

Presentación de COLA, un corpus oral de lenguaje adolescente en línea

Resumen

El corpus COLA, Corpus Oral de Lenguaje Adolescente, une el texto y el sonido de las conversaciones grabadas, de modo que se puede escuchar la conversación transcrita y leerla a la vez y se halla en línea: www.colam.org. Presentamos los pasos dados para construir el Corpus COLA de la Universidad de Bergen, Noruega, así como los diferentes retos que este mismo trabajo ha supuesto.

Palabras clave: lenguaje adolescente, recolección de datos, construcción de corpus oral, transcripción, anonimización

Abstract

In the COLA-corpus, *Corpus Oral de Lenguaje Adolescente*, the sound and text are put together so that it is possible to read the transcribed text and listen to the sound at the same time. The text and sound are online: www.colam.org. We present the steps taken to build the COLA corpus of the Bergen University, Norway, as well as the different challenges arisen from this work.

Keywords: teenage language, recollection of data, oral corpus building, transcription, anonymization

1. Introducción

Este trabajo describe los pasos dados para construir el Corpus COLA, Corpus Oral de Lenguaje Adolescente de la Universidad de Bergen, Noruega,¹ así como los diferentes retos que este mismo trabajo ha supuesto al unir el texto y el sonido del material grabado, de modo que se puede escuchar la conversación transcrita y leerla a la vez en línea. Asimismo presentamos algunos aspectos de su uso actual en la investigación y en la enseñanza. Este trabajo consta de tres partes: la presentación del corpus y sus usos, la recolección de datos y la transcripción con la anonimización del material.

2. El corpus COLA

El Corpus COLA, Corpus Oral de Lenguaje Adolescente, contiene el habla juvenil y conversaciones espontáneas e informales, recogidas en Madrid (COLAM), Santiago de Chile (COLAS) y Buenos Aires (COLABA). Abarcando un total de 700.000 palabras, cuenta con 500.000 de Madrid, 100.000 de Santiago de Chile y otras 100.000 de Buenos Aires. Dicho corpus se encuentra en línea en un interfaz www.colam.org, donde se recogen además distintos artículos e información sobre este corpus en especial y el lenguaje juvenil en general.

Las capitales y metrópolis de Madrid, Buenos Aires y Santiago de Chile han sido elegidas para la recopilación de datos por su relevancia, influencia y prestigio sobre la lengua del resto del país.² Por tratarse de capitales de países constituyen además entidades comparables en los estudios contrastivos.

Dicha recolección de datos se ha obtenido por medio de la grabación de conversaciones espontáneas de los jóvenes, realizadas por ellos mismos, sin presencia de persona adulta. Las conversaciones resultantes han sido trasladadas a formato de audio digital WAV o MP3 y posteriormente transcritas con el Programa Transcriber, de tal manera que se puede leer la transcripción y escuchar el sonido

¹ <http://www.colam.org>

² Annette Myre Jørgensen y Eli-Marie Drange, "La lengua juvenil de las metrópolis Madrid y Santiago de Chile", *Arena Romanística*, 9, 2012, 74–96.

conjuntamente. En esto radica la novedad del Corpus COLA.

Los archivos de texto de las conversaciones están en formato HTML y enlazados a los archivos de sonido WAV por el programa Corpus Workbench y colgados en el interfaz de www.colam.org.³ En dichas transcripciones se respetan las reglas ortográficas, independientemente de la pronunciación de las palabras, aunque la palabra también aparece escrita tal cual suena, resaltando un rasgo juvenil, por ejemplo: {OscarlOjcar}. El tamaño del corpus, así como su organización y diseño permite estudiar los diferentes elementos del lenguaje juvenil.

Mediante la aplicación de los programas informáticos KWIC⁴ y WORKBENCH⁵ a los archivos transcritos se pueden realizar búsquedas de una, dos y tres palabras consecutivas, la frecuencia de su uso, la edad, el sexo y el nivel social del/de la que habla.

COLA es un recurso que ofrece la posibilidad a lingüistas y a otros interesados en el lenguaje juvenil/oral de investigar el estilo comunicativo, la interacción, el marcaje del discurso con silencios, solapamientos, vocativos típicamente juveniles *tíaltío*, *troncal/tronco*, así como sus frecuentes marcadores del discurso⁶ *en plan*, *como*, etc. Permite también realizar estudios del discurso y el léxico de los jóvenes de habla castellana, haciendo distintas comparaciones entre los jóvenes de Madrid, Buenos Aires y Santiago de Chile.⁷ Asimismo es posible realizar estudios comparativos con otros corpus de lenguaje juvenil, como el inglés COLT, Corpus of London Teen-

³ Knut Hofland et al., “Cola: A Spanish Spoken Corpus of Youth Language”, *Proceedings from the Corpus Linguistics Conference Series*, Birmingham: Centre for Corpus Research, University of Birmingham, 2005.

⁴ <http://kwic-concordance.software.informer.com/5.0/>

⁵ <http://www.mysql.com/products/workbench/>

⁶ Annette Myre Jørgensen and Juan Antonio Martínez, “Vocatives and Phatic Communion in Spanish Teenage Talk”, ed. Jens Normann Jørgensen, *Love Ya Hate Ya-the Sociolinguistic Study of Youth Language and Youth Identities*, Cambridge: Cambridge Scholars Publishing, 2010, pp. 193–209.

⁷ Annette Myre Jørgensen, “Los Vocativos en el Lenguaje Juvenil de Santiago de Chile y Madrid”, ed. Carmen García and María Elena Placencia, *Estudios de Variación Pragmática en Español*, Buenos Aires: Dunken, 2011, pp. 141-166

ge Language, y el noruego UNO, Ungdomsspråk i Norden.⁸

Se describen en lo que sigue los pasos dados en la construcción del corpus como la recolección de datos: entrevistas en los centros escolares, la transcripción del lenguaje del material grabado así como el tratamiento de la anonimización obligada por la última legislación europea y noruega.⁹

En dicha realización hemos ido solventando diferentes escollos, como obtener los permisos para poder ser grabados, la irresponsabilidad de la no devolución del material de grabación, la inutilidad de las grabaciones o robos y pérdidas. La principal dificultad a la hora de transcribir radicó en entender sus sonidos y casi adivinar sus palabras, acentuándose el problema en los solapamientos de dos o más hablantes. Cuando ríen, gritan, comen, hablan con la boca llena, oyen música o hay ruidos de fondo como ladridos de perros, sus motos, tráfico, tecleo, avisos o alarmas, vídeojuegos, este trabajo resulta bastante arduo y poco sereno. De hecho, son reconocidos los casos de personas que sienten que van perdiendo su paciencia e incluso en ocasiones la cordura realizando este trabajo. En la descripción de estos ruidos así como de los distintos sonidos humanos estornudos, toses, carraspeos, sorbidos, o ruidos de boca en general hemos evolucionado desde escribirlo con su nombre hasta la forma actual, eliminarlo. Dado que se trata de un corpus que se oye a la par que se lee, dejar que el sonido hable por sí solo.

Para el tratamiento de la anonimización de datos, obligada por ley, que pudieran identificar a los informantes ha sido necesario aprender el manejo de programas adecuados.

3. Justificación del Corpus Cola

3.1 El lenguaje juvenil

Briz define el lenguaje juvenil o adolescente como “la interacción

⁸ Anna-Brita Stenström and Annette Myre Jørgensen, “La Pragmática Contrastiva Basada En El Análisis de Corpus”, ed. Lars Fant y Ana María Harvey, *El diálogo oral en el mundo hispanohablante*, Frankfurt am Main: Vervuert-Iberoamericana, 2011, pp. 251–276

⁹ Presentada en: <http://www.nsd.no>

coloquial de, o entre los jóvenes.”¹⁰ Es un lenguaje específico, que parte del lenguaje estándar, si bien se diferencia de éste: “No puede negarse la evidencia de que el habla de los jóvenes, en ciertas modalidades de la comunicación contrasta vivamente con la de los mayores.”¹¹ Este lenguaje, cuando se desarrolla entre jóvenes de 13 a 19 años¹² en una situación informal, adquiere una serie de rasgos particulares, un estilo comunicativo especial que lo convierte en un interesantísimo objeto de estudio. Hablantes de diferentes variedades del español emplean diversos estilos comunicativos en interacciones similares y ese estilo comunicativo particular, ese *tatuaje*, es lo que ha llamado la atención de los investigadores, desde Platón hasta nuestros días. La razón principal de este interés probablemente se deba a que “los jóvenes son grandes propulsores de cambios lingüísticos, del papel de la lengua en el intercambio social”, como dice Eckert.¹³ Esta influencia es corroborada por varios investigadores del lenguaje juvenil como Briz, Zimmermann y Stenström.

3.2 ¿Por qué estudiar el lenguaje juvenil?

En las últimas décadas la juventud ha ido constituyendo un grupo social cada vez más fuerte, que goza de prestigio y que marca ideales para el resto de la sociedad. Como dice Zimmermann¹⁴ “[...] hay cada vez más adultos que pretenden ‘rejuvenecerse’ a través del uso de expresiones tomadas del lenguaje juvenil.” A los jóvenes se les ha estudiado desde muchos ángulos, pero poco se ha hecho desde el lingüístico.¹⁵

¹⁰ “La interacción entre jóvenes. Español coloquial, argot y lenguaje juvenil”, ed. María Teresa Echenique Elizondo y José Pedro Sánchez Méndez, *Lexicografía y lexicología en Europa y América: Homenaje a Günther Hensch*, Madrid: Gredos, 2003, pp. 141–154, aquí p. 142.

¹¹ Ibid.

¹² Definición de la adolescencia. http://www.unicef.org/honduras/Estado_mundial_infancia_2011.pdf

¹³ Penelope Eckert, “Adolescent Social Structure and the Spread of Linguistic Change”, *Language in Society*, 17/2 1988, pp. 183–207.

¹⁴ Klaus Zimmermann, “La Variedad Juvenil y la Interacción Verbal entre Jóvenes”, ed. Félix Rodríguez, *El Lenguaje De Los Jóvenes*, Barcelona: Ariel, 2002, pp. 137–61.

¹⁵ Félix Rodríguez, *El Lenguaje de Los Jóvenes*, Barcelona: Ariel, 2002.

Los jóvenes no respetan la normativa como lo hace un adulto. En añadidura tienen una situación comunicativa especial: su inseguridad y cambiante competencia conlleva uso frecuente de: malas palabras o palabras tabús, acrónimos, abreviaciones, marcadores del discurso, vocativos.¹⁶ Además, dice Drague,¹⁷ la influencia del inglés es patente en el lenguaje y es interesante ver el impacto que tiene en el juvenil. Las palabras nuevas entran con gran facilidad en el lenguaje juvenil, y pasan luego al lenguaje adulto.

A diferencia del inglés, del alemán, del italiano y del noruego, el lenguaje de los jóvenes de habla española no ha hecho correr mucha tinta. Sin embargo, Briz opina que la incidencia del lenguaje juvenil en la historia de la lengua puede llegar a ser mayor en el futuro, gracias a su presencia en los medios de comunicación y en ciertos productos literarios de adolescentes. Podremos, asimismo, predecir los cambios que se producirán en nuestro lenguaje en el futuro, mediante el conocimiento de la forma de hablar de los jóvenes, ya que éstos desempeñan la función de “filtros”. Con lo descrito anteriormente, el interés por la investigación de esta habla, queda justificado.

3.3 ¿Cómo estudiar el lenguaje juvenil?

Varios autores coinciden en las ventajas de la existencia de un corpus para el estudio del lenguaje de este grupo etario, aunque esta tarea no está exenta de retos.¹⁸ Debido a las características particulares de la adolescencia, el análisis de su lenguaje debe partir de su interacción, que es fundamentalmente oral, espontánea y tiene lugar en situaciones distendidas. Según Briz, “La base de partida para su estudio ha de ser el acto comunicativo, la conversación de o entre jóvenes: no pueden tomarse los elementos por separado o de manera

¹⁶ Ibid.

¹⁷ Eli-Marie Drange, *Anglicismos en el Lenguaje Juvenil Chileno y Noruego: Un Análisis Comparativo*, Bergen, Universidad de Bergen, 2009.

¹⁸ Jan Svartvik, Karin Aijmer, and Bengt Altenberg, *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, London ; New York: Longman, 1991.

aislada [...].¹⁹

La lingüística de corpus se caracteriza por una metodología empírica que realiza sus investigaciones sobre la base de “colecciones extensas de textos naturales”, los llamados *corpus*, línea de trabajo cada vez más considerada en la lingüística actual. Lo citado por los autores anteriores, en añadidura a lo que dice Biber, “La investigación empírica de los corpus puede verter luz en cuestiones lingüísticas no detectadas”,²⁰ justificaría el uso de un método inductivo empírico. Así el corpus COLA brinda ventajas para la investigación de la lengua en uso, porque parte de enunciados reales, y como tal facilita información acerca de la lengua, a la que de otro modo no hubiéramos accedido por la intuición, proporcionando resultados más objetivos.

Por lo expuesto, la premisa imprescindible para el escrutinio de un lenguaje juvenil, es la elaboración de un corpus adecuado, que refleje su estilo propio. Es bien sabido que un joven no habla de la misma manera con un igual que con un adulto, cambia de registro. La grabación por los propios informantes sin presencia de adulto resulta el método más eficaz a la hora de la recopilación de un lenguaje informal y espontáneo ya que se evita la “paradoja del observador”,²¹ es decir, observar como habla la gente cuando no está siendo observada.

4. Construcción del Corpus Cola

4.1 La recolección de datos

Las grabaciones comenzaron a hacerse en el 2002 con una grabadora SONY-Net, minidisk con un micrófono de solapa o de mosquito cuyo paso de los archivos grabados a los utilizables en el programa Transcriber resultaba muy lento y caro. A partir del 2007 se usaron grabadoras MP3, lo que facilita el trasvase de los archivos de sonido

¹⁹ Antonio Briz, “La Interacción entre Jóvenes. Español Coloquial, Argot y Lenguaje Juvenil”, p. 142.

²⁰ Douglas Biber, Ulla Connor, and Thomas A. Upton, *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*, Amsterdam: John Benjamins Publishing Company, 2007, p. ix.

²¹ William Labov, *Sociolinguistic Patterns*, Oxford: Blackwell, 1972, p. 209.

directamente al ordenador.

El método de grabación de COLA es el ya utilizado por COLT, Corpus de lenguaje juvenil inglés (www.uib.hf.aksis/colt), basándose ambos corpus en el modelo Longman del BNC, Corpus Nacional Británico, *British National Corpus*. El primer paso a dar en la recolección de los datos fue conseguir contactar con los centros escolares, seleccionados con anterioridad por ingresos económicos, zona residencial y formación académica y sus permisos para grabar en ellos. Una vez obtenidos éstos, se comienza con la captación de informantes. Para ello se seleccionan diferentes unidades de cada nivel académico, según unos criterios previos, encaminados a obtener datos equilibrados de sexo, edad y nivel social. El criterio básico es la edad de la adolescencia comprendida entre los 13 y 19 años. A continuación, en cada aula explicamos el proyecto invitando a los jóvenes a que colaboren como informantes. A aquellos que aceptan a ser grabados se les pasa unos cuestionarios sobre sus datos, que deben firmar sus padres. Cada participante menor de 18 años, debe contar con el consentimiento expreso de su padre, madre o tutor familiar, aprobando que el habla de su hijo pueda ser utilizado como material de investigación lingüística. De entre los informantes se escogieron a tres o cuatro como *reclutas*, que así se denomina a los chicos encargados de realizar la grabación. Estos reclutas, después de haberlas realizado deben rellenar además, otros cuestionarios con los datos de los compañeros que han participado en cada conversación grabada por ellos como su relación con ellos: compañero de clase, amigo, vecino, hermano, primo; el lugar donde se lleva a cabo la grabación: recreos, parques, botellones, cibercafés, sus casas; actividad que realizan: deberes escolares, juegos, tertulias; edad y nacionalidad, etc.

Una vez entregado el material necesario a cada recluta, éste o ésta graba durante tres o cuatro días las conversaciones que va manteniendo con sus compañeros, sin presencia de adultos. Posteriormente devuelve el material grabado a las personas responsables del proyecto. Hemos comprobado que a las chicas de esa edad, cuando se reúnen, les resulta fácil hablar largo y tendido. Sin embargo los chicos, sobre todo en edades de 14 a 16 años, prefieren jugar a la 'playstation', o juegos de ordenador o hacer deporte, por lo que resulta más difícil conseguir que conversen. Un joven del corpus no-

ruego UNO comenta antes de proceder a las grabaciones:

det pleier egentlig vi aldri å gjøre, bare sitte og snakke (ØSVGGU1)
 es algo que no hacemos nunca, lo de estar así como hablando

4.2 La transcripción del material

Para la transcripción ortográfica del sonido grabado hemos usado el programa Transcriber²² que permite a la par que se oye el sonido, escribir lo que se dice en un archivo de texto, quedando así sincronizados los archivos sonoros y transcritos.²³ Esto lleva a poder leer la transcripción HTML y escuchar el sonido al mismo tiempo. La transcripción y los archivos de sonido están insertados en un sistema Web hecho para el proyecto COLA. La transcripción se convierte en un archivo HTML, con enlaces a fragmentos de sonido. Un ejemplo se encuentra en el siguiente sitio: <http://clu.uni.no/humfak/cola/txt-lyd/mabpe2-01c.htm>

IMAG-1

T	L	10	MABPE2J01:	que guarro tronco
T	L	10	MABPE2G01:	si pues es sin comer sabes je je je
T	L	10	MABPE2J01:	basta tron
T	L	10		no me acordaba que tenía el micro
T	L	10	MABPE2G01:	pero no lo tienej conectado
T	L	10	MABPE2J01:	siii
T	L	10	MABPE2G01:	ah ah ah
T	L	10	MABPE2J01:	pero que me da igual tronco
T	L	10	MABPE2J01:	1[yo soy a mi modo]
T	L	10	MABPE2G01:	1[si pues entonces a mi mi]
T	L	10	MABPE2J01:	a quien no le guste que no lo oiga
T	L	10	MABPE2G01:	ja ja ja ja
T	L	10	MABPE2G01:	punto final y se acabó
T	L	10	MABPE2G01:	te enteras @nombre
T	L	10	MABPE2J01:	2[cállate gilipollas ja ja ja]
			MABPE2G01:	2[je je je je je je]

²² Accesible en esta página <http://trans.sourceforge.net/en/presentation.php>

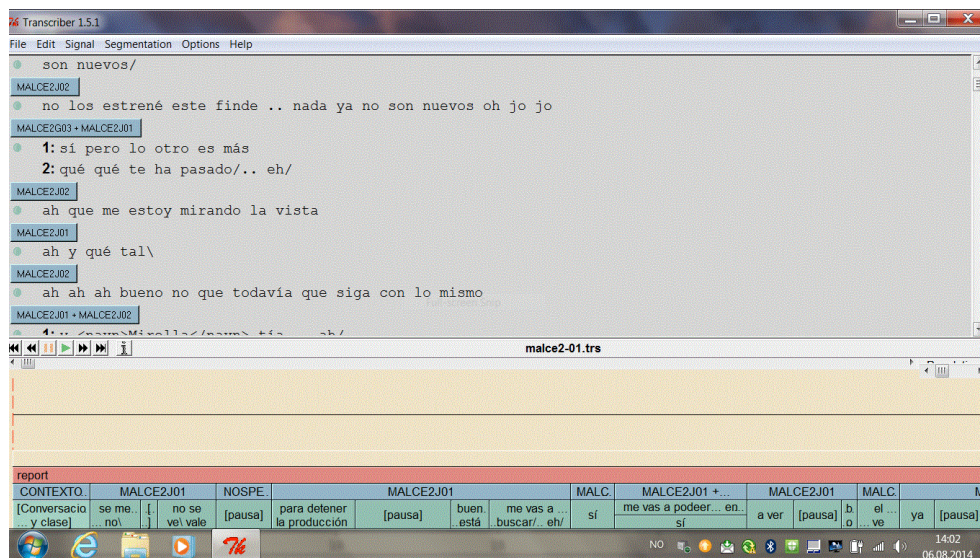
²³ Manual: <http://trans.sourceforge.net/en/usermanUS.php>

T	L	10	MABPE2G02:	ja ja ja
T	L	10	MABPE2J01:	o tú ú @nombre te enteraaas/
T	L	10	MABPE2J01:	un mínimo respeto
T	L	10	NOSPEAKER:	<pausa/>
T	L	10	MABPE2G01:	@nombre esa te tocas el chochete ja ja ja ji ji
T	L	10	MABPE2G01:	ji ji ji ja ja ja ja ja y por eso también me reía porque mice
T	L	10	MABPE2G01:	yo me llamo
T	L	10	MABPE2J01:	ah ji ji yo me llamo no sé tu mismo
T	L	10	MABPE2G01:	porque es que dice yo me llamo @nombre
T	L	10		y claro me vino al acabeza me tocas el chochete
T	L	10	MABPE2J01:	yo ya lo había oído @nombre yo pensé chúpate el culete
T	L	10	MABPE2G01:	el chochete tía
T	L	10	MABPE2J01:	una polla

Con los distintos comandos del programa Transcriber se pueden ir editando sonidos y eventos que se entretajan en una conversación. Este programa deja ir aplicando uno y otro turno según vaya cambiando el hablante, adjudicando un nuevo renglón para cada uno. Los hablantes se identifican con unos códigos definidos para cada grabación que incluyen, ciudad, nombre del centro, nivel escolar, chico o chica y orden de turno en la conversación. Ej.: MALCE2J01 significa: MA es abreviación de Madrid, LC son las siglas del nombre del colegio, E2 es 2º de ESO, J es la chica que comienza hablando la primera (01) en esta grabación, y MALCE2G03 es el chico que entra como el tercero en la conversación.

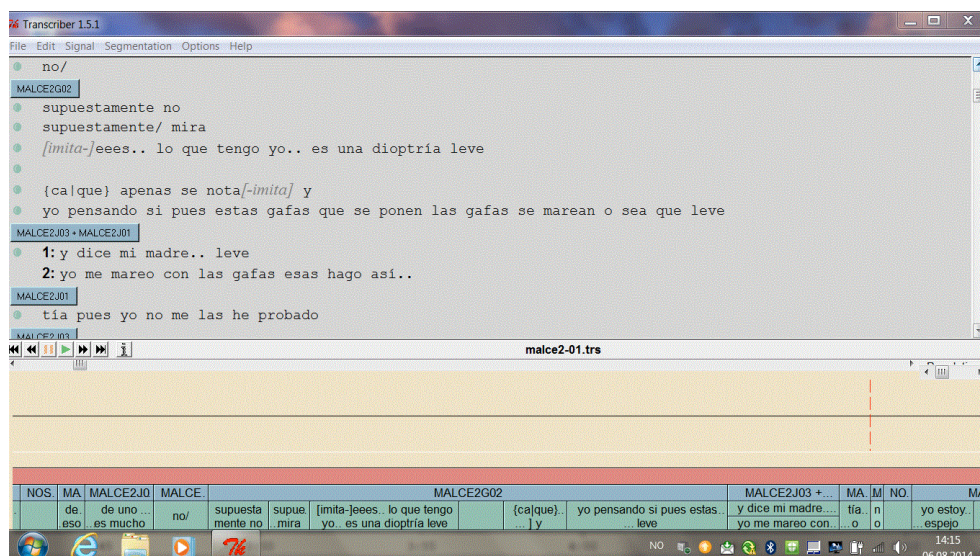
Como en una conversación informal, y más si es entre jóvenes, es frecuente quitarse la palabra, se crean solapamientos de palabras o enunciados entre dos o más hablantes. El programa Transcriber cuenta con un comando que subdivide un turno en dos entradas o renglones para escribir por separado las dos palabras o enunciados que se emiten a la vez. Sin embargo, cuando aparecen más de dos solapamientos, en el caso de más de dos hablando a la vez, no resulta posible transcribirlas todas. En las transcripciones del corpus COLA aparecen los dos hablantes más sobresalientes, MALCE2G03 y MALCE2J01:

IMAG-2



Algunos comandos, como “insertar evento”, que sirven para indicar información extra lingüística del modo de enunciar el texto como reírse, gritar, etc., no se usan, ya que esta información se obtiene escuchando la transcripción con sonido en la red. Se han marcado los anglicismos con [ing], cuando cantan canciones, con [C]. Cuando imitan a compañeros, padres y profesores se ha marcado con [imi], ya que es interesante ver los diferentes modos de introducir lenguaje directo:

IMAG-3



El archivo de texto transcrito se sincroniza con su archivo de sonido al presionar la tecla return/enter quedando ambos unidos en el

tiempo, de tal manera que se va leyendo lo que el parlante va diciendo.

4.2.1 Las búsquedas y las variables

Los datos de cada hablante, edad, sexo, nivel social, etc., se encuentran registrados en otro archivo que a su vez se conectan con las transcripciones en un formato Corpus Workbench²⁴ por el cual se pueden realizar búsquedas concretas. Tenemos una imagen en pantalla para el programa de búsqueda en <http://clu.uni.no/humfak/cola/s9>

IMAG-4

The screenshot shows a web browser window titled "Searching with IMS CWB" at the URL clu.uni.no/humfak/cola/s9.html. The page displays search filters for various variables: Corpus (COLA), Informant (MALCE2J01), alder (All), Kjønn (J), Sosialstatus (All), Skole (All), Skoletype (All), Trinn (All), +/- (+), and Fil (All). Below the filters, there are search options for Word 1, Word 2, and Word 3, with "Whole word" selected for each. There are also checkboxes for "BLANK" and "Minusord". A search button is visible. Below the search options, there is a "Filnavn" field with the value "x". At the bottom, a table shows search results with columns for List, KWIC context (char), Sound (sec), lines per page, and Case insensitive.

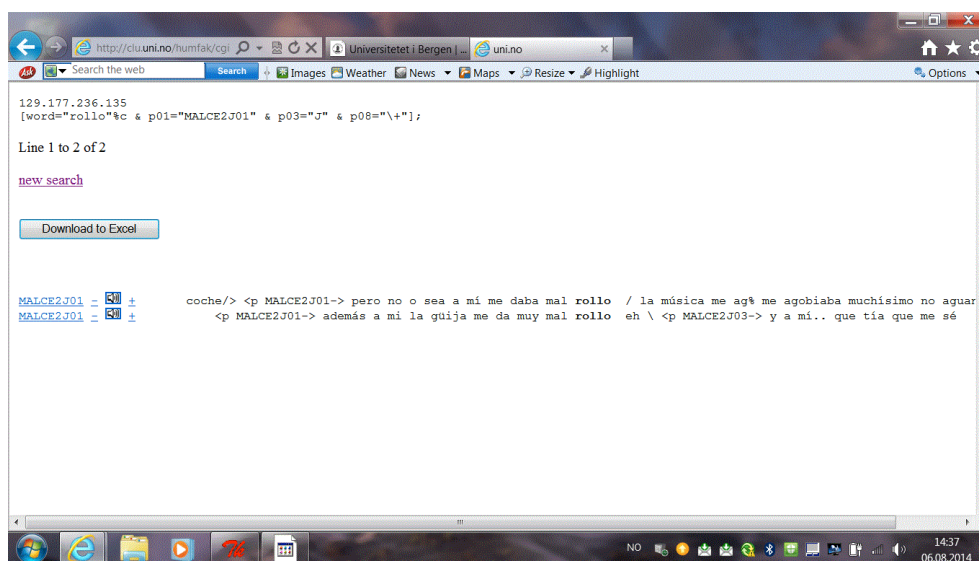
List	KWIC context (char)		Sound (sec)		lines per page	Case insensitive
	Left	Right	Left	Right		
KWIC-concordance (right sort)	60	70	3	3	43	<input checked="" type="checkbox"/>

4.2.2 Las concordancias

En los resultados aparecidos después de una búsqueda, las concordancias, se encuentran sincronizados el texto y el sonido en varios puntos. Los fragmentos de sonido de la concordancia se pueden presentar en archivos de sonidos WAV o MP3. Hay una versión especial donde el usuario puede hacer una clasificación de cada línea de concordancia, <http://clu.uni.no/humfak/cola/s9.html>

²⁴ Oliver Christ, "A modular and flexible architecture for an integrated corpus query system", Proceedings of COMPLEX 94: 3rd Conference on Computational Lexicography and Text Research, Budapest: 1994, pp. 23–32.

IMAG-5



Los textos han sido convertidos al formato usado por Wordsmith, un programa de concordancia de ventanas. Esta versión puede emitir el sonido de la concordancia y también tener un mejor soporte para los textos descifrados XML, aunque las facilidades de la búsqueda no son tan flexibles como con el Corpus Workbench.

4.3 La anonimización

Los archivos, tanto el escrito como el de audio, obligados por la legislación europea y noruega han sido anonimizados con el Programa Audacity²⁵ que permite anular determinados sonidos, insertando silencios en los momentos convenientes. Así se han eliminado los datos que pudieran dar pie al reconocimiento de los hablantes o de personas referidas por ellos: nombres, centros escolares, direcciones, números telefónicos, etc... En algunos casos se mantienen los moteles como *Mary*, *Quique*, *Pepe*, etc., que por ser muy comunes, no particularizan a nadie.

Aunque la anonimización es una tarea lenta, porque para anonimizar las conversaciones se requiere volver a escuchar todos los archivos de sonido, en sentido positivo, cabe decir que el hecho de haber garantizado la anonimidad, ha supuesto una mayor libertad de expresión para los hablantes.

²⁵ <http://audacity.sourceforge.net/?lang=nb>

5. Conclusiones

Para concluir podríamos decir que, aunque la elaboración del corpus COLA exige mucho tiempo y haya supuesto tener que vencer dificultades como: permisos para obtener hablantes dispuestos a grabar sus conversaciones, eliminación de material basura, tiempos inútiles de grabación no aptos para ser transcritos por ruidos, no entendibles, falta de espontaneidad por la presencia del micrófono, etc., el lento proceso de transcripción (un minuto hablado corresponde al menos a diez minutos de transcripción) y anonimización (posterior repaso de los archivos de sonido para la anonimización), hoy en día son más de 700 personas que lo usan ya que al estar en línea es utilizado actualmente por lingüistas de 34 países, así como en distintas universidades españolas: Barcelona, Madrid, Alicante, Vigo, Valencia. A su vez, ofrece gran rentabilidad, ya que puede ser usado en clases de español como segunda lengua, trabajos de grado, postgrado, tesis doctorales y artículos en varias universidades del mundo.²⁶ Es usado con este fin sobre todo en: Oslo, Kristiansand, Estocolmo, Lovaina, Londres, Bergen y Berlín.

Bibliografía

- Albelda, Marta, “Rentabilidad de los corpus discursivos en la didáctica de lenguas extranjeras,” ed. Javier de Santiago Guervós, *Actas del XXI Congreso Internacional de ASELE, 2010*, Salamanca.
- Barras, Claude et al., “Transcriber: a Free Tool for Segmenting, Labeling and Transcribing Speech”, *First International Conference on Language Resources and Evaluation (LREC), ELRA*, [Paris: 1994] Granada: 1998, pp.1373-1376.
- Biber, Douglas, Ulla Connor, y Thomas A. Upton, *Discourse on the move using corpus analysis to describe discourse structure*. Studies in

²⁶ Marta Albelda, “Rentabilidad de los corpus discursivos en la didáctica de lenguas extranjeras,” ed. Javier de Santiago Guervós, *Actas del XXI Congreso Internacional de ASELE, 2010*, Salamanca.

- corpus linguistics vol. 28; Amsterdam: John Benjamins Publishing Company, 2007.
- Briz, Antonio, “La interacción entre jóvenes. Español coloquial, argot y lenguaje juvenil”, ed. María Teresa Echenique Elizondo and Juan Pedro Sánchez Méndez, *Lexicografía y Lexicología en Europa y América: Homenaje a Günther Hensch*, Madrid: Gredos, 2003, pp. 141–154.
- Christ, Oliver, “A modular and flexible architecture for an integrated corpus query system”, Proceedings of COMPLEX 94: 3rd conference on Computational Lexicography and Text Research, Budapest: 1994, pp. 23–32.
- Crowdy, Steve, “Spoken Corpus transcription”, *Literary & Linguistic Computing*, 9/1, 1994, pp. 25–28.
- Drange, Eli-Marie, *Anglicismos en el lenguaje juvenil chileno y noruego: un análisis comparativo*, Bergen: Universidad de Bergen, 2009.
- Eckert, Penelope, “Adolescent Social Structure and the Spread of Linguistic Change”, *Language in Society*, 17/2, 1988, pp. 183–207.
- Haslerud, Vibecke, “The variable (t) in Sydney adolescent speech”, University of Bergen, 1995.
- Hofland, Knut et al., “COLA: A Spanish spoken corpus of youth language”, ed. University of Birmingham Centre for Corpus Research, *Proceedings from the Corpus Linguistics Conference Series*, Birmingham: 2005.
- Hummel, Martin, “Sincronía y diacronía de los llamados adjetivos adverbializados y de los adverbios en –mente”, *Anuario de Letras. Lingüística y Filología*, 1/2, 2013. pp. 217–281.
- Jørgensen, Annette Myre y Juan Antonio Martínez, “Vocatives and phatic communion in Spanish teenage talk”, ed. J. Normann Jørgensen, *Love ya hate ya-The sociolinguistic study of youth language and youth identities*. Cambridge: Cambridge Scholars Publishing, 2010, pp. 193–209.
- Jørgensen, Annette Myre, “Las formas de tratamiento en el lenguaje juvenil”, ed. Leticia Rebollo and Celia Lopes, *Las formas de*

tratamiento en español y en portugués: Variación, cambios y funciones conversacionales, San Pablo: 2011, pp. 1–19.

- Jørgensen, Annette Myre, “Los vocativos en el lenguaje juvenil de Santiago de Chile y Madrid”, ed. Carmen García and Maria Elena Placencia, *Estudios de variación pragmática en español*, Buenos Aires: Dunken, 2011, pp. 141–166.
- Jørgensen, Annette Myre y Eli-Marie Drange, “La lengua juvenil de las metrópolis Madrid y Santiago de Chile”, *Arena Romanistica* 9, 2012, pp. 74–96.
- Labov, William, *Sociolinguistic Patterns*, Oxford: Blackwell, 1972.
- Martínez López, Juan Antonio, “Lexical innovations in Madrid's teenage talk: some intensifiers”, ed. Anna-Brita Stenström & Annette Myre Jørgensen, *Youngspeak in a Multilingual Perspective*, Amsterdam: John Benjamin's Publishing Company, 2009, pp. 81–95.
- Mura, Angela, *La fraseología del desacuerdo: los esquemas fraseológicos en español y en italiano*, E-Prints Complutense. Recuperada en Abril 8, 2013, del sitio Web tema: Portal Recursos Educativos Abiertos (REA) <http://www.temoa.info/es/node/678548>
- Palacios Martínez, Ignacio y Paloma Núñez Pertejo, “Strategies used by English and Spanish teenagers to intensify language”, *Spanish in Context*, 11/2 2014, pp. 175–201.
- Placencia, María Elena y Carmen García, *Research on politeness in the Spanish-speaking world*, Mahwah, New Jersey: Lawrence Erlbaum, 2007, XVII.
- Recalde, Montserrat y Victoria Vázquez Rozas, “Problemas metodológicos en la formación de corpus orales”, ed. Pascual Cantos Gómez and Aquilino Sánchez Pérez, *A Survey of Corpus-based Research: Panorama de investigaciones basadas en corpus*, Murcia: AELINCO, 2009, pp. 37–49.
- Rodríguez, Felix, *El lenguaje de los jóvenes*, Barcelona: Ariel, 2002.
- Stenström, Anna-Brita, Gisle Andersen, and Kristine Hasund, *Trends in Teenage Talk: Corpus compilation, analysis and research* Amsterdam: John Benjamin's Publishing Company, 1998.

- Stenström, Anna-Brita, “Taboo words in teenage talk: London and Madrid girls’ conversations compared.” ed. Clare Mar-Moliner, and Miranda Stewart, *Language Variation and Change: Historical and contemporary perspectives. Special issue of Spanish in Context* 3:1, 2006, 168, pp. 115–138.
- Stenström, Anna-Brita, “Pragmatic Markers in contrast: Spanish *pues nada* and English *anyway*”, ed. Anna Brita Stenström and Annette Myre Jørgensen, *Youngspeak in a Multilingual Perspective* Amsterdam: John Benjamin's Publishing Company, 2009, pp. 137–160.
- Stenström, Anna-Brita y Annette Myre Jørgensen, “La pragmática contrastiva basada en el análisis de corpus”, ed. Lars Fant and Ana María Harvey, *El diálogo oral en el mundo hispanohablante*, Frankfurt am Main: Vervuert-Iberoamericana, 2011, pp. 251–276.
- Svartvik, Jan, Karin Aijmer, y Bengt Altenberg, *English corpus linguistics: studies in honour of Jan Svartvik*, London: New York: Longman, 1991.
- Zimmermann, Klaus, “La variedad juvenil y la interacción verbal entre jóvenes”, ed. Felix Rodríguez, *El lenguaje de los jóvenes*, Barcelona: Ariel, 2002, pp. 137–161.