todos) 1.- Ciencias y Tecnología. 101.- Biología. 102.- Veterinaria. 103.- Ecología. 104.- Tecnología.

Resultado de la consulta al banco de datos

Cómo citar el CORPUS

Consulta:

Resultado:

Ilustración 4. Resultados de la búsqueda de la forma “*?i?idí*” en el CREA.

ortografía o pronunciación similar, por si hemos cometido una equivocación al teclear el término deseado. No es posible utilizar comodines en el diccionario académico.

²⁰ Lo mismo ocurre en el resto de diccionarios electrónicos para el español. En ningún caso se ofrecen alternativas válidas (palabras con una forma parecida, bien por su pronunciación o por su grafía), ya que la variación fonético-ortográfica que refleja esta palabra no parece ajustarse a las normas del español, sino a la pronunciación del fonema /d/ como [r] en posición intervocálica en el inglés americano (cf. EDWARDS 2003:92), lo que nos da indicios de que su forma de entrar en el español es por vía oral y no escrita.

²¹ Los resultados de las consultas al CREA y el CORDE son válidos hasta la fecha de redacción de este artículo.

Al recuperar las concordancias de estos 16 casos, descubrimos que, una vez excluidos los no pertinentes (*dividí* como forma del verbo *dividir*), en el CREA hay cuatro casos de la variante *bividí* (ya que ni *bibidí*, la forma que buscamos, ni *bivirí*, la entrada que después encontramos en el *DRAE*, aparecen documentadas):

Concordancias (RAE)

Consulta: */?i?id/, en todos los medios, en CREA*
 Resultados: 16 casos en 13 documentos.

OBTENCIÓN DE EJEMPLOS

Recuperar Concordancias Nomel Clasificación

Agrupación Marcas

Cómo citar el CORPUS Concordancias. Pantalla: 1 de 1. Ver párrafos

Nº	CONCORDANCIA	AÑO
1	ñez, mi adolescencia, ni juventud. Desde entonces dividí el mundo en dos, el poeta y los demás; y elegí	1996
2	mostración para mí no fue convencerlo, por lo que dividí el tiempo en partes iguales y sumé los espaci	1988
3	mi gran obra al-Burhan fi asrar ilm al-mizan: la dividí en cuatro partes y traté en ella de numerosas	1981
4	octubre de 1964. Llevo en prisión cinco años, que dividí en dos etapas. La tercera etapa prefiero exten	2002
5	cosa en cualquier momento. "A partir de entonces dividí mis horas de trabajo entre Juan Bravo, sede de	1992
6	eroso que me dio un conocimiento rápido y gratis. Dividí mi trabajo en dos partes: durante el verano y	1992
7	nos distrajo. La historia de Edmundo me distraía. Dividí mi vida por el denominador de los pasos que él	2001
8	Emploi cu Temps que ahora me hace sonreír, porque dividí las veinticuatro horas del día en tal for	1993
9	licidad del éxito. Aprendí la lección de Polibio. Dividí al ejército en siete partes y puse un camp	1993
10	Hube de comprar un fardo de alfalfa arenada. Lo dividí en cuatro porciones y las coloqué distanta	1993
11	, y recibí a cambio una ración de sire congelado. Dividí mis fuerzas entre volver a cerjería e intentar	1975
12	. La soledad les conveía a mis meditaciones: las dividí entre las motivadas por el culto nostálgico de	1982
13	ta, levantó la cadena y volvió a abrir. Estaba en bividí, calzoncillos y medias negras. Tenía puestos u	1996
14	er el espectáculo de su ex jefe en calzoncillos y bividí. - Jodido, pues, jodido -dijo Zemorano. Cogió	1996
15	tana. Tenía la camisa abierta, desabotonada, y un bividí blanco, lleno de manchas de grasa. - A la diez	1996
16	en el suelo de su oficina, en pantalón, medias y bividí, rodeado de veinte o treinta botellas de vod	1996

Página 160

Cerró la puerta, levantó la cadena y volvió a abrir. Estaba en **bividí**, calzoncillos y medias negras.

AÑO: 1996
 AUTOR: Bayly, Jaime
 TÍTULO: Los últimos días de "La Prensa"
 PAÍS: PERÚ
 TEMA: 07.Novela
 PUBLICACIÓN: Seix Barral (Barcelona), 1996

Ilustración 5. Concordancias correspondientes a la búsqueda de “/?i?id/” en el CREA.

Los cuatro casos relevantes que mostraron las estadísticas aportan la siguiente información: documentan el uso del término en una época determinada, mediados de los noventa, pues pertenecen todos al año 1996²²; lo circunscriben a Perú (lo que concuerda con la marca geográfica de la entrada *bivirí* en el *DRAE*); y, por último, se refieren al ámbito de la ficción, pues todos ellos aparecen en la obra del autor peruano Jaime Bayly. Asimismo, nos proporcionan una aproximación a su significado: por el contexto (concordancias), esta forma parece referirse a una prenda de ropa, posiblemente de ropa interior:

- “En bividí, calzoncillos y medias negras”
- “En calzoncillos y bividí”
- “Un bividí blanco”
- “Pantalón, medias y bividí”

²² Lo que parece dar cuenta de que la incorporación de esta palabra al caudal léxico del español ha debido de ser reciente. Sin embargo, queremos señalar la aportación realizada recientemente por Mercedes Sánchez Sánchez (Jefa de Estudios de la Escuela de Lexicografía Hispánica, Real Academia Española): utilizando el buscador de la Biblioteca Virtual Miguel de Cervantes (<http://www.cervantesvirtual.com/>), localiza el término *bibidí* en la obra *Crónicas diabólicas* del autor mexicano Jorge Ulica, pseudónimo de Julio G. Arce (1870-1926). Estas *Crónicas*, publicadas entre 1916 y 1926, adelantan en unos setenta años la fecha de introducción del término en español; en concreto, a 1924, fecha en que aparece la crónica “¿Do you speak pocho?...” en *El Tucsonense*, Tucson, 18 octubre 1924, p. 5, col. 1-4, lo que reflejaría que los hispanos que habían emigrado a California ya usaban el término en el primer cuarto del siglo XX.

Eso no importa. Lo que yo quiero es que usted me diga qué puedo hacer con la *lanled* del *bordo* donde vivo, que después de rentarme un *jausquipinrrun*, no quiere ni que caliente *guora* porque dice que le *esmoqueo* la *parlor*. Ayer, a la hora del *bricfast*, iba a guisar *jamanegs*, y se levantó de la *carna* furiosa, en *blummers* y *bibidí*, amenazándome con llamar por el *telefon* al *patrol* para que me llevaran a la *yeil*. Yo quise decirle nada a mi compadre Goyo cuando volvió de la *canería* en donde es *boss*, para no *levantar el infierno*, pero si estas cosas *no vienen a un stop*, va a haber *jel*. No puedo seguir comiendo únicamente *jaidogs cofi an donas* y *asicrim* a riesgo de coger una *maladia*. A veces tengo que ir, casi en ayunas, *al otro lado de la bahía* y si no fuera porque *en donde Don Taun* tomo unos *sandwiches*, de *bicon* y otros de *chis* me moriría.

Ilustración 6. Concordancia de “bibidí” en las “Crónicas diabólicas”.

Por otra parte, esta aportación pone de relieve la necesidad de compaginar y complementar el empleo de corpus con otros recursos, electrónicos o no, para evitar caer en la parcialidad.

Con esta información inicial sobre sus variantes formales y su ámbito de uso, extendimos a Google²³ la búsqueda de las dos variantes localizadas, *bividí* y *bibidí*, ya que el significado no se deducía con claridad de los ejemplos del CREA. Hubo que descartar mucho “ruido”, en especial con la forma *bibidí* debido a la existencia de un personaje homónimo del *anime* japonés *Dragon Ball Z*. Pero los datos extraídos del buscador nos permitieron dar con el origen de la palabra, obtener una definición clara del término, constatar su empleo en otros países diferentes de Perú y encontrar nuevas variantes formales.

Por lo que respecta al origen del término, se trata de una palabra formada a partir de una marca registrada, que pertenece a una firma de ropa interior masculina creada en 1876 en Estados Unidos por Bradley, Voorhees y Day, de cuyas iniciales proviene el nombre (BVD):

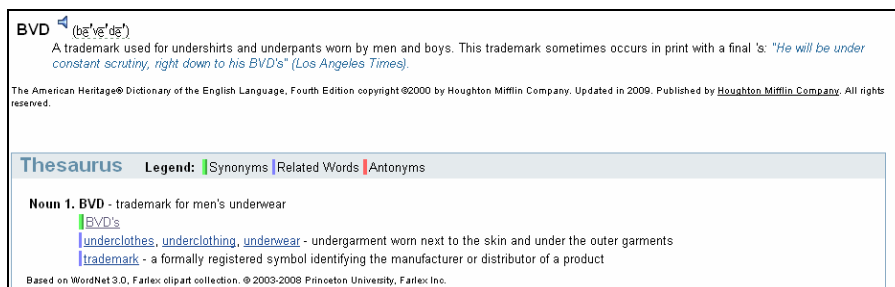


Ilustración 7. Captura de pantalla de la página web de la marca BVD²⁴.

²³ Que nos ofrece, además, la posibilidad de buscar imágenes, lo cual en este caso es particularmente interesante. Los resultados de las búsquedas en Google son válidos hasta la fecha de redacción de este artículo.

²⁴ URL: <http://www.bvd.com/>

Este nombre, *BVD*, es propio del inglés americano, de lo que da fe su inclusión en diversos diccionarios de esta lengua. Véanse dos ejemplos bien conocidos:



BVD (bē've'dēz)
 A trademark used for undershirts and underpants worn by men and boys. This trademark sometimes occurs in print with a final 's': "*He will be under constant scrutiny, right down to his BVD's*" (*Los Angeles Times*).

The American Heritage® Dictionary of the English Language, Fourth Edition copyright ©2000 by Houghton Mifflin Company. Updated in 2009. Published by Houghton Mifflin Company. All rights reserved.

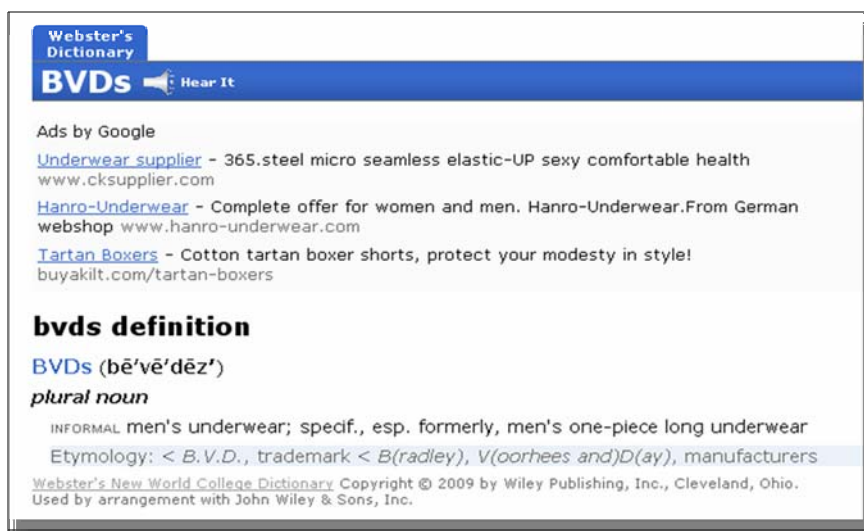
Thesaurus Legend: | Synonyms | Related Words | Antonyms

Noun 1. BVD - trademark for men's underwear

BVD's
 | [underclothes](#), [underclothing](#), [underwear](#) - undergarment worn next to the skin and under the outer garments
 | [trademark](#) - a formally registered symbol identifying the manufacturer or distributor of a product

Based on WordNet 3.0, Farlex clipart collection. © 2003-2008 Princeton University, Farlex Inc.

Ilustración 8. "BVD" en el American Heritage Dictionary of the English Language²⁵.



Webster's Dictionary
BVDs Hear It

Ads by Google
[Underwear supplier](#) - 365.steel micro seamless elastic-UP sexy comfortable health
 www.cksupplier.com
[Hanro-Underwear](#) - Complete offer for women and men. Hanro-Underwear.From German webshop www.hanro-underwear.com
[Tartan Boxers](#) - Cotton tartan boxer shorts, protect your modesty in style!
 buyakilt.com/tartan-boxers

bvds definition
BVDs (bē've'dēz)
plural noun
 INFORMAL men's underwear; specif., esp. formerly, men's one-piece long underwear
 Etymology: < B.V.D., trademark < B(*radley*), V(*oorhees and*)D(*ay*), manufacturers
 Webster's New World College Dictionary Copyright © 2009 by Wiley Publishing, Inc., Cleveland, Ohio.
 Used by arrangement with John Wiley & Sons, Inc.

Ilustración 9. "BVD" en el Webster's Dictionary²⁶.

²⁵ URL: <http://www.thefreedictionary.com/BVD>

²⁶ URL: <http://www.yourdictionary.com/bvds>

Como dato curioso, pero al mismo tiempo representativo de su importancia, llama la atención el hecho de que esta palabra está recogida en WordNet, la conocida base de datos léxica del inglés:

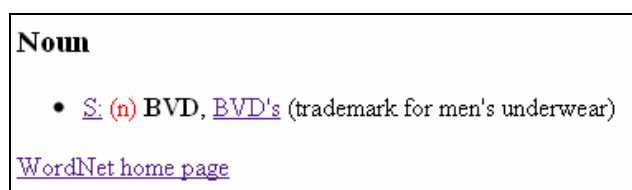


Ilustración 10. “BVD” en WordNet²⁷.

El empleo del término en español documenta este origen, como el siguiente ejemplo, extraído de una página web²⁸ que recoge *bvd* como una variante de *bibidi*:

Al Sr. Enrique “No paro de meter cizana” Martínez: seguramente Frank debería encontrar una foto del recuerdo tuya de los años 1988-1989 para comprobar que tus prendas favoritas eran: *andar en bvd* (o mejor conocidas bajo *bibidi*) y en jeans como los que turqueaba Vanilla Ice.

Cuando buscamos en Google definiciones en español de alguno de estos términos, las dos primeras apariciones son de WordReference, un conocido sitio de referencia para diccionarios de traducción: en concreto, el término *bividi* aparece en el diccionario español-inglés (que incluye la marca de uso Perú antes de proporcionar las traducciones al inglés, ¡de un término

²⁷ URL: <http://wordnet.princeton.edu/>. Acceso a la consulta en línea a través de la URL: <http://wordnetweb.princeton.edu/perl/webwn>.

²⁸ Datos aportados por la estudiante Nuria Martínez. En la actualidad, la mayoría de estos textos no se pueden recuperar, debido a que estaban extraídos de *blogs* o páginas personales cuyos enlaces han caducado, uno de los inconvenientes que presenta el uso de Internet como banco de datos: la vida efímera de muchas de las páginas, sobre todo las de índole personal, aunque sean las que más proliferan en la actualidad.

inglés!) y en un comentario de su foro, donde un hablante pregunta por el significado de la palabra inglesa *BVD*²⁹.



Ilustración 11. “BVD” en WordReference³⁰.

Los datos obtenidos a través de Google también nos ofrecieron muestras de uso del término que sugieren una ampliación de su definición respecto a la que ofrece el *DRAE*, ya que se emplea no solo como prenda interior, y la utilizan tanto hombres como mujeres. Como botón de muestra, véase el resultado de buscar “*bibidí*” en “Imágenes” de Google. Muchas de las imágenes pertenecen a firmas de venta de ropa que describen las características de sus productos (camisetas interiores, camisetas deportivas, camisetas de vestir...):



²⁹ Los resultados son los mismos tanto si introducimos en el buscador *bibidí* como *bibidí*.

³⁰ URL: <http://www.wordreference.com>.

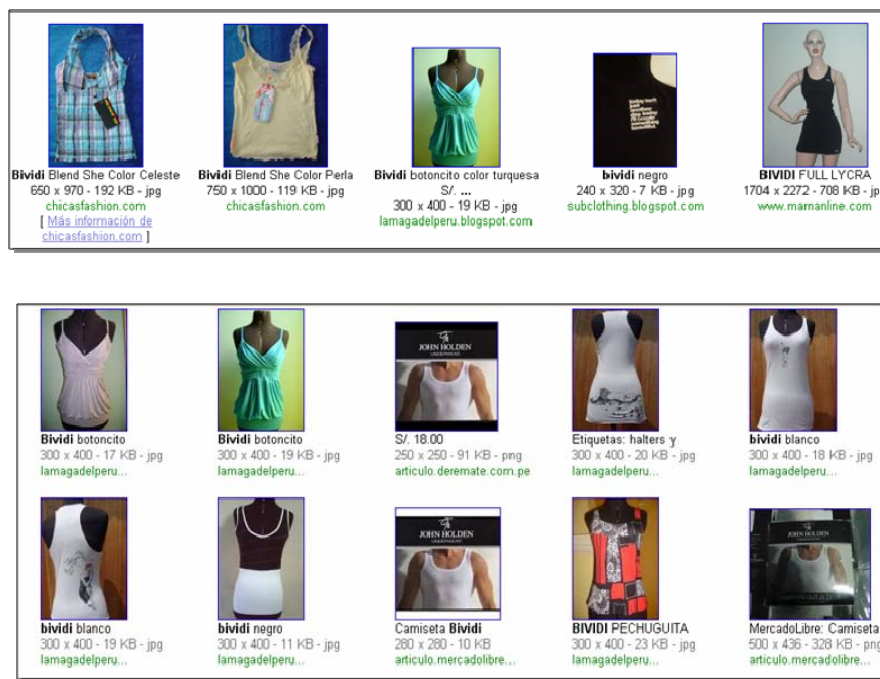


Ilustración 12. Algunos resultados de la búsqueda de “bividi” en Google Imágenes³¹.

Por otra parte, ya en su propio empleo en la variedad de inglés americano las siglas se han generalizado como un nombre común³², empleo que se ha transferido al español, en concreto, al español hablado en Perú:

En el español del Perú muchas marcas extranjeras ya registradas en el mercado internacional han llegado a hacerse nombre comunes: [...] **bividi** (B.V.D.), cuya

³¹ URL: <http://images.google.es>

³² De acuerdo con la respuesta que proporciona el *Daily Mail* (Londres) a una consulta de un lector a propósito del origen de este término que ha oído en algunas películas ambientadas en el Oeste (*QUESTION In some westerns, an old man is seen capering in a suit of long underwear that he calls his 'BVDs' or 'bivvy dees'. Where does this name come from?*), a principios del siglo XX la marca BVD dominaba de tal manera el mercado que el término pasó de designar un producto específico de ropa interior a ser el genérico para cualquier prenda de este tipo. Incluso, ya se había perdido entonces la conciencia de su origen en unas siglas (cf. URL: <http://www.thefreedictionary.com/BVD>, sección *References in periodicals archive*).

variante popular es *bivirí* atribuida a pronunciación de japoneses, frecuentemente empleados o dueños de lavanderías... (Hildebrandt 1994:155-158, *apud* Universidad de Piura PLANCAD 2001).

J. Calvo Pérez también comenta esta tendencia del español hablado en Perú a la generalización de marcas comerciales, citando la misma fuente: “En el español del Perú, existen diversas marcas ya registradas en el mercado internacional, que se han convertido en nombres comunes. Hildebrandt (1994, s.v. *cuáquer*), anota las siguientes”, y recoge una enumeración de diferentes grupos (*id.*): “las perfectamente integradas”, “las muy difundidas”, “las usuales”, “las anticuadas”, “las modernas” y “las no totalmente integradas”. *Bividí* está integrada en el grupo de las usuales y lleva, además, una nota aclaratoria sobre su inexistencia en España. Más adelante (*ibid.*:42), J. Calvo vuelve a incluir el término objeto de análisis en una lista de entradas “en que la marca se ha tomado como generalizador del producto, perdiendo su especificidad y ganando en universalidad”. Se indica adicionalmente que es un peruanismo³³.

A continuación, incluimos un fragmento del testimonio –que ya no se encuentra en la web– de un hablante de Lima en el que compara la palabra estudiada con la variante que se encuentra en otra zona del español (Santa Cruz de la Sierra, Bolivia), con plena conciencia de las diferencias léxicas entre su variedad del español y la de este lugar de Bolivia, por otra parte, país limítrofe con Perú:

Hay otros nombres también de cosas diferentes, de las que me acuerdo como achujcha es caigua, churikis es molleja, no se dice gaseosa sino soda, polera es polo, **bibidí** es camiseta, casaca es campera. Despues voy a apuntar más nombres.

³³ Es interesante la entrada porque recoge todas las variantes que hemos localizado del término: *bividí*, *bibidí*, *bivirí*, *bibirí* y *BVD*.

BIVIRÍ (bividí) ¶ (m.c.: BVD) [pr.] (camiseta {interior}), *bibirí* (*bibidí*).

Resulta curiosa la evolución de la lengua, que utiliza las iniciales de un nombre propio para formar las siglas de una marca comercial que, con el tiempo, se convierten en nombre común, tanto que puede llegar a ser el “apodo-apellido” de alguien y, así, volver a la esfera del nombre propio, como en la historia sobre Pedro Vivirí³⁴ que encontramos en el diario oficial del departamento de Cajamarca, en el norte de Perú:

Pedro era un hombre de cierta edad, esa que se vuelve indefinida después que uno pasa los 40. Él vivía en Hualgayoc, un pueblo no muy lejos de aquí, Hualgayoc es un pueblo que amanecía frío y anocheecía igual. Sus calles sin luz, apenas eran iluminadas por las luces tenues de lámparas de camiseta, llamadas Petromax, porque esa era su marca registrada y nadie decía una lámpara a kerosene, simplemente se decía Petromax. La marca se había convertido en un nombre, como sucede a veces con algunos apodos.

Pedro tenía una cantina y vendía aguardiente en las mañanas frías, en las tardes heladas y en las noches gélidas, sin embargo él tenía una rara afición, usaba esas (sospecho que solo era una) camisetas llamadas *bivirí*. Pedro usaba esas camisetas que alguna vez aparecieron en Norteamérica y se registraron bajo la marca *BVD* (*bividi*, según su pronunciación), marca registrada, como el frío lo era a Hualgayoc el nombre de ese tipo de camisetas se peruanizó en *bivirí*.

Camiseta interior masculina (como la denomina la Real Academia de la Lengua Española y le antepone Perú) se convirtió en la prenda favorita de Pedro. Cuando por las noches los parroquianos bebían sendas copas del licor en su cantina, Pedro atendía como si fuese un cantinero del oeste, sin piano, claro; sin chicas con faldas de encajes; sin pistoleros, pero sí con una pléyade de mineros que gastaban su semana y Pedro los atendía en *bivirí*.

Era una rareza, un hombre atendiendo en ese frío casi infinito llevando puesto una camiseta sin mangas, sin temblar siquiera. Los hualgayoquinos, dados a poner sobrenombres no tardaron en llamarlo “*Pedro bivirí*” y todos olvidaron su apellido, él mismo empezó a olvidarlo y muchas veces decía “Soy *Pedro bivirí*” como quien dice soy “Juan casaca” o “Luis Pantalón”. Otros olvidaron hasta su nombre y simplemente lo llamaban “*El bivirí*”.

Entonces, el pueblo asumió a *bivirí* como un nuevo apellido y su esposa era la señora de *bivirí* y sus hijos eran los *bivirís* y la calle era del *bivirí* y el pueblo era el pueblo del *bivirí*.

Pero la edad no pasa en vano y Pedro empezó a sufrir de achaques primero, ya no podía ponerse su *bivirí* de antaño y se metía bajo un poncho inmenso que lo abrigaba. La edad no perdona. Nadie sabe por qué cierto día empezó a perder la vista y se quedó ciego totalmente, ya no era ni la sombra de aquel cantinero rudo que atendía en la madrugada puesto el *bivirí* americano.

³⁴ URL: <http://balconinterior.blogspot.com/>

Las viejas del pueblo entonces empezaron a especular, decían que el frío intenso producía ceguera, otras decían que era el *bivirí*, otras creían que vender cañazo causaba la atroz ceguera.

Pedro Bivirí murió una tarde cuando otra generación había tomado el pueblo, cuando casi nadie ya lo conocía porque él no podía ver a nadie y los demás no lo veían porque no querían ver a un viejo ciego del que nada sabían. Y hoy que es abril y han pasado años de tu muerte me acuerdo de ti, Pedrito, de tu camiseta con marca registrada y sin mangas, de tu última y vacía mirada, literalmente vacía como esta soledad que hoy me ha hecho recordarte.

Tomado de “Hualgayoc, historia y Tragedia de un pueblo minero” (JAP)

Por lo tanto, parece que hay pruebas suficientes de que el término objeto de estudio, procedente del inglés (*BVD*), se introdujo en español por vía oral a través de Perú (peruanismo), mediante un proceso de adaptación de la pronunciación inglesa (*bividí-bibidí*) a la local (bien por “peruanización” de aquella o por deformación popular atribuida a los japoneses, de ahí *bivirí*). En este sentido, es interesante el siguiente comentario de un usuario que mantiene un blog sobre jerga peruana³⁵, puesto que, además de redundar en el origen de la palabra, registra la variante formal *bivirí*, que, por otra parte, sabemos que es la recogida en el *DRAE-01*:

bivirí. s. prenda de ropa interior masculina consistente en una camiseta escotada, sin mangas y sin cuello. Esta palabra viene de *BVD*, siglas de una marca registrada estadounidense de ropa interior para hombres, que se generalizó en el Perú en la década del cincuenta. En inglés estas éstas se leen */bividí/*, lo que se ha adaptado como */bibirí/* en la pronunciación local.

Sin entrar a discutir las razones por las que se ha producido esta adaptación fónica, algunos de los textos encontrados dan muestras de una conciencia del hablante sobre el carácter normativo de la forma *bivirí*, pese a no ser la más frecuente en el uso, como comentaremos a continuación:

Toquemos la cuestion fashion, agg que palabra más rosquetona. Dime que no has sufrid@ como cojud@ para bajar de peso y te entre esa blusita o puedas tú compadre, andar hecho un imbécil con tu *bivirí* (de *BVD*, así se escribe. no quiero responder a bestias que me corrigen sin saber) por la calle, ridículo de mierda³⁶.

³⁵ URL: <http://cultureando.blogspot.com/>. El término se encuentra también en otros sitios web que lo mencionan entre los peruanismos y americanismos del *DRAE*.

³⁶ Tomado del blog personal Memorias del Olvidado, de un peruano. URL: blog: <http://mystic-place.blogspot.com/>

RESPONDERLE A ESTOS TRICICLEROS, ES ENSUCIARSE LA BOCA. FELIZMENTE QUE SON DEMASIADO BESTIAS PARA GANAR, YA LES DIJE QUE LES DEN GRACIAS A SUS AMOS ESPAÑOLES QUE LOS DEJARON BAILANDO YUNSA Y COMIENDO QUESO CON LLAMA, JAJAJAJA, PORQUE DE HABER SIDO ALEMANES O MIS ANTEPASADOS GALESES LA MAYORIA NO EXISTIRIA. PERO LO QUE MAS ME CALIENTA ES QUE SE EXPRESAN COMO SABELOTODOS RIDICULOS HIJOS DE PUTA (“ERUROPA”, “PAGINA WED”, “NGNORANTE”, “ALOGADO” “OFENZA” “ANTISIPACION”, SEGURO QUE TAMBIEN DICEN “**BIBIDI**” JAJAJAJA TREMENDOS INDIASOS COQUEROS)³⁷.

Por lo que se puede observar, parece que hay hablantes para los que *bivirí* es la variante escrita que se debe usar, porque esa es la normativa, lo cual no quiere decir que sea la más usada. Además, como se colige del segundo texto, parece ser que para algunos hablantes *bibidí* es usada por personas occidentalizas o cholos y que, por lo tanto, dicho término no debe emplearse. Sin embargo, los datos de uso de nuevo contradicen esta censura, ya que la forma *bibidí* aparece ampliamente documentada.

Si indagamos un poco más con Google, pronto encontramos muestras procedentes de otras geografías (España, Argentina, Estados Unidos...), con un papel destacado para Ecuador, expansión que parece lógica dada la proximidad de este último país a Perú. De hecho, la entrada de la Wikipedia³⁸ inglesa para *BVD* alude a su uso en ambos lugares: “In Ecuadorian and Peruvian Spanish, the term *bividi*, pronounced like the English initials, is an eponym for a man’s sleeveless underwear T-shirt”.

En el momento de realizar la búsqueda³⁹, se encontraron ejemplos no dudosos en Ecuador y Argentina, y en otros lugares (México, Chile) donde resultaba más difícil la adscripción geográfica. En cualquier caso, parece claro que el territorio en que se emplea el término se ha ampliado. Sin

³⁷ Del Centro de Medios independientes Perú. URL: <http://peru.indymedia.org/>

³⁸ URL: <http://en.wikipedia.org/wiki/BVD>

³⁹ El origen se determinó bien a partir de la extensión de las páginas web (.pe), bien a partir de información explícita sobre el país que aparecía en los sitios de Internet, aunque lógicamente no tenemos certeza de la procedencia del autor del texto.

embargo, es cierto que Perú es el país que arroja más casos, lo que parece volver a confirmar este como país de entrada del término en español. Además, en Perú conviven *bibidí*, *bivirí*, *bividí* e, incluso, *bvd*.

En una rápida comparación de los resultados obtenidos por región en Google⁴⁰, destaca el empleo del término en Perú, Estados Unidos y

⁴⁰ Utilizando la “Búsqueda avanzada”, opción “Región”, que busca solo páginas ubicadas en la región especificada. A continuación resumimos los datos:

	Argentina	Bolivia	Chile	Colombia	Costa Rica
bibidí	434 (ruido)	2 (ruido)	268 (ruido)	56	2 (ruido)
bividí	336	1	5	1	0
	Cuba	Ecuador	El Salvador	España	EE.UU.
bibidí	0	452	3 (ruido)	1730 (ruido)	20400 (ruido)
bividí	0	148	0	1300	3600
	Honduras	México	Nicaragua	Panamá	Paraguay
bibidí	1 (ruido)	208 (ruido)	1 (ruido)	85 (ruido)	6 (ruido)
bividí	0	9	0	4	2
	Perú	Puerto Rico	R. Dominic.	Uruguay	Venezuela
bibidí	457	2 (ruido)	5 (ruido)	132 (ruido)	115 (ruido)
bividí	6030	0	0	0	4

Tabla 1. Datos para la búsqueda de “bibidí” y “bividí”.

España, por orden de casos, lugares donde además se registran las diferentes variantes formales, como se observa en el gráfico que sigue:

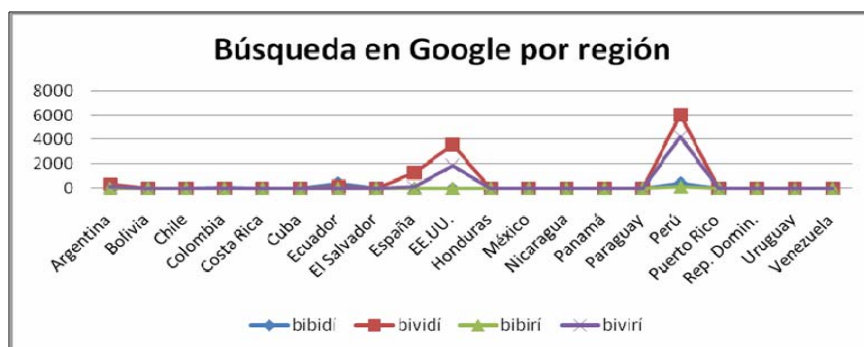


Ilustración 13. Resultados de la búsqueda por región en Google de “bibidí”, “bividí”, “bibirí” y “bivirí”.

En segundo lugar, hay que comentar que, aunque inicialmente las variantes *bibidí* y *bibirí* arrojaban un número más elevado de casos –sobre

	Argentina	Bolivia	Chile	Colombia	Costa Rica
bibirí	80 (ruido)	0	3 (ruido)	4 (ruido)	0
bivirí	151	0	1	0	0
	Cuba	Ecuador	El Salvador	España	EE.UU.
bibirí	1 (ruido)	0	0	118 (ruido)	5050 (ruido)
bivirí	1	0	0	118	1860
	Honduras	México	Nicaragua	Panamá	Paraguay
bibirí	0	614 (ruido)	0	6 (ruido)	1 (ruido)
bivirí	0	7	0	2 (ruido)	0
	Perú	Puerto Rico	R. Dominic.	Uruguay	Venezuela
bibirí	114	1 (ruido)	0	0	1 (ruido)
bivirí	4220	0	0	2	4

Tabla 2. Datos para la búsqueda de “bibirí” y “bivirí”.

todo la primera forma— que *bividí* y *bivirí*, hubo que descartar mucho ruido. En un análisis más profundo de dichos resultados, se detectó que los primeros (los de *bibidí* y *bibirí*) no se correspondían con el empleo del término para designar una camiseta —con las excepciones notables de Perú (457 casos), Ecuador (452 casos) y Colombia (56 casos)—. De estos tres países, en Ecuador (452-148 casos) y Colombia (56-1 casos) la variante *bibidí*, de hecho, supera a *bividí*⁴¹.

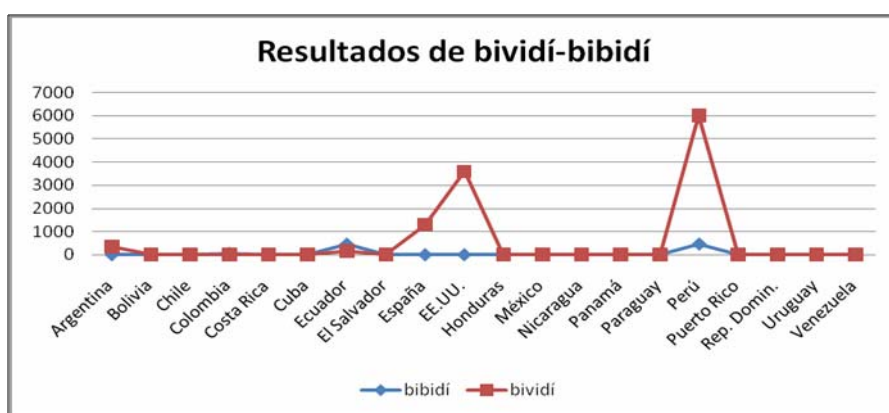


Ilustración 14. Comparación de los resultados de “*bibidí*” y “*bividí*”.

Por lo tanto, la forma *bividí*, la que más se ajusta a la pronunciación original inglesa (de *BVD*), es la que cuenta con más casos, pese a no ser la forma normativa recogida en el *DRAE*⁴². Perú (6030 casos) encabeza la

⁴¹ El tamaño reducido de la escala del gráfico impide apreciar claramente el caso de Colombia.

⁴² Recordemos que es *bivirí*. Este hecho resulta doblemente llamativo: por un lado, no es la forma más empleada, tal y como se desprende de los datos aportados —y ni siquiera la documentada en los bancos de datos académicos (que es precisamente *bividí*)—; por otro lado, contradice la doctrina académica a propósito de términos que siguen un patrón muy similar, como *DVD* o *CD*. Si bien para estas siglas, a partir de las cuales se han creado también nombres comunes, la forma de entrada en español parece ser la lengua escrita, en ambas se documentan pronunciaciones “a la inglesa” ([*dividí*], [*sídí*]) en América, que la RAE, a través del

lista de registros, seguida de Estados Unidos (3600 casos), España (1300 casos), Argentina (336 casos) y Ecuador (148 casos). La propagación del término parece obedecer a una doble tendencia: por una parte, hacia el norte, de la zona andina al Caribe continental (de Perú a Ecuador y Colombia) –en estos dos países con predominio, además, de la variante *bibidí* frente a la forma más extendida en el resto de lugares, *bividí*–; y, por otra, hacia lugares tan distantes como Estados Unidos, España y Argentina, en los que la proximidad geográfica no puede ser el criterio, sino la emigración u otros factores relacionados con las nuevas tecnologías (como la ubicación de servidores).

La forma normativa *bivirí* sigue el mismo patrón de distribución de los datos que *bividí*, aunque con un recuento menor de muestras en todos los países: en primer lugar, destaca de nuevo Perú (4220 casos), seguido de Estados Unidos (1860 casos), Argentina (151 casos) y España (118 casos). Además, Perú es el único país donde alterna con *bibirí* (114 casos), aunque a una distancia considerable de la otra variable. En el resto de países ha habido que descartar las apariciones de *bibirí* por no ser pertinentes para la búsqueda.

Diccionario panhispánico de dudas, desaconseja, igual que las grafías que se corresponden con sus lecturas inglesas (*dividí* y *cidí* respectivamente):

DVD. [...] En español debe leerse [deubedé] o [debedé], dependiendo del nombre con que se denomine la letra *v* (→ *v*, 1); se desaconseja la pronunciación [dividí], propia del inglés, a pesar de su extensión en algunas zonas de América. [...] A partir de la lectura española de la sigla, se han creado los sustantivos *devedé* (pl. *devedés*), en América, y *deuvedé* (pl. *deuvedés*), en España: «En los *devedés*, memorables representaciones operísticas» (Glantz Rastro [Méx. 2002]); «Yo podría haber vivido con Eduardo comprando *deuvedés*» (Gopegui Lado [Esp. 2004] 133); se desaconseja la forma [dividí], por corresponder a la lectura inglesa de la sigla.

CD. [...] En español debe leerse [sedé, zedé]; se desaconseja la pronunciación [sidí], propia del inglés, a pesar de su extensión en algunas zonas de América. [...] A partir de la lectura española de la sigla se ha creado el sustantivo *cedé* (pl. *cedés*): «En las tiendas ya se vendían *cedés* con canciones sobre el tema» (PzReverte Reina [Esp. 2002]). Se desaconseja la forma [cidí], por corresponder a la lectura inglesa de la sigla. En cualquier caso, se recomienda usar con preferencia el equivalente español (*disco*) *compacto*.

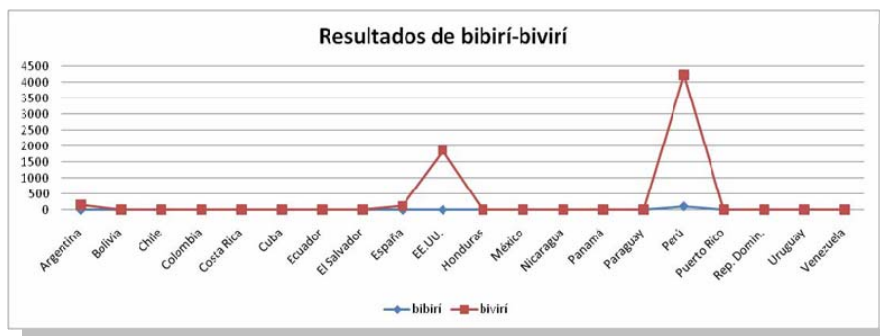


Ilustración 15. Comparación de los resultados de “bibirí” y “bivirí”.

Por lo tanto, de forma global, se observa que las dos variantes más frecuentes, *bividí* y *bivirí*, se encuentran concentradas en los mismos lugares –excepto en Ecuador, donde no hay casos de la segunda– con una marcada supremacía de la primera.

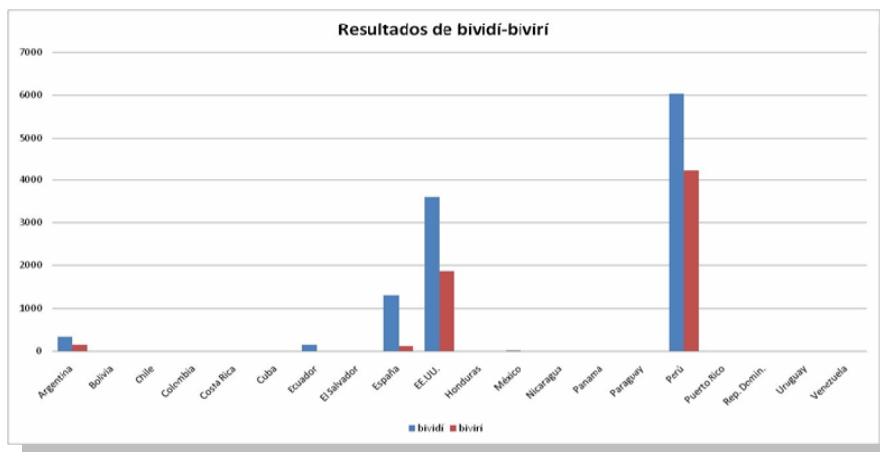


Ilustración 16. Comparación de los resultados de “bividí” y “bivirí”.

3. Conclusiones

En resumen, con este pequeño ejemplo, hemos querido demostrar que las posibilidades que ofrecen para la investigación los corpus son apasionantes.

Los inconvenientes de utilizar la web como corpus se pueden resumir en los siguientes puntos:

- 1) Vida efímera de los textos en la web: muchas páginas desaparecen en un lapso relativamente breve de tiempo, por lo que ni siquiera pueden ser citadas.
- 2) Caducidad de los enlaces, sobre todo los de las páginas de índole personal (*blogs*), que son las que más proliferación tienen en la actualidad.
- 3) Falta de información sobre los textos: autores anónimos, ausencia de datación, etc., factores todos ellos que no permiten identificar con certeza la procedencia de las páginas.
- 4) No aval institucional, lo que puede disminuir la relevancia de los datos obtenidos y arrojar dudas sobre su fiabilidad, al no estar respaldados por ningún tipo de autoridad.
- 5) Ausencia de planificación y control sobre la selección de los textos (páginas), que no obedecen a unos criterios lingüísticos previamente establecidos, por lo que no existe garantía de que sean representativos de la lengua objeto de estudio.
- 6) Imposibilidad de determinar la frecuencia relativa de una forma dado que el número total de formas que integran Internet es desconocido, con las consecuencias que esto tiene para validar ciertos resultados.

Sin embargo, no todo son obstáculos; el uso de la web como corpus implica ventajas, en especial en lo que se refiere al número ingente de muestras reales de lengua que proporciona y a las diferentes posibilidades para realizar búsquedas que ofrece: por región, mediante imágenes, por dominio (término *site:pe*, por ejemplo), en noticias, en vídeos, en libros, en definiciones (define: término), con comodines, etc.

REFERENCIAS BIBLIOGRÁFICAS

- EDWARDS, H. T. (2003): *Applied Phonetics: The sounds of American English*, NT: Delmar Thomson Learning, 3ª ed.
- FRANCIS, W. N. (1992): "Language Corpora B.C.", en J. SVARTVIK (ed.), *Directions in Linguistics: Proceedings of Nobel Symposium 82 (Stockholm, 4-8 August 1991)*, Berlin/New York: Mouton de Gruyter, 17-32.
- MARTÍ ANTONÍN, Mª. A. y CASTELLÓN MASALLES, I. (2000): *Lingüística computacional*, Barcelona: Universitat de Barcelona.
- MCENERY, T. (2003): "Corpus Linguistics", en R. MITKOV (ed.), *The Oxford Handbook of Computational Linguistics*, Oxford: Oxford University Press, 448-463.
- MCENERY, T. y WILSON, A. (1996): *Corpus Linguistics*, Edinburgh: Edinburgh University Press. Suplemento web: <http://www.lancs.ac.uk/fss/courses/ling/corpus/>
- MCENERY, T. y WILSON, A. (2001): *Corpus Linguistics*, Edinburgh: Edinburgh University Press, 2ª ed.
- MORALA, J. R. (2002): "Nuevas tecnologías y recursos lexicográficos: *fuereño*", en G. CLAVERÍA (coord.), *Filología en Internet*, Bellaterra, Barcelona: Servei de Publicacions, Universitat Autònoma de Barcelona, 45-53.
- REAL ACADEMIA ESPAÑOLA (2001): *Diccionario de la lengua española*, Madrid: Espasa, 22ª edición. URL: <http://www.rae.es>
- REAL ACADEMIA ESPAÑOLA (2005): *Diccionario panhispánico de dudas*, Madrid: Santillana. URL: <http://www.rae.es>
- SINCLAIR, J. (1994): *EAGLES*, Document EAG-CWG-IR-2.
- SINCLAIR, J. (1996): *EAGLES Preliminary recommendations on Corpus Typology*. Documento electrónico disponible en URL: <http://www.ilc.cnr.it/EAGLES96/corpusstyp/corpusstyp.html>
- TORRUELLA, J. y LLISTERRI, J. (1999): "Diseño de corpus textuales y orales", en J. M. Bleca et al. (eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos*, Barcelona: Milenio y Universidad Autónoma de Barcelona, Dpto. de Filología Española, 45-77. Disponible electrónicamente en .pdf: http://liceu.uab.es/~joaquim/publicacions/Torruella_Llisterri_99.pdf
- Universidad de Piura PLANCAD, Facultad de Ciencias de la Educación (2001): *Fascículo autoinstructivo. Abordar la realidad lingüística*, Lima: Ministerio de Educación/DINFOCAD/UCAD/PLANCAD. Disponible versión electrónica en URL: http://ciberdocencia.gob.pe/archivos/fasciculo_Comunicacion_abordar_la_realidad.doc