

FUNCIONES Y DURACIÓN DE *O SEA*: DATOS DEL CORPUS CONVERSACIONAL TAPATÍO

FUNCTIONS AND DURATION OF *O SEA*: DATA FROM THE ‘TAPATIO’ CONVERSATIONAL CORPUS

Rebeca Martínez Gómez y Karol Ibarra Zetter
Universidad de Nuevo México

Este ensayo describe la recolección de un corpus conversacional de la Zona Metropolitana de Guadalajara. El corpus consta de más de 40 horas de grabación de conversación espontánea entre 114 hablantes. Describimos la metodología utilizada en la recolección de materiales y finalmente proveemos un breve estudio del marcador discursivo *o sea*. Desde el punto de vista de la fonología basada en el uso (Bybee, 2001), probamos si la duración de este marcador discursivo tiene una relación con la función que realiza en el discurso. Se extrajeron 100 ejemplares de *o sea* producidos por 10 hablantes y se analizaron en cuanto a su función discursiva. La duración de estos fue medida utilizando el programa Praat (Boersma y Weenink, 2016). Los resultados de un análisis de varianza muestran que efectivamente el marcador discursivo *o sea* se produce con una duración significativamente diferente dependiendo de su función discursiva.

PALABRAS CLAVE: corpus, conversación espontánea, español de Guadalajara, marcadores discursivos, fonología basada en el uso

The present work describes the building of a conversational corpus from the Metropolitan Area of Guadalajara. This corpus contains more than 40 hours of audio recordings of spontaneous conversations from 114 speakers. The methodology for data collection is described here, followed by a brief study of the discourse marker *o sea*. From the perspective of usage based phonology (Bybee 2001), we test if the duration of the marker is related to its functions in discourse. We extracted a 100 tokens of *o sea* produced by 10 speakers and we analyzed them according to their function. We then measured the duration of these tokens using Praat (Boersma y Weenink 2016). An ANOVA test shows that there is a significant difference on duration among the functions *o sea* has in discourse.

KEYWORDS: corpus, spontaneous conversation, Spanish from Guadalajara, discourse markers, usage based phonology

RECEPCIÓN: 13/06/2016

ACEPTACIÓN: 02/11/2016



Introducción

Desde finales del siglo xx, muchos de los trabajos empíricos de producción lingüística se han basado en datos provenientes de fuentes “naturales”, especialmente desde el punto de vista de la lingüística funcional (e.g. Hopper, 1987; Bybee, 2002). Por fuentes naturales, se entiende que las investigaciones se basan en el habla de un grupo social, y por lo tanto que reflejan la realidad lingüística de dicha comunidad. Esto contrasta con estudios lingüísticos previos, cuyos análisis se basaban en la intuición de los hablantes de la lengua estudiada o de los mismos investigadores. Es decir, ahora la corriente en la lingüística es crear muestras de la lengua en uso real en los cuales se fundamenten los estudios. Este conjunto de ejemplos reales de uso de la lengua se denomina corpus lingüístico, y consta de una colección de textos naturales (en un amplio sentido de la palabra) compilados ya sea de manera escrita u oral en grabaciones de audio o video. Por este motivo, la creación de corpora es un trabajo necesario para los estudios del lenguaje en su contexto natural.

En este breve artículo se describirá la recolección de un corpus conversacional de materiales provenientes de la Zona Metropolitana de la ciudad de Guadalajara, México. En esta primera sección, hablaremos sobre aspectos generales de los corpora, y en la sección posterior, daremos detalles sobre la metodología utilizada y sobre las características generales del corpus presentado. Por último, para ejemplificar el tipo de análisis lingüístico que se puede hacer con un corpus de estas características, exponemos un análisis de las funciones del marcador discursivo *o sea* y su correlación con su duración.

El desarrollo de la lingüística computacional y la lingüística de corpus ha conllevado el incremento en la elaboración de corpora (e.g., Baker, 2009; Garside, 1988; Marcus *et al.*, 1993). Existe una gran variedad de corpora diseñados según el tipo de análisis para el que se recaban los datos. Hay distintos enfoques, como compilaciones sobre adquisición de la lengua (e.g., MacWhinney, 2000), o corpora con características específicas que permiten búsquedas complejas de datos. Un ejemplo de esto último sería el Corpus of Contemporary American English (COCA, por sus

siglas en inglés), enfocado al inglés norteamericano o estadounidense, que contiene más de 520 millones de palabras provenientes de diversos géneros discursivos (Davies, 2008). Con este tipo de corpora es posible realizar análisis específicos, como los efectos de la frecuencia en la lengua (véase Bybee y Hopper, 2001).

Como decíamos anteriormente, un corpus puede estar formado por textos escritos o por grabaciones de distinta índole. Por ejemplo, en el COCA todos los datos se presentan de forma escrita aunque algunas de sus fuentes son orales —es decir, se incluyen transcripciones de entrevistas, programas de televisión y radio, películas, etc.— Sin embargo, no se proporciona los archivos de audio correspondientes en la base de datos. Un corpus similar al COCA existe en español, creado por el mismo grupo de investigación, es el Corpus del Español (Davies, 2002), el cual sigue un formato similar al del corpus del inglés. Actualmente contiene aproximadamente dos mil millones de palabras, incluyendo varios géneros y dialectos o variedades del español. Sin embargo, este tipo de corpus, si bien es muy importante para cierto tipo de investigaciones, deja de lado otros aspectos relevantes. Por ejemplo, el hecho de que solamente contenga textos escritos impide hacer investigaciones de índole fonética/fonológica. También, debido a que los datos de distintas variedades de español se mezclan, permite analizar datos a gran escala, pero dejando de lado las características lingüísticas específicas de ciertos registros o sociolectos. De esto se puede concluir que las características de cada corpus definen el tipo de investigación lingüística que se puede lograr con él.

Los corpora orales son de gran relevancia porque dan cuenta de procesos que solamente ocurren en la oralidad (vs. escrituralidad), tales como patrones de cambio morfosintácticos que se van incorporando a la gramática (Du Bois, 2003; Bybee, 2006) o reducciones fonéticas (e.g., Johnson, 2004; File-Muriel, 2009).

Dentro de los corpora orales se encuentran generalmente dos tipos: los que se obtienen a base de entrevistas estructuradas o semiestructuradas —por ejemplo, la famosa entrevista sociolingüística desarrollada por Labov (1981)— y los que se obtienen únicamente de conversaciones espontáneas o sin ningún tipo de intervención del investigador.

Por cuestiones metodológicas (e.g., la necesidad de obtener materiales comparables aunque de distintos grupos sociales), muchos de los corpora orales se basan en entrevistas. Algunos ejemplos son el New Mexico and Colorado Spanish Survey, (NMCOS 2002) que recopila encuestas de vocabulario y entrevistas en el sur de Colorado y al norte de Nuevo México, en los Estados Unidos, o el Pear Film World Corpus (Chafe, 1980), el cual utiliza una narrativa fílmica en que se presenta una historia sin diálogo. El filme se mostró a los miembros de la comunidad de habla estudiada, con el fin de semidirigir y elicitar oralmente construcciones lingüísticas a partir de los eventos que se presentan en el filme. El resultado es una amplia colección de datos de la historia que se presenta, pero narradas en diversas lenguas y dialectos.

Por otro lado, existen corpora orales que son grabaciones de interacciones espontáneas, es decir, sin ningún tipo de estructura fijada a priori por los investigadores. Ejemplos de estos son el Santa Barbara Corpus of Spoken American English (Du Bois *et al.*, 2000), que recoge interacciones orales en distintas situaciones, y el corpus en español Val.Es.Co (Briz, 1995), compuesto por conversaciones coloquiales entre otros tipos de textos.

Una de las diferencias más significativas entre un corpus oral de entrevistas y uno de conversaciones espontáneas es que en el último es posible analizar la estructura de las conversaciones naturales o no estructuradas (Schegloff, 2007) y reducir el efecto de la paradoja del observador (“Observer’s paradox” de Labov, 1981), entre otras cosas. De esta forma, la conversación espontánea muestra características de su género en comparación con otro tipo de textos orales.

En cuanto a corpora existentes del español, además de los ya mencionados, existen otros como el Corpus del Proyecto para el estudio sociolingüístico del español de España y de América (PRESEEA), el cual trata de representar la lengua española hablada en el mundo hispánico “atendiendo a la diversidad sociolingüística de las comunidades de habla hispanohablantes”. (PRESEEA, 2014). Estos corpora recaban datos de diferentes equipos de investigación en diversos países de habla hispana y se enfoca en el habla oral de entrevistas semiestructuradas. Den-

tro de este mega corpus se encuentran diferentes variedades de español mexicano de las siguientes ciudades: Culiacán, Mérida, Monterrey (Rodríguez Alfano *et al.*, 2010), y México (Martín Butragueño y Lastra, 2011).

Algo notable es que todavía no existe ningún corpus oral de Guadalajara, la segunda zona metropolitana más grande de México, con excepción del proyecto “El habla culta de la zona metropolitana de Guadalajara” (Helas Poncela, 1999). Si bien este corpus provee información relevante, su enfoque buscaba a un grupo específico de hablantes de Guadalajara. Además, las grabaciones no se encuentran disponibles para otros investigadores. Finalmente, los audios no cuentan con la calidad de hoy en día, con el fin de llevar a cabo investigaciones de fonética. Consecuentemente, y teniendo en cuenta lo que se ha dicho hasta el momento, se ha creado este corpus con datos de hablantes de la Zona Metropolitana de Guadalajara y que a continuación describiremos.

Objetivo y diseño del Corpus Conversacional Tapatío (CCT)

Ya que la variación en la lengua es una de las áreas más amplias de investigación dentro de la sociolingüística, y que la variación en Guadalajara ha sido escasamente estudiada en corpora (con excepción de Helas Poncela, 1999), el Corpus Conversacional Tapatío (CCT de ahora en adelante) ha sido creado con la intención de estudiar diversos fenómenos lingüísticos propios de esa Zona Metropolitana de Guadalajara (ZMG de ahora en adelante). La ZMG cuenta con más de 4,000,000 de habitantes y está compuesta por seis municipios centrales: Guadalajara, Zapopan, San Pedro Tlaquepaque, Tonalá, Tlajomulco de Zúñiga y El Salto (véase mapa).

En estos materiales se pretende estudiar fenómenos pertenecientes a diversos niveles lingüísticos, desde fonológicos hasta pragmáticos. Así, el CCT recoge grabaciones de interacciones espontáneas cara a cara de hablantes de la ZMG. El CCT representa una muestra de la población de la ZMG. Los hablantes comprenden un rango variado de edades, ocupaciones, géne-



Figura 1. Mapa de la zona metropolitana de Guadalajara incluyendo los municipios de la zona conurbada. Fuente: iconogdl.wordpress.com

ros y distintos trasfondos socioculturales. Las conversaciones documentan interacciones del día a día en distintos contextos, por ejemplo: conversaciones en oficinas, actividades de casa, en traslados por la ciudad, cafés, reuniones sociales, etc. A continuación, describimos los materiales y procedimientos utilizados para la recolección de los datos.

Materiales

De acuerdo con los lineamientos de la Universidad de Nuevo México, se requiere que toda investigación que involucre recolección de datos con participantes humanos tenga la aprobación del Consejo Institucional de Evaluación (Institutional Review Board o IRB, por sus siglas en inglés). La oficina del IRB se encarga de proteger los derechos y el bienestar de los

participantes humanos de una investigación con el fin de lograr investigaciones de calidad desde el punto de vista científico y ético. Siguiendo las políticas del IRB de la Universidad de Nuevo México, para la presente investigación se requirió obtener el permiso de dicha oficina, además de utilizar un documento que se presentaba a los participantes, quienes leyeron y expresaron su consentimiento por escrito. En este consentimiento se explicó a los participantes en qué consistía la actividad (explicada a continuación en el apartado *procedimiento*), cuáles eran sus derechos y se estipulaba que daban consentimiento al uso de sus materiales para investigaciones lingüísticas.

Los materiales utilizados para la recolección de datos fueron dos tipos de grabadoras, una Zoom H4N y otra Sony PCM-M10. Estas grabadoras profesionales graban audio de alta calidad, de modo que análisis acústicos detallados de estas grabaciones son posibles en programas como Praat (Boersma y Weenink, 2016).

El único material para llenar por parte de los participantes fue una breve forma demográfica. Esta incluyó los datos: sexo, rango de edad, nivel educativo, ocupación, colonia, zona metropolitana de Guadalajara y años vividos en la región.

Participantes

En este corpus participaron en total 114 habitantes de la ZMG. La metodología utilizada para reclutar a posibles participantes fue la de muestreo de las redes sociales, o “amigo de un amigo” (Milroy y Gordon, 2003). El proceso consistió en que las dos investigadoras, hablantes de la ZMG, contactaron a sus redes sociales, que a su vez, contactaron a más voluntarios para ser grabados. Esta metodología no solamente es ventajosa para la recolección de los datos sino útil para el análisis de los mismos, ya que los diferentes vínculos de una red social tiene implicaciones en cuanto al comportamiento lingüístico de una comunidad. De esta forma, hay participantes de diferentes generaciones, diversas ocupaciones, regiones de la ZMG y niveles educativos. A continuación se presentan las características generales de los participantes del corpus (las características específicas de los hablantes se describen en el anexo 3).

- Sexo. Este corpus no cuantificó distintos géneros o constructos sociales. Por simplificación de los datos se utilizó un pequeño cuestionario que los participantes contestaban, el cual sólo incluía los campos de “mujer” u “hombre”. En total, el CCT reunió 68 mujeres y 46 hombres.
- Edades. Uno de los requisitos para participar en el corpus era que los participantes fueran mayores de edad. Las edades comprendieron un rango de 18-98 . La edad que se vio mayormente representada en la muestra fue de 25-30 años.
- Zonas. El CCT contiene en su mayoría hablantes de los municipios de Guadalajara y Zapopan (que viven respectivamente en 20 y 19 colonias de esos municipios) y en menor medida hablantes de Tlaquepaque que viven en seis colonias diferentes así como de Tlajomulco de Zúñiga, otra zona donde más recientemente se ha incrementado la población por la creación de zonas habitacionales. Algunas de las colonias donde más participantes viven son: La colonia Centro y Jardines de la Cruz en el municipio de Guadalajara; Las Fuentes, Rinconada Santa Rita y La Estancia, en el municipio de Zapopan; Nueva Galicia en el municipio de Tlajomulco; y El Parián y el Fraccionamiento Revolución en Tlaquepaque. En el anexo 3 se encuentra la lista detallada de colonias donde viven los participantes y los respectivos municipios a los que pertenecen.
- Niveles de educación. El nivel de educación de los participantes en este corpus varía, desde la educación básica: primaria, secundaria, la educación media: preparatoria, superior y licenciatura, estudiantes y egresados de postgrados como maestrías y doctorados. El nivel educativo más común que se encontró en esta muestra es el de licenciatura (49 de los 114 participantes), seguido por 20 participantes con estudios de preparatoria. Los niveles más escasos encontrados fueron la primaria (3 participantes), la educación técnica (3 participantes), cursando un doctorado (3 participantes) o habiendo completado un doctorado (1 participante).

- Ocupaciones. Las ocupaciones de los hablantes son muy variadas: estudiantes, biólogos, contadores, chefs, taxistas, amas de casa, instructores de preescolar y de idiomas, profesores universitarios, recepcionistas, abogados, gerentes, médicos, etc. En el anexo 3 se muestran las ocupaciones de los 114 hablantes del corpus.

Procedimiento

La recolección de los datos comenzó desde diciembre del 2010 hasta el 2015, recaudando información dependiendo de los proyectos y diferentes estancias de las investigadoras. Sin embargo, las dos primeras recolecciones fueron las que generaron mayor cantidad de datos en el invierno del 2010-2011 y el verano del 2011.¹ Para el reclutamiento de participantes se utilizó una invitación oral aprobada por el IRB para atraer a posibles participantes y a sus redes sociales. Esta invitación explicaba a grandes rasgos el objetivo de las grabaciones de audio —hacer análisis del habla de la ZMG en cuanto a diversos niveles lingüísticos—. En especial, se hacía hincapié en que las investigadoras no juzgarían su forma de hablar sino que lo que se buscaba era oír las formas oriundas de hablar en la ZMG. Se enfatizó esto con el fin de tener un mínimo de efecto de algo similar a la paradoja del observador, la cual se refiere al cambio de comportamiento, en este caso lingüístico, del participante, a causa de sentirse observado y desear que su conducta sea aprobada por la investigadora. Una vez que los hablantes aceptaban participar, se les daba un pequeño entrenamiento para utilizar las grabadoras profesionales² y se les indicaba que realizaran las grabaciones en un contexto conversacional en el que ellos se sintieran a gusto, en un ambiente no ruidoso y con un máximo de cinco participantes a la vez de ser posible.

¹ Estas dos estancias fueron patrocinadas por el Instituto Latinoamericano y de Iberia (LAI, por sus siglas en inglés).

² Dejar que los participantes hicieran la grabación por ellos mismos fue parte de la estrategia para disminuir el efecto de la paradoja del observador.

Características del CCT

El CCT recolectó más de 40 horas de conversaciones a través de más de 80 conversaciones. Estas conversaciones son de distintas duraciones (de 10 minutos hasta 2 horas) y abarca una variedad de temas de la vida cotidiana seleccionados por los mismos participantes. Parte de estas conversaciones se han transcrito mientras que otras están en proceso de transcripción. El modelo de transcripción del CCT sigue el de Du Bois (1993), con algunas adaptaciones. De este modo, quedan representadas las unidades entonativas junto con otros elementos importantes para la comprensión y análisis de los textos, tales como los silencios (...), los enunciados truncos (--), alargamientos vocálicos al final de la unidad entonativa (=), entre otros. El ejemplo (1) da muestra de este modelo de transcripción.

Ejemplo (1)

Sofía: ... entonces --
... la verdad si alguien que --
ya no= --
... entra dentro de tu --
código de= ética y de honor y de --
dentro de= --
no sé,
o sea= --
de lo que es digno para ti,
Tomás: no.
aparte --
Sofía: adiós.

Función y forma fónica: el caso de *o sea* en el CCT

En las secciones anteriores hemos descrito la metodología utilizada en la recolección de materiales del CCT y sus características generales. En este último apartado, hacemos un breve análisis del marcador discursivo *o sea* con el objetivo de ejem-

plificar el tipo de análisis lingüístico que se puede llevar a cabo utilizando un corpus de estas características.

Los estudios de fonología basada en el uso (Bybee, 2001) han resaltado casos en los que existe una relación entre la función discursiva de una forma lingüística con su forma fónica (e.g., Scheibman, 2000, con la frase del inglés *I don't know*). A continuación probamos que este también es el caso del marcador discursivo *o sea*. La reducción fonológica en el discurso ha sido analizada desde diferentes perspectivas (cf. Jurafsky *et al.*, 2001; Johnson, 2004; Plug, 2007), mientras otros autores específicamente se han enfocado en la reducción de marcadores discursivos (Villameriel García, 2008; Romera y Gorka, 2002). Lo que proponemos en este trabajo es observar, desde el marco teórico de la fonología basada en el uso de la lengua (Bybee, 2010), no solamente la duración del marcador discursivo *o sea* en general, sino también la correlación entre la función específica del marcador y su duración. Con este objetivo, hablaremos sobre las funciones que se le han adjudicado al marcador y después presentaremos nuestro análisis dando cuenta de las funciones encontradas en el CCT así como de su forma fonética.

O sea, en sus orígenes, consiste de la conjunción *o* y *sea*, de la tercera persona singular del presente de subjuntivo. Haciendo un repaso de lo que se ha escrito sobre este marcador discursivo, encontramos que se le ha considerado desde una simple muletilla hasta una partícula con un sinnúmero de funciones como reparador, reformulador, expletivo, epistémico, entre otros (Briz, 2001a; Briz, 2001b; Galán Rodríguez, 1998; Travis, 2005; Félix-Brasdefer, 2006; Romera, 2001). En su función canónica, se le adjudica clarificar un segmento A con un segmento B (Galán Rodríguez, 1998; Schwenter, 1996; Travis, 2005) como se ve en el siguiente ejemplo:

Ejemplo (2)

Tomás:	aparte el cuate iba mal,	←	Segmento A
	o sea,		
	iba -- ... estancado.	←	Segmento B

Sin embargo, se encuentra en otras posiciones más controvertidas, como simple muletilla (Cortés Rodríguez, 1991) como en el siguiente ejemplo, donde aparentemente no cumple ninguna otra función ya que no hay una reformulación de algo dicho:

Ejemplo (3)

Ana: ... (4.0) el día que lo vi con el chupete,
el día que dije ah,
estás saliendo con alguien?
sí,
con quien?
no la conoces.
y dijo así como que o sea - - ←
y sí= o sea, ←
o sea, ←
uno se da cuenta o sea, ←
y yo esa semana dije bueno,

Y como marcador epistémico (Schwenter, 1996) donde no hay ningún tipo de segmento B después de *o sea*:

Ejemplo (4)

Marylin: ... (1.9) pero así.
Asqueroso.
... (1.2) o sea.
...(1.1) todo charpeado los baño=s,
con alfombra a un lado=,
...(1.8) o sea -- ←
y luego, hay un letrerito?

En el siguiente análisis sobre *o sea* enfatizaremos dos cosas: una, las funciones que encontramos en este corpus a partir de los segmentos anteriores y posteriores al marcador; y dos, cómo estas funciones aparecen fonéticamente con distintas duraciones. A continuación, damos cuenta de lo que se ha dicho con anterioridad sobre este marcador.

Funciones de *o sea*

Dentro de los análisis lingüísticos que se han hecho sobre *o sea*, encontramos el de Schwenter (1996) quien hace un análisis a partir de tres variedades del español peninsular. El autor divide las funciones de *o sea* en dos: como conectivo y como marcador epistémico (aunque argumenta que pueden ocurrir al mismo tiempo). Como conectivo “hearers are guided to an interpretation of a set of utterances in the way speakers want those utterances to be understood”, y como marcador epistémico “the speakers mark their commitment to the truth of the propositions expressed in the utterances” (Schwenter, 1996: 87).

Otra contribución al entendimiento de *o sea* es el estudio de Travis (2005). Ella propone cinco usos vinculados a la reformulación: 1) clarificación, 2) anteceder a un comentario al margen, 3) conclusión, 4) introducir el resto de un enunciado y 5) anteceder un reparo discursivo. Sin embargo, aunque Travis (2005: 132) encuentra que la función epistémica se lleva a cabo a través de las otras distintas funciones de *o sea*, no lo encuentra como función independiente en sus datos de Cali, Colombia, como lo arguye Schwenter con sus datos de España.

El estudio de Félix-Brasdefer (2006) utilizando datos del español de México del Corpus de Monterrey, divide su análisis en tres relaciones que puede presentar *o sea*: semántica, pragmática y textual. Sin embargo, el trabajo no recurre a un método de transcripción que indique patrones de entonación u otros elementos importantes en la decisión sobre el funcionamiento de *o sea* en esos contextos específicos. El primer uso de *o sea* que encuentra este autor es el canónico, donde el segmento después de *o sea* se aproxima al significado anterior a *o sea*. En cuanto a la relación pragmática, el autor argumenta que *o sea* tiene tres funciones: explicar, dar una conclusión o marcar epistemicidad (la cual arguye debe producir un efecto de cortesía en el oyente). La relación textual es cuando el hablante emplea *o sea* para contribuir a la organización y manejo del discurso. En esta categoría, se encuentran otras tres funciones: como reinicio, como intensificador y como cambio de turno. Sin embargo, el análisis que ofrece el autor tiene el problema principal de hacer una di-

visión demasiado categórica entre las funciones, cuando él mismo admite que estas pueden ocurrir al mismo tiempo. Por otro lado, algunas de sus explicaciones parecen centrarse en todo el acto de habla más que en la función específica de *o sea*. Por ejemplo, la función epistémica, que, de acuerdo con este autor, debe producir un efecto de cortesía, no es clara en el siguiente ejemplo que él proporciona:

Ejemplo (5)

No/ se me hace que no/ nunca/ o sea que no/ nunca me ha gustado a mí eso.

Este ejemplo parece más una conclusión del discurso previo (que no se provee), mientras que el efecto de cortesía podría más bien provenir del hecho de que la gente hace clarificaciones al oyente, no necesariamente de una función específica de *o sea*. En general, este tipo de conclusiones sobre *o sea* pareciera un análisis del discurso en el que se encuentra el marcador más que sus funciones. Para poner otro ejemplo, Briz (2001b: 175) presenta el siguiente uso de *o sea* como demandando información:

Ejemplo (6)

¿Estás segura?/ ¿seguro?/ o sea ¿lo tienes claro?

En este caso, vemos que *o sea* no está demandando información, sino simplemente reformulando el segmento anterior (i.e. lo tienes claro está parafraseando las preguntas anteriores). Con el fin de sobrepasar estas limitaciones, proponemos encontrar las funciones de *o sea* a través de la comparación de los segmentos que le preceden y proceden.

Otra función de *o sea* que mencionaremos es como expletivo o muletilla. De acuerdo con Briz (2001b: 183) este uso es un marcador social de un grupo de edad específico (25 años o menores) porque son los que aparecen utilizándolo más frecuentemente. El ejemplo que da el autor es el siguiente:

Ejemplo (7)

Pues o sea el chaval/ o sea/ estaba estudiando o sea EGB/ pero era pues o sea muy malo para estudiar/ un ceporro/ además o sea que es un animal/ es muy BASTO y y o sea

El autor señala que aquí *o sea* no está conectando nada, por lo que lo llama “apoyatura formulativa”. En este breve estudio compararemos los segmentos anteriores y posteriores a *o sea* para observar las funciones del marcador en el CCT.

Por último mencionaremos el estudio de Vázquez Carranza (2012), que usa la metodología del Análisis Conversacional (Schegloff, 2007) en conversaciones espontáneas del español de Toluca, México. A través de un análisis interaccional Vázquez Carranza clasifica *o sea* según el tipo de reparos que hace. Principalmente resuelve el problema de encontrar *o sea* en posiciones donde pareciera que no hay una clarificación (como en el ejemplo 3) y argumenta que “*o sea* can not only do correction or replacement but in general it can be considered to be part of the speaker’s self-adjustments of their talk” (2012: 166). Una segunda contribución importante de este estudio es reconocer y tratar de explicar los casos de *o sea* al final de un turno (o donde hay un lugar relevante de transición de turno, como en el ejemplo 4). Vázquez Carranza no solamente observa cuestiones de turno en estos casos sino el nuevo significado que *o sea* está indexando: la indignación o decepción con respecto a algo.

Análisis de o sea en el CCT

Se extrajeron los primeros mil segundos de seis conversaciones del CCT, y se identificaron todos los ejemplares del marcador *o sea*. De aquí resultaron cien de estos, provenientes de diez hablantes, cuatro mujeres y seis hombres. Los hablantes de los cuales se extrajeron los ejemplares fueron Raquel, Ramiro, Yolanda, Sofía, Paco, Irma, Tomás, Adolfo, David, Gilberto y sus datos demográficos se pueden ver en el anexo 3.

Para la clasificación de funciones de *o sea*, se tomó en cuenta la relación que estos ejemplares tenían con los segmentos anteriores y posteriores, así como con los procesos de cambio lin-

güístico. Esta aproximación distingue a este estudio de estudios previos, ya que consideramos que *o sea* se gramaticalizó a partir de la conjunción *o* y la tercera persona singular del presente de subjuntivo *sea*, y a partir de ahí ha tenido otros cambios semánticos relacionados con la subjetivización y la intersubjetivización (Traugott, 2004). Así, de los datos extraídos distinguimos cuatro funciones de *o sea*: reformulación textual, reformulación subjetiva, marcador en contra de las expectativas y reparador discursivo. A continuación explicamos cada uno.

Reformulación textual. Esta es la función canónica donde *o sea* presenta una reformulación de algo dicho anteriormente, producido por el mismo hablante o por el interlocutor.

Reformulación o marcador subjetivo. La característica de esta función es que no existen claramente segmentos A o B como en la función prototípica; es decir, hay una reformulación pero solamente a nivel cognitivo del hablante (el mismo al que Vázquez Carranza llama *self-initiated self repair* en el mismo turno); puede ocurrir a la mitad de un enunciado, como en el siguiente ejemplo:

Ejemplo (8)

Sofía: la verdad si es alguien que ya no==
entra dentro de tu... código de ética y de honor y de --
dentro de,
no sé o sea,
de lo que es digno para ti--
adiós.

Reparador discursivo. Esta función implica la completa cancelación del segmento A y el remplazo con el segmento B (i.e., no reformulación de lo mismo, sino un cambio total de contenido textual):

Ejemplo (9)

Ana: [digo al llegar manda un men- --] A
o sea,
[márcame de --
ya sé que estás en tu casa,] B

Marcador en contra de las expectativas. La característica de esta función es que el hablante tiende a no proveer más comentario después de o sea; en cambio, el hablante indica con el marcador que la situación narrada anteriormente no es esperada.

Por ejemplo, en el siguiente fragmento, Sofía no espera que un idioma pueda ser tan bonito (según su apreciación) pero tan difícil a la vez:

Ejemplo (10)

Sofía: es--
un cagadero ese idioma.
...(1.4) o sea,
está !muy bonito,
suena !precioso,
pero la verdad es que sí está así de,
a ver espérame.
o sea--

El análisis de esta función coincide con los ejemplos que Vázquez Carranza provee y con lo que él argumenta, excepto que le llamamos marcador en contra de las expectativas con la intención de abstraer el significado (a diferencia de Vázquez Carranza, quien simplemente lo considera como apareciendo al final del turno) . Algo que hay que notar también, es que consideramos que este significado proviene de las frecuentes instancias donde los hablantes utilizan o sea para reformular algo increíble o inesperado. Esto concuerda con lo que dicen Bybee *et al.* (1994) en cuanto a que una forma lingüística puede adquirir un nuevo significado a partir de las instancias en las que aparece con más frecuencia, emancipándose así de su contexto original.

Por otro lado, la hipótesis de la fonología basada en el uso es que hay una relación entre la función que tiene una palabra con las características fonéticas con las que se produce. Es decir que en este caso, presuponemos que cada una de las cuatro funciones tendrá una duración distinta. Para probar esta hipótesis, además de codificar cada uno de los cien ejemplares según su función, se midió su duración utilizando el software Praat (Boersma y Weenink, 2016). Este paso es lo que precisamente

enriquece el estudio de esta partícula lingüística. En comparación con trabajos que solamente mencionan anecdóticamente cuestiones fonológicas como Vázquez Carranza (2012), la calidad del sonido de estas grabaciones permite realizar este tipo de análisis fonológico. Los ejemplares se segmentaron de modo que cada uno quedara seleccionado individualmente a partir del principio de la vocal [o] (o a partir del principio del marcador, en casos donde hubiera reducción de la vocal) hasta el final de la vocal [a] utilizando la forma de onda mostrada por el programa, como se muestra en la Figura 2.

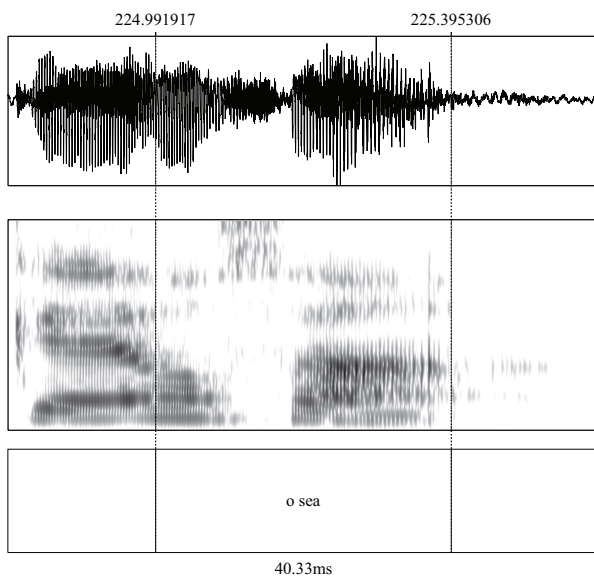


Figura 2. Muestra la segmentación de un ejemplar de *o sea* en una de las conversaciones.

Resultados

Primeramente, la distribución de las funciones y la media de duración de cada una se muestra en el cuadro 1.

Función de <i>o sea</i>	Tamaño muestral	Media de duración
Reformulación textual	49	0.30741
Reparador discursivo	14	0.28844
Reformulación subjetiva	23	0.40348
Marcador en contra de las expectativas	14	0.46527

Cuadro 1. Muestra el número de ejemplares de cada función de *o sea* y la media de duración.

Se realizó un análisis estadístico ANOVA para probar si cada función tenía una duración distinta. Los resultados muestran un valor $p=0.00005$, por lo que podemos decir que, efectivamente, los datos muestran que las funciones de *o sea* se producen con una duración significativamente diferente entre ellas. Para identificar entre qué grupos existen diferencias, se realizó una prueba post hoc de Tukey. Los resultados se resumen en el cuadro 2.

Función	vs. otras funciones	Nivel de significación
Reformulación textual	-Reformulador Subjetivo	.016
	-Contra las expectativas	.000
	-Reparador	.959
Reparador discursivo	-Reformulador Textual	.959
	-Reformulador Subjetivo	.039
	-Contra las expectativas	.002
Reformulación subjetiva	-Reformulador Textual	.016
	-Contra las expectativas	.469
	-Reparador	.039
Marcador en contra de las expectativas	-Reformulador Textual	.000
	-Reformulador Subjetivo	.469
	-Reparador	.002

Cuadro 2. Muestra los resultados de una prueba de Tukey que muestra qué funciones de *o sea* tienen duraciones significativamente diferentes (en rojo). Por ejemplo, cuando *o sea* es un marcador en contra de las expectativas, la duración es definitivamente diferente a cuando funciona como reformulador textual o como reparador, pero no cuando es reformulador subjetivo.

Algo interesante en estos resultados es que *o sea* como marcador en contra de las expectativas es el más largo en duración (media de 0.46527). Aunque esto a principio pareciera ir en contra de lo que normalmente se encuentra en procesos de gramaticalización y forma fónica (i.e. reducción fonológica, Bybee 2006), argumentamos que lo que sucede es lo que Vázquez Carranza encuentra en sus ejemplos de esta función: una reducción de [o'sea] to [o'sa] (p. 179) al mismo tiempo que hay un alargamiento de la última vocal. Es decir, que aunque la duración sea la más larga de todas las funciones, sí existe una reducción a nivel segmental. Esto podría analizarse detalladamente en esta misma base de datos.

Conclusiones

En este artículo hemos descrito el proceso de recolección del CCT, el cual recoge un conjunto de conversaciones espontáneas de la Zona Metropolitana de Guadalajara, México. Además, con base en datos del CCT presentamos un análisis de las funciones del marcador discursivo *o sea* y cómo su forma fónica (i.e., duración) es significativamente diferente dependiendo de la función de cada instancia. Una de las ventajas de contar con este corpus, como hemos ejemplificado brevemente, es el poder realizar estudios desde un enfoque funcionalista que abarque la relación que hay entre la fonética, y otros niveles lingüísticos, e incluso factores extralingüísticos a partir de los datos demográficos. Por ejemplo, también se podría observar si existe una distribución social de marcadores discursivos como *o sea* (altamente asociado al estilo fresa, véase Martínez Gómez 2014) o pues, cuya supuesta reducción pos se vincula a un nivel educativo bajo. Por otro lado, es evidente que la diversidad de contextos cotidianos y hablantes en las grabaciones del CCT también puede ser de utilidad para estudios dentro de áreas más recientes como la de la variación pragmática. En resumen, esperamos que el CCT sea de gran contribución para la lingüística del español mexicano.

Apéndices

Anexo 1: Edades

Edades	Número de mujeres	Número de hombres	Número de mujeres y hombres combinados
	F	M	
18 - 25	16	7	23
26 - 35	28	20	48
36 - 45	7	5	12
46 - 55	5	3	8
56 - 60	10	9	19
60 +	3	1	4
TOTAL	69	45	114

Anexo 2: Educación

Profesional y asociados y/o carrera técnica (y pasante)	7
Licenciatura	49
Primaria	3
Preparatoria	20
Maestría	6
Secundaria	11
Estudiantes / candidatos de doctorado	3
Doctor	1
Técnicos	3

***Nota: los municipios se abreviaron en la tabla y a continuación se ofrece el nombre completo: GDL= Guadalajara; TLAJ Z= Tlajomulco de Zúñiga; SP TLAQ = San Pedro Tlaquepaque; ZAP = Zapopan.

Anexo 3: Datos demográficos por participante (seudónimos)

Pseudónimo	Sexo	Edad	Educación	Ocupación	Colonia	Municipio
Aleida	F	18-25	Preparatoria	Estudiante	Las Fuentes	ZAP
Alejandro	M	18-25	Licenciatura	Estudiante	La Estancia	ZAP
Alicia	F	18-25	Asociados	Ama de Casa	La Estancia	ZAP
Ana	F	18-25	Licenciatura	Estudiante	Las Fuentes	ZAP
Camila	F	18-25	Secundaria	Ventas	Centro	TLAJ Z
Claudia	F	18-25	Preparatoria	Estudiante	Lomas Altas	ZAP
Dan	M	18-25	Licenciatura	Estudiante	Jard. De la Cruz	GDL
Diana	F	18-25	Licenciatura	Estudiante	El parian	SP TLAQ
Elisa	F	18-25	Preparatoria	Estudiante	Hacienda Casa Fuerte	TLAJ Z
Estela	F	18-25	Licenciatura	Estudiante	Las Fuentes	ZAP
Eugenio	M	18-25	Preparatoria	Estudiante	Las Fuentes	ZAP
Humberto	M	18-25	Preparatoria	Lucha libre	San Rafael	GDL
José Luis	M	18-25	Preparatoria	Estudiante	Jard. De la Cruz	GDL
Liliana	F	18-25	Licenciatura	Estudiante	Nueva Galicia	TLAJ Z
Lucy	F	18-25	Licenciatura	Estudiante	El Parián	SP TLAQ
Mariano	M	18-25	Licenciatura	Estudiante	Las Fuentes	ZAP
Miriam	F	18-25	Licenciatura	Estudiante	Residencial Victoria	ZAP
Raquel	F	18-25	Maestría	Psicóloga	Rinc. Santa Rita	ZAP
Rosalba	F	18-25	Estudiante Doctoral	Estudiante/ Instructor	Vallarta San Jorge	ZAP
Rubí	F	18-25	Preparatoria	Estudiante	Jard. De la Cruz	GDL
Saul	M	18-25	Licenciatura	Arquitecto	Lomas Altas	ZAP
Tere	F	18-25	Preparatoria	Estudiante	Chapalita	GDL
Yadira	F	18-25	Preparatoria	Ama de Casa	Zona Centro	GDL
Adela	F	26-35	Secundaria	Cuidadora Adultos Mayores	Zona Centro	GDL
Alfonso	M	26-35	Licenciatura	Estilista	Jardines de Chapalita	ZAP
Alondra	F	26-35	Licenciatura	Estudiante	Las Fuentes	ZAP

Pseudónimo	Sexo	Edad	Educación	Ocupación	Colonia	Municipio
Antonio	M	26-35	Secundaria	Empleado	Barranquitas	GDL
Ariana	F	26-35	Licenciatura	Diseño	Ciudad Granja	ZAP
Carlos	M	26-35	Licenciatura	Electricista	La Estancia	ZAP
Carmen	F	26-35	Técnico	Pintor	San Felipe De Jesús	GDL
Cassandra	F	26-35	Licenciatura	Ama de Casa	Puerta de Hierro	ZAP
César	M	26-35	Licenciatura	Instructor danza	Santa Tere	GDL
Daniel	F	26-35	Licenciatura	Estudiante	Las Fuentes	ZAP
Eduardo	M	26-35	Licenciatura	Planeador Urbano	Chapalita	GDL
Edwin	M	26-35	Preparatoria	Instructor danza	La Normal	GDL
Estrella	F	26-35	Licenciatura	Empleado	La Estancia	ZAP
Fátima	F	26-35	Preparatoria	Empleado	La Capacha	SP TLAQ
Flor	F	26-35	Licenciatura	Soporte técnico	Jard. De la Cruz	GDL
Gonzálo	M	26-35	Secundaria	Chofér	Centro	GDL
Graciela	F	26-35	Licenciatura	Ama de Casa	Santa Tere	GDL
Guillermo	F	26-35	Licenciatura	Comerciante	Isla Raza	ZAP
Ileana	F	26-35	Maestría	Instructor universitario	Zapopan Centro	ZAP
Ines	F	26-35	Licenciatura	Ama de Casa	Santa Tere	GDL
Javier	M	26-35	Licenciatura	Ingeniero	Lomas de Zapopan	ZAP
Jessica	F	26-35	Licenciatura	Acróbata	Seattle	ZAP
Juan Carlos	M	26-35	Licenciatura	Ingeniero	Lomas Altas	ZAP
Karla	F	26-35	Licenciatura	Ama de Casa	La Estancia	ZAP
Laura	F	26-35	Licenciatura	Ama de Casa	Providencia	ZAP
Lázaro	M	26-35	Preparatoria	Cantante	La Normal	GDL
Leonardo	M	26-35	Maestría.	Prof. Investigador	Fracc. Revolución	SP TLAQ
Luigi	M	26-35	Licenciatura	Empleado	Benito Juárez Norte	GDL
Manuel	M	26-35	Licenciatura	Fotógrafo Cine	Seattle	ZAP
Mariela	F	26-35	Licenciatura	Marketing	Residencial Victoria	GDL

Pseudónimo	Sexo	Edad	Educación	Ocupación	Colonia	Municipio
Pablo	M	26-35	Secundaria	Electricista	Balcones de Oblatos	GDL
Paco	M	26-35	Licenciatura	Abogado	Colonia Moderna	GDL
Patricia	F	26-35	Licenciatura	Marketing	Chapalita	GDL
Paula	F	26-35	Licenciatura	Cheff/dueño negocio	Vallarta Universidad	ZAP
Pedro	M	26-35	Licenciatura	Ingeniero	La Estancia	ZAP
Pilar	F	26-35	Maestría	Estudiante/ Instructor	Lomas Altas	ZAP
Ramiro	M	26-35	Licenciatura	Empresario	Jardines de la Patria	ZAP
Rebeca	F	26-35	Asociados	Ama de Casa	Lomas de Zapopan	ZAP
Samuel	M	26-35	Licenciatura	Comerciante	Residencial Victoria	GDL
Socorro	F	26-35	Licenciatura	Receptionista	Arcos de Zapopan	ZAP
Sofía	F	26-35	Licenciatura	Restauradora	Chapultepec	GDL
Susana	F	26-35	Licenciatura	Secretaria	Jard. De la Cruz	GDL
Tomás	M	26-35	Licenciatura	Comerciante	tba	ZAP
Valeria	F	26-35	Secundaria	Empleado	Barranquitas	GDL
Virginia	F	26-35	Asociados	Instructora de danza	Jardines del Bosque	GDL
Yngrid	F	26-35	Licenciatura	Maestra preescolar	La Calma	GDL
Yolanda	F	26-35	Licenciatura	Diseño Gráfico	Rinc. Santa Rita	ZAP
Zeus	M	26-35	Maestría	Empleado de Gobierno	Revolución	SP TLAQ
Aldo	M	36-45	Médico	Doctor gastro	Nueva Galicia	TLAJ Z
Alma	F	36-45	Licenciatura	Empleado	Colina del Pueblo	ZAP
David	M	36-45	Maestría	Prof. Investigador	La Normal	GDL
Enrique	M	36-45	Licenciatura	Biólogo	San Felipe De Jesús	GDL
Esmeralda	F	36-45	Secundaria	Aux. Enfermería	Zona Centro	GDL
Eva	F	36-45	Preparatoria	Ama de Casa	Nueva Galicia	TLAJ Z

Pseudónimo	Sexo	Edad	Educación	Ocupación	Colonia	Municipio
Gerardo	M	36-45	Preparatoria	Técnico electricista	Zona Centro	GDL
Jazmín	F	36-45	Preparatoria	Ama de Casa	Zona Centro	GDL
Julio	F	36-45	Candidato Doctoral	Prof. Investigador	San Felipe De Jesús	GDL
Laura	F	36-45	Licenciatura	Comerciante	Vallarta San Jorge	GDL
Rita	F	36-45	Licenciatura	Medical Practitioner	Rinc. Santa Rita	GDL
Sergio	M	36-45	Secundaria	Músico	Centro	GDL
Alex	M	46-55	Licenciatura	Gte. Recursos Humanos	Jard. Del Rosario	SP TLAQ
Beto	M	46-55	Preparatoria	Taxista	Centro	GDL
Carla	F	46-55	Licenciatura	Cirujano Dentista	Jardín Real	ZAP
Eliana	F	46-55	Secundaria	Vendedora jubilada	Centro	GDL
Erika	F	46-55	Candidato Doctoral	Prof. Investigador	Colina del Pueblo	ZAP
Juan	M	46-55	Secundaria	Mecánico	Centro	GDL
Lorena	F	46-55	Preparatoria	Ama de Casa	Vallarta San Jorge	GDL
Sonia	F	46-55	Primaria	Empleado	Barranquitas	GDL
Adolfo	M	56-60	Preparatoria	Gerencia	La Estancia	ZAP
Alfredo	M	56-60	Licenciatura	Arquitecto	Juan Manuel Vallarta	GDL
Cecilia	F	56-60	Asociados	Instructor de español	San Felipe De Jesús	GDL
Cristina	M	56-60	Primaria	Aseo hogar	Sta. Cecilia/Oblatos	GDL
Edwina	F	56-60	Primaria	Empleada jubilada	La Estancia	ZAP
Francisco	M	56-60	Preparatoria	Empleado	Zona Centro	GDL
Gabriela	F	56-60	Licenciatura	Ama de Casa	Rinc. Santa Rita	ZAP
Gilberto	M	56-60	Secundaria	Jubilado	Mezquitán	GDL
Gustavo	M	56-60	Pasante Ing.	Jubilado	Revolución	SP TLAQ
Irma	F	56-60	Técnico	Jubilado	Vallarta San Jorge	ZAP
Luis	M	56-60	Preparatoria	Asist. Admin. Jubilado	La Estancia	ZAP

Pseudónimo	Sexo	Edad	Educación	Ocupación	Colonia	Municipio
Luis Miguel	M	56-60	Licenciatura	Contador	Centro	GDL
Ma. De la Luz	F	56-60	Técnico	Ama de Casa	Lomas Altas	ZAP
Mario	M	56-60	Asociados	Profesor jubilado	Analco (Centro)	SP TLAQ
Martha	F	56-60	Secundaria	Ama de Casa	La Estancia	ZAP
Pamela	F	56-60	Asociados	Comerciante	Lomas Altas	ZAP
Rosa	F	56-60	Secundaria	Jubilado	La Estancia	ZAP
Yanet	F	56-60	Licenciatura	Instructor de español	Centro Tlaquepaque	SP TLAQ
Yuliana	F	56-60	Secundaria	Ama de Casa	Rinc. Santa Rita	GDL
Guadalupe	F	61+	Primaria	Ama de Casa	Mezquitán Country	GDL
Jorge	M	61+	Licenciatura	Instructor de idiomas	Talpita	GDL
Paty	F	61+	Preparatoria	Ama de Casa	Chapalita	GDL
Sara	F	61+	Preparatoria	Tienda de ropa	Jardines del Bosque	GDL

Bibliografía

- BAKER, PAUL (2009), *Contemporary Corpus Linguistics*, Londres, Continuum (Contemporary Studies in Linguistics).
- BOERSMA, PAUL y DAVID WEENINK (2016), *Praat: Doing Phonetics by Computer* [Software para computadora]. Version 6.0.17, (2016) Institute of Phonetic Sciences, University of Amsterdam, <http://www.fon.hum.uva.nl/praat/> [consultada el 21 de abril de 2016].
- BRIZ GÓMEZ, ANTONIO (coord.) (1995), *La conversación coloquial (Materiales para su estudio)*, Valencia, Universidad de Valencia. Libro agotado Anejo XVI de la revista *Cuadernos de Filología*.
- BRIZ GÓMEZ, ANTONIO (2001a), “El uso de *o sea* en la conversación”, en J. de Kock (ed.), *Gramática española, enseñanza e investigación: lingüística con corpus*. Catorce aplicaciones

- sobre el Español, Salamanca, Universidad de Salamanca, pp. 287-318.
- BRIZ GÓMEZ, ANTONIO (2001b), "Otra vez sobre *o sea*", en Carmen Saralegui Platero y Manuel Casado Velarde (eds.), *Pulchre, bene, recte. Estudios en homenaje al Prof. Fernando González Ollé*, Navarra, Universidad de Navarra: EUNSA, pp. 169-190.
- BYBEE, JOAN (2001), *Phonology and language use*, Cambridge, Cambridge University Press.
- BYBEE, JOAN (2002), "Word Frequency and context of Use in the Lexical Diffusion of Phonetically Conditioned Sound Change", *Language Variation*, 14, pp. 261-290.
- BYBEE, JOAN (2006), "From Usage to Grammar: The Mind's Response to Repetition", *Language*, 82 (4), pp. 711-33.
- BYBEE, JOAN (2010), *Language, Usage and Cognition*, Cambridge, Cambridge University Press.
- BYBEE, JOAN L., y PAUL J. HOPPER (2001), *Frequency and the Emergence of Linguistic Structure*, Ámsterdam, John Benjamins (Typological studies in language, 45).
- CHAFE, WALLACE L. (1980), *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, Norwood, New Jersey, Ablex (Advances in discourse processes, 3).
- CORTÉS RODRÍGUEZ, LUIS (1991), *Sobre conectores, expletivos y muletillas en el español hablado*, Málaga, Librería Agora.
- DAVIES, MARK (2002), *Corpus del español: 100 million words, 1200s-1900s*, <http://www.corpusdelespanol.org> [consultado el 21 de abril de 2016].
- DAVIES, MARK (2008), *The Corpus of Contemporary American English: 520 million words, 1990-present*, <http://corpus.byu.edu/coca/> [consultado el 21 de abril de 2016].
- DU BOIS, JOHN W., STEPHAN SCHUETZE-COBURN, SUSANNA CUMMING y DANAE PAOLINO (1993), "Outline of discourse transcription", en Jane Anne Edwards y Martin D. Lampert (eds.), *Talking data: Transcription and coding in discourse*, Hillsdale, NJ, Lawrence Erlbaum Associates, pp. 45-89.
- DU BOIS, JOHN W., WALLACE L. CHAFE, CHARLES MEYER, SANDRA A. THOMPSON, ROBERT ENGLEBRETSON y NII MAR-

- TEY (2000), *Santa Barbara corpus of spoken American English*, Parte 1, Filadelfia, Linguistic Data Consortium.
- DU BOIS, JOHN (2003), "Discourse and Grammar", en Michael Tomasello (ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, vol. 2, Londres, Erlbaum, pp. 47-88.
- FÉLIX-BRASDEFER, J. CÉSAR (2006), "Pragmatic and Textual Functions of *O sea*: evidence from Mexican Spanish", en Timothy L. Face y Carol A. Klee (eds.), *Selected Proceedings of the 8th Hispanic Linguistics Symposium*, pp. 191-203.
- FILE-MURIEL, RICHARD J. (2009), "The Role of Lexical Frequency in the Weakening of Syllable-Final Lexical /s/ in the Spanish of Barranquilla, Colombia", *Hispania*, 92(2), pp. 348-360.
- GALÁN RODRÍGUEZ, CARMEN (1998), "La dimensión explicativa y deóntica de los conectores *o sea* y *es decir*", *Anuario de Estudios Filológicos*, 21, pp. 85-104.
- GARSIDE, ROGER, GEOFFREY SAMPSON Y GEOFFREY LEECH (eds.) (1988), *The computational analysis of English: A corpus-based approach*, Londres, Longman.
- HELAS PONCELA, MARÍA DEL ROSARIO (1999), *El habla culta de la zona metropolitana de Guadalajara*, Guadalajara, Universidad de Guadalajara.
- HOPPER, PAUL (1987), "Emergent Grammar", *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 13, pp. 139-157.
- JOHNSON, KEITH (2004), "Massive reduction in conversational American English", en *Spontaneous speech: Data and analysis*. Proceedings of the 1st session of the 10th international symposium, Tokio, The National International Institute for Japanese Language.
- JURAFSKY, DANIEL, ALAN BELL, MICHELLE GREGORY Y WILLIAM D. RAYMOND (2001), "Probabilistic relations between words: Evidence from reduction in lexical production", en Joan L. Bybee y Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, Ámsterdam, John Benjamins, pp. 229-254.

- LABOV, WILLIAM (1981), "Field Methods of the Project on Linguistic Change and Variation", en *Southwest Educational Development Laboratory, and National Institute of Education* (U.S.), Austin, Southwest Educational Development Laboratory (Sociolinguistic working paper, 81).
- MACWHINNEY, BRIAN (2000), "The CHILDES Project: The database", vol 2. *Psychology Press*.
- MARCUS, MITCHELL P., MARY ANN MARCINKIEWICZ y BEATRICE SANTORINI (1993), "Building a large annotated corpus of English: The Penn Treebank", *Computational Linguistics*, 19(2), pp. 313-330.
- MARTÍN BUTRAGUEÑO, PEDRO y YOLANDA LASTRA (2011), *Corpus sociolingüístico de la ciudad de México: Materiales de PRESEEA - México*, vol. 1: Nivel alto, México, El Colegio de México.
- MARTÍNEZ GÓMEZ, REBECA (2014), "Language Ideology in Mexico: The Case of Fresa Style in Mexican Spanish", en *Texas Linguistic Forum. Proceedings of the Symposium About Language and Society - Austin 15*.
- MILROY, LESLEY y MATTHEW J. GORDON (2003), *Sociolinguistics: Method and Interpretation*, Malden, MA, Blackwell Pub (Language in society, 34).
- PRESEEA (2014), *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*, Alcalá de Henares, Universidad de Alcalá <http://preseea.linguas.net> [consultado el 1 de marzo 2016].
- PLUG, LEENDERT (2007), "Phonetic Reduction and Pragmatic Organisation in Dutch Conversation: A Usage-Based Account", *Dissertation Abstracts International, Section C: Worldwide*, 68(3), pp. 647-648.
- RODRÍGUEZ ALFANO, LIDIA, MARÍA EUGENIA FLORES TREVIÑO y TZITEL PÉREZ AGUIRRE (2010), *Corpus El habla de Monterrey-PRESEEA*, México, Secretaría de Educación Pública.
- ROMERA, MAGDALENA (2001), "Discourse functional units: a re-examination of discourse markers with particular reference to Spanish", tesis doctoral, Los Ángeles, University of Southern California.

- ROMERA, MAGDALENA y GORKA ELORDIETA (2002), “Características prosódicas de la unidad funcional del discurso entonces: implicaciones teóricas”, *Oralia: Análisis del Discurso Oral*, 5, pp. 247-263.
- SCHEGLOFF, EMANUEL A. (2007), *Sequence Organization in Interaction: A Primer in Conversation Analysis*, Cambridge, Cambridge University Press.
- SCHEIBMAN, JOANNE (2000), “I Dunno: A Usage-Based Account of the Phonological Reduction of Don’t in American English Conversation”, *Journal of Pragmatics*, 32(1), pp. 105-124.
- SCHWENTER, SCOTT A. (1996), “Some Reflections on O Sea: A Discourse Marker in Spanish”, *Journal of Pragmatics*, 25(6), pp. 855-874.
- TRAUGOTT, ELIZABETH C. (2004), “From Subjectification to Intersubjectification”, en Raymond Hickey (ed.), *Motives for Language Change*, Cambridge, Cambridge University Press, pp. 124-140.
- TRAVIS, CATHERINE E. (2005), *Discourse Markers in Colombian Spanish: A Study in Polysemy*, Berlín, Mouton de Gruyter (Cognitive Linguistics Research).
- VÁZQUEZ CARRANZA, ARIEL (2012), “O sea in Talk: a study of Mexican Spanish interactions”, *Essex Graduate Student Papers in Language and Linguistics On-line*, 13, pp. 156-188.
- VILLAMERIEL GARCÍA, SAÚL (2008), “Marcadores del discurso en la lengua de signos española y en el español oral: un estudio comparativo”, Actas del 8º Congreso de Lingüística General, USILSE. UVa. Centro Transdisciplinar de Investigación en Educación (CETIE-UVa), Madrid, Universidad Autónoma de Madrid.

Agradecimientos

Queremos agradecer primeramente a todos los hablantes de Guadalajara que participaron por hacer posible este proyecto. Agradecemos también el financiamiento para la recolección del

presente corpus por parte de Tinker Foundation Field Research Grant, otorgada por el Latin American & Iberian Institute (LAI) en el invierno del 2010 y el verano del 2011. Rebeca Martínez Gómez también agradece al Conacyt, del que era becaria durante la recolección del corpus en el verano del 2011. Finalmente, agradecemos a la doctora Catherine Travis, Australian National University, y a la doctora Naomi Shin, Universidad de Nuevo México, por su guía en el planeamiento y desarrollo de este proyecto.