Corpus españoles dialógicos para el análisis de la conversación

Inmaculada Solís García
Universidad de Florencia

The aim of this paper is to highlight the useful features of Spanish oral corpora for research in conversation analysis. A review on the oral sections of the more general corpora will help us to see how in these transcriptions some of the parameters that must have been taken into account, such as taking turns, interruptions, overlaps, backchannels, etc. are not always respected. Some current parallel and multilingual corpus that will boost the conversation contrastive analysis will be described.

Keywords: Spanish oral corpora; Conversation Analysis; Task-oriented Dialogue; Contrastive Linguistics; Multilingual corpora.

1. Introducción

Existen distintas tipologías de corpus de lengua oral. Según el tipo de texto, Llisterri (2018) diferencia entre corpus orales espontáneos (conversaciones, monólogos) o no espontáneos (revisado, leído, escrito para ser oralizado o escrito para ser oralizado como si no estuviera escrito). También existen tipologías de corpus en función de las situaciones comunicativas en las que se recogen los datos: monológica (un emisor y varios receptores) con textos como las conferencias, clases, discursos políticos; dialógica (varios emisores y varios receptores) que incluye la conversación espontánea, el debate, la mesa redonda o las tertulias en medios de comunicación y, por último, la entrevista, es decir, diálogos con papeles fijados en los que está prevista una intervención mínima del entrevistador y una participación máxima del entrevistado.

Pues bien, la mayoría de los macrocorpus de lengua oral recogidos en ámbito hispánico consisten en entrevistas semidirigidas. Los recuentos más exhaustivos, realizados por Llisterrri *et al.* (2005), Briz y Albelda (2009) y Rojo (2016),

CHIMERA. Romance Corpora and Linguistic Studies 5.1 (2018), 117-129. ISSN 2386-2629

describen los siguientes macrocorpus de entrevistas, o con predominio de entrevistas, que citamos a continuación: Proyecto PILEI y Macrocorpus de la norma lingüística culta de las principales ciudades de España y América (MC-NC)[1]; Corpus del Proyecto para el estudio sociolingüístico del Español de España y de América (PRESEEA)[2]; Corpus del Proyecto Estudio gramatical del español hablado en América (EGREHA)[3]; Corpus Oral y Sonoro del Español Rural (COSER)[4] y numerosos corpus más reducidos que se han incorporado al corpus PRESEEA[5]. En algunos proyectos las entrevistas se denominan "conversaciones semidirigidas", como en el Corpus sociolingüístico de la ciudad de Mérida (Perú) o "relatos semilibres" como en el Corpus de Bogotá o en el de Tucumán[6].

Como Antonio Briz observa (2009 :1), este tipo de corpus permite controlar el equilibrio en los parámetros sociolingüísticos y asegurar la representatividad de los informantes en el total de la muestra, por lo que ofrece una cierta sistematicidad al lingüista. Ahora bien, para el objetivo que hemos asumido, es decir, señalar los corpus que pueden usarse para el análisis de la conversación, hemos de observar las limitaciones de este tipo de corpus. Aunque las entrevistas llegan a ser espontáneas, una de las dos personas dirige el diálogo y el informante habla para que su testimonio lingüístico sea almacenado. Por este motivo no se dan las mismas dinámicas y estrategias comunicativas que en las conversaciones de tipo espontáneo en las que hay un reparto libre de papeles. Así pues, vamos a

[1] Bajo el impulso de Juan M. Lope Blanch, desde 1964 se recogieron muestras habladas de la norma lingüística culta de las principales ciudades de Iberoamérica y de la Península Ibérica.

[2] El proyecto PRESEEA agrupa unos 40 equipos de investigación sociolingüística. El corpus es accesible en http://preseea.linguas.net/ [Fecha de consulta 2.09.2018]

[3] En el marco del Proyecto EGREHA (Estudio gramatical del español hablado en América), dirigido por César Hernández Alonso, se ha recogido un macrocorpus que contiene dos tipos de archivos: los materiales del corpus de la norma culta MC-NC recién descrito y un nuevo grupo de entrevistas orales sin transcripción. No se encuentran publicadas. Los resultados de las investigaciones sobre este corpus figuran en cuatro volúmenes editados en Hernández Alonso 2009.

[4] Es un corpus dialectal de entrevistas en 636 enclaves rurales de la Península Ibérica, coordinado por Inés Fernández-Ordóñez. Actualmente es accesible y es posible consultar 163 transcripciones con archivos sonoros, correspondientes a 210 horas de grabación, que conforman un corpus interrogable de 2.573.719 palabras [http://www.corpusrural.es/descripcion.php - Fecha de consulta 2.09.2018].

[5] Como el Corpus del habla de Almería, el de Jaén y el Corpus Oral de Asturias, entre muchos otros.

[6] El Corpus sociolingüístico de la ciudad de Mérida fue coordinado por Carmen Luisa Domínguez y Elsa Mora; José Joaquín Montes Giraldo publicó una parte del corpus de entrevistas semidirigidas recogido en la ciudad de Bogotá y el corpus de Tucumán fue dirigido por M. Elena Rojas Maye.

centrar nuestra atención en los corpus de conversaciones espontáneas, de debates, mesas redondas o tertulias en medios de comunicación.

2. Corpus conversacionales del español

Los mayores macrocorpus conversacionales actualmente a disposición de los investigadores en ámbito hispánico son los siguientes:

a. El Corpus de conversaciones coloquiales Val.Es.Co. Este corpus, bajo la dirección de Antonio Briz, es el primer corpus publicado en español de conversaciones coloquiales espontáneas y de grabación secreta en situaciones reales de comunicación, con un sistema propio de transcripción que intenta reflejar lo más fielmente posible la oralidad sin dificultar la lectura del texto. Consta en la actualidad de 341 horas de grabación de las que han sido transliteradas más de treinta conversaciones y publicadas diecinueve en Briz y Grupo Val.Es.Co. (2002). Todo el material se halla almacenado en soporte digital pero es necesario solicitar el audio de las grabaciones para poder consultarlo. [http://www.valesco.es - Fecha de consulta 2.09.2018].
b. El Alicante Corpus Oral del Español (ALCORE). Bajo la dirección de Dolores Azorín se han elaborado dos corpus orales de esta zona geográfica: el corpus ALCORE y el COVJA (Corpus Oral de la Variedad Juvenil Universitaria del Español de Alicante), incluido como subcorpus del primero. Se entrevista a los informantes y después, en grupos, se crean conversaciones de cuatro participantes (Azorín Fernández 2002). Ha entrado a formar parte del banco de datos del CREA. No es consultable el audio. [http://corpus.rae.es/creanet.html - Fecha de consulta 2.09.2018].
c. El Corpus del español conversacional de Barcelona y su área metropolitana. El responsable de este corpus es el grupo GRIESBA (Grupo de Investigación del Español de Barcelona), dirigido por M.ª Rosa Vila Pujol. Las transcripciones se han publicado en Vila 2001, pero no es consultable el audio.
d. El Corpus Oral del Lenguaje Adolescente (COLA), coordinado por Annette Myre Jørgensen, recopila más de 300 conversaciones espontáneas entre jóvenes adolescentes (13 a 19 años) de las ciudades de Madrid, Santiago de Chile, Buenos Aires y Managua. Las transcripciones están alineadas al sonido y el audio es consultable. [http://www.colam.org/index-espannol.html - Fecha de consulta 2.09.2018].

e. El C-Or-DiAL está dirigido por Carlota Nicolás (Nicolás 2012). Se trata de una colección de conversaciones de habla espontánea en español transcritas y anotadas con etiquetas prosódicas y de funciones comunicativas. [http://lablita.dit.unifi.it/corpora/cordial/ - Fecha de consulta 2.09.2018].
f. El Corpus del Grupo de Investigación Lingüística Aplicada (COGILA). Las transcripciones y el audio de las conversaciones, producidas en su mayoría por jóvenes universitarios andaluces, aparecen publicadas en Barros 2012.

Asimismo, hemos de citar corpus más generalistas que combinan distintos géneros discursivos orales, entre los que encontramos conversaciones, como son:

g. El Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC) que contiene la transcripción de 126 conversaciones de español hablado peninsular procedentes de programas de la televisión, grabaciones entre amigos, etc. Fue recogido bajo la dirección de Francisco Marcos Marín. Está incorporado al CREA y no es posible consultar los audios. [http://www.lllf.uam.es/ESP/Corlec.html - Fecha de consulta 2.09.2018].
h. C-ORAL-ROM es un macrocorpus de habla espontánea, que, además del español, incorpora otras tres lenguas romances: italiano, francés y portugués. El proyecto fue coordinado por Emanuela Cresti y Massimo Moneglia; el subcorpus del español corrió a cargo de Antonio Moreno Sandoval. Los audios y las transcripciones son consultables, pero su acceso no es libre (Cresti – Moneglia 2005).
i. El Corpus de Referencia del Español Actual (CREA) es más bien un banco de datos que cuenta con una parte oral con transcripciones propias de documentos sonoros extraídas de medios de comunicación y transcripciones incorporadas de corpus orales cedidos a la RAE[7] y recodificados. No es posible consultar los audios. [http://corpus.rae.es/creanet.html - Fecha de consulta 2.09.2018].
j. El Corpus del Español del siglo XXI (CORPES XXI) intenta superar las limitaciones del CREA. Un 10 % de los textos corresponden a la lengua oral

[7] Estos son algunos de los corpus que han entrado a formar parte del CREA: ACUAH (Análisis de la conversación de la Universidad de Alcalá de Henares), COVJA (Corpus Oral de la Variedad Juvenil Universitaria del Español Hablado en Alicante), CSC (Corpus para el Estudio del Español Hablado en Santiago de Compostela), ALFAL (Macrocorpus de la Norma Lingüística Culta de las Principales Ciudades del Mundo Hispánico), CORLEC (Corpus Oral de Referencia del Español Contemporáneo, cedido por la Universidad Autónoma de Madrid), Caracas-77, Caracas-87 (Estudios sociolingüísticos de Caracas), CEAP (Corpus de Encuestas en Asunción de Paraguay) y CSMV (Corpus Sociolingüístico de la Ciudad de Mérida, cedido por la Universidad de los Andes, Venezuela).

con dos millones de palabras. Algunos de los textos disponen de sonido alineado; en otros casos no se puede consultar el alineado, pero se dispone del audio completo e incluso del vídeo completo. La tipología de textos orales que es posible consultar es la siguiente: debates, discursos, entrevistas, magacines y variedades, retransmisiones deportivas, sorteos y concursos, tertulias y "otros", correspondientes a diversos programas de TVE, de Youtube o a páginas de medios de comunicación con disponibilidad de audios, etc. No es posible, sin embargo, consultar directamente textos de conversaciones espontáneas.
[http://www.rae.es/recursos/banco-de-datos/corpes-xxi - Fecha de consulta 2.09.2018].

k. El Corpus El Grial constituye una base de datos formada por diversos corpus. Está construida por el grupo de investigación ALADE bajo la dirección de Giovanni Parodi. Esta base de datos recoge más de 100 millones de palabras, su acceso es público y gratuito. Entre los diversos corpus que lo integran, se encuentran transliteraciones ortográficas de corpus orales (corpus oral de políticas públicas, noticias de televisión, guías didácticas, entrevistas orales, etc.). No se puede acceder directamente a los audios. [http://www.elgrial.cl/index2015.php - Fecha de consulta 2.09.2018]
l. El Corpus del español, creado y mantenido por Mark Davies, contiene 20 millones de palabras de textos contemporáneos, de las que 5 millones suponen el corpus oral. Los documentos orales proceden de la codificación de diversos archivos sonoros o de la cesión de otros corpus orales a este macrocorpus. No es posible consultar los audios. [http://www.corpusdelespanol.org/ - Fecha de consulta 2.09.2018].

Ahora bien, para la investigación en análisis de la conversación, a diferencia de lo que sucede en otro tipo de estudios lingüísticos que pueden llevarse a cabo a partir de transcripciones ortográficas de la lengua oral, es imprescindible disponer de la señal sonora, ya que las transcripciones no siempre respetan los estándares necesarios de transcripción de la oralidad válidos para el análisis de la conversación. A saber, en muchos casos no se presta la suficiente atención a la transcripción de estrategias comunicativas como la toma de turnos, la interrupción, los solapamientos, las señales de *feedback* (marcadores cognitivos, etc.). Es lo que ocurre, por ejemplo, en la mayoría de los bancos de datos generalistas que poseen una sección oral, como el CREA, el CORPES XXI, etc.
Si bien los corpus de conversaciones que hemos citado proporcionan espontaneidad en los interlocutores y variedad situacional, el número de hablantes, la heterogeneidad e imprevisibilidad en los rasgos de edad, sexo y nivel

sociocultural (no siempre estas variables se describen en fichas sociolingüísticas), así como distintos tipos de relación entre los interlocutores, diversidad de temáticas, de espacios físicos, etc. en ocasiones dificultan la comparación intralingüística e interlingüística entre ellos.

Por otro lado, existen otras colecciones de textos conversacionales que no han sido diseñados y recogidos para la investigación lingüística sino para desarrollar aplicaciones en el ámbito de las tecnologías del habla, pero que pueden interesarnos: hablamos de corpus específicamente recogidos para la creación de sistemas de diálogo entre personas y ordenadores.

Como señalan Llisterri *et al.* (2005: 301), el uso cada vez más creciente de sistemas de diálogo, mediante los que una persona puede acceder a un servicio telefónico automático para obtener un determinado tipo de información o realizar una transacción, hace que cuestiones tales como la constitución de recursos para desarrollarlos sean las más abordadas en la actualidad en el campo de las tecnologías del habla. Para el AC son especialmente útiles los corpus "persona-persona", que intentan recoger una muestra amplia y realista de la situación comunicativa propia de la aplicación, en la que se establecen estrategias de gestión del diálogo en función del comportamiento natural observado en un determinado servicio. En esta tipología de corpus conversacionales encontramos:

m. El corpus CALLHOME, de Alexandra Canavan y George Zipperlen. Consiste en 120 conversaciones telefónicas entre miembros de la familia o amigos de área caribeña de una duración de más de 30 minutos. Las transcripciones y los audios no tienen acceso libre (Canavan – Zipperlen 1996a).
n. El corpus CALLFRIEND, de Alexandra Canavan y George Zipperlen. Consiste en 60 conversaciones telefónicas entre miembros de la familia o amigos de países de Caribe (Puerto Rico, República Dominicana) de 5-30 minutos de duración. Contiene documentación sobre los hablantes (edad, sexo, educación). Las transcripciones y los audios no tienen acceso libre (Canavan – Zipperlen 1996b).
o. El Corpus DIME (Diálogos Inteligentes Multimodales en Español), 26 diálogos multimodales del español de México que cuenta tanto con la señal sonora como con la filmación de acciones que tienen lugar delante de la pantalla de un ordenador en un entorno para el diseño de cocina. Cada diálogo consiste en la solución de un problema de diseño propuesto con la técnica del mago de Oz a través del mismo escenario, resuelto en la conversación (Pineda *et al.* 2002).

p. El Corpus Ferrovial, recogido en el proyecto BASURDE, que desarrolla un sistema de diálogo oral en un dominio restringido: la información sobre viajes en ferrocarril. Consiste en 204 diálogos (Bonafonte *et al.*, 2000).

Este tipo de corpus más limitados facilitarían el control de algunas variables comunicativas, por lo que podrían ser muy útiles en el ámbito de análisis de la conversación. Desgraciadamente, algunos de ellos se han elaborado para necesidades de empresas y por este motivo no son fácilmente accesibles. En otros casos, se han creado en el marco de proyectos de I+D que han contado con fuentes de financiación pública españolas o europeas. En ese caso debería ser posible alcanzar acuerdos para su uso sin fines comerciales.

3. Corpus conversacionales y análisis contrastivo de la conversación

Los microcorpus paralelos, multilingües y monolingües, que ilustraremos a continuación pueden favorecer el análisis contrastivo de la conversación, ya que en ellos también es posible aislar variables comunes como el tema, la finalidad comunicativa, etc. Nacen en diferentes contextos de estudio: enseñanza de segundas lenguas, traducción automática, etc. Distinguiremos según el medio de elicitación entre corpus paralelos espontáneos y semi-espontáneos. Entre los primeros encontramos:

q. AKSAM (Aktivitetstyper och samtalsstruktur hos L1-och L2-talare av spanska - Activity types and conversational structure in L1 and L2 users of Spanish). Bajo la dirección de Lars Fant, se creó este corpus compuesto por grabaciones de negociaciones entre empresarios, por una parte, y estudiantes universitarios, por otra. Contiene también material auténtico en sueco. No es accesible actualmente.
r. Corpus en agencias de viajes. Se trata de dos corpus de conversaciones auténticas en agencias de viajes, uno en español y otro en francés, que no forman parte del mismo proyecto de recogida pero cuyo uso contrastivo está dando lugar a fructíferas investigaciones sobre estilos comunicativos. El corpus español se compone de diez interacciones, pertenecientes a dos series: una recogida en Valencia por Josefa Contreras (Contreras 2005) y otra en Málaga por Federica Nonelli (esta última no está publicada). El corpus francés – Lancom, LAngue et COMmunication - es un corpus dedicado a la didáctica del francés como L2 y contiene una sección dedicada a textos

auténticos francófonos en agencias de viajes con una duración de 5 horas. Se poseen los archivos sonoros. En este momento no es posible consultarlo pues está en fase de transmigración al sitio Ortolang.

s. ClInt **(**Clinical Interview) es un corpus oral bilingüe español-catalán con quince horas de entrevistas clínicas dirigido por M. Antònia Martí. Posee archivos sonoros alineados con las transcripciones. Es el resultado de un proyecto financiado por el Ministerio de Educación e Innovación (FFI2009-06252-E/FILO) y, por lo que sabemos, es el primer corpus de este tipo en español. Forma parte de un proyecto más amplio, el Text-knowledge 2.0, cuyo objetivo es el estudio del uso lingüístico. En este marco el equipo está desarrollando varios corpus representativos de diferentes situaciones comunicativas. (Llisteri 2010: 109).

Dentro de los corpus semiespontáneos ilustraremos algunos pertenecientes a la tipología *task-oriented*. Este tipo de textos consiste en interacciones dialógicas de tema no libre, elicitadas y orientadas a la realización de una tarea: reconstrucción de un recorrido en un mapa (*Map-task*, Anderson *et al.* 1991), consulta de una agenda de citas (*Schedule corpus*), el juego de encontrar diferencias entre distintos dibujos (*Encuentra la diferencia*, Péan *et al.* 1993), etc. Estas técnicas de elicitación sirven para obtener interacciones que se caracterizan por un alto grado de espontaneidad en la realización fonético-prosódica y discursiva, manteniendo bajo control otras variables, como el contexto situacional y pragmático. Hasta donde nosotros sabemos, en español se han recogido escasos corpus de esta tipología dialógica:

t. El primer corpus paralelo del que tenemos noticia es un corpus de nueve conversaciones *task-oriented* que forman parte del *Interactive Systems Lab scheduling corpus*, recogido en la Carnegie Mellon University en varios idiomas. Se trata de conversaciones entre dos hablantes que han de encontrar un momento adecuado para encontrarse. Están agrupadas por sexo (tres mujeres, tres hombres y tres mixtas). Los datos son propiedad del Interactive Systems Lab, dirigido por Alex Waibel y tienen acceso restringido. Se ha usado en investigaciones sobre cohesión y coherencia en el discurso conversacional (Taboada 2001).

u. El Corpus *Pratid nelle lingue europee*. Se trata de un corpus recogido como parte de un proyecto de análisis pragmático de diálogos *task-oriented* en algunas de las principales lenguas europeas (español, inglés, francés, alemán y portugués). En este momento son de acceso libre doce diálogos *task-oriented* del tipo “Encuentra las diferencias”. Se pueden consultar los audios

y las transcripciones de cuatro diálogos en español. El Proyecto está dirigido por Renata Savy y ha dado lugar a numerosos trabajos de investigación en ámbito pragmático y prosódico. Para el contraste con el italiano es posible utilizar este mismo tipo de textos recogidos en el corpus CLIPS. [http://www.parlaritaliano.it/index.php/it/corpora-di-parlato/672-corpus-pratid-nelle-lingue-europee - Fecha de consulta 2.09.2018].

v. El corpus Diálogos en Español (DiESPA) está constituido por dieciséis diálogos *task- oriented* ("Encuentra las diferencias") en distintas variedades peninsulares del español. Se pueden consultar en acceso abierto los audios y las transcripciones anotadas de cuatro diálogos de Barcelona, tres de Sevilla y uno de Almería anotados pragmáticamente. En este número de la revista es posible consultar un trabajo de investigación basado en este corpus (Alfano *et al.* 2018), dirigido, como el anterior, por Renata Savy. [http://www.parlaritaliano.it/index.php/it/corpora-di-parlato/792-corpus-diespa-dialogos-en-espanol - Fecha de consulta 2.09.2018].

w. El corpus *Glissando* contiene un subcorpus español / catalán de 28 hablantes por idioma y más de 25 h. de duración con tres tipos de peticiones *task-oriented*: a) información de viaje, b) información para un intercambio en un curso universitario y c) información para un recorrido turístico. La primera consiste en una conversación telefónica entre un operador y un cliente que desea información sobre precios y horarios de un recorrido específico; la segunda tiene lugar entre un funcionario administrativo de la universidad y un estudiante que le solicita información y la tercera es un diálogo inspirado en los *Map Task*, aunque se desarrolla de forma diferente. En los corpus *Map Task*, los hablantes deben cooperar para reproducir en el mapa del *follower* la ruta que está impresa en mapa del *giver* y el éxito de la comunicación se cuantifica por el grado de coincidencia de las dos rutas. Sin embargo, en estas conversaciones uno de los hablantes desempeña el papel de alguien que está planificando un viaje a Corfú y llama a un colega que vivió cinco años en Grecia. No se ha de reproducir ninguna ruta y se supone que entre los dos colegas hay una cierta familiaridad. Las transcripciones y el audio no son de libre acceso (Garrido *et al.* 2013).

Por último, citaremos, dentro de los corpus útiles en AC, los corpus de transmisiones televisivas o radiofónicas con transcripciones incorporadas, entre los que encontramos, por nombrar un ejemplo, el corpus de Investigación en Español de México del Posgrado de Ingeniería Eléctrica y Servicio Social (CIEMPIESS - UNAM) a cargo de Carlos Mena, Daniel Hernández y Abel Herrera. Este corpus contiene 18 h. de transmisiones de discurso radiofónico, de

acceso libre actualmente [http://www.ciempiess.org/downloads - Fecha de consulta 2.09.2018].

En la siguiente tabla proponemos un resumen de las características más relevantes de cada corpus:

Tabla 1. Características de los corpus orales conversacionales

Corpus	**Transcripciones**	**Audio**	**Alineamiento**	**Etiquetado**	**Disponible en línea gratuitamente**	**En comercio**
Corpus Val.Es.C	sí	consulta en sede	no	no	sí	
ALCORE	sí	no	no	no	sí	
Corpus del español conversacional de Barcelona y su área metropolitana	sí	no	no	no	no	sí
COLA	sí	sí	sí	no	sí	
C-Or-DiAL	sí	sí	sí	sí	sí	
COGILA	sí	sí	no	no	no	sí
CORLEC	sí	no	no	no	sí	
C-ORAL-ROM	sí	sí	sí	no	no	sí
CREA	sí	no	no	no	sí	
CORPES XXI	sí	en parte	en parte	no	sí	
Corpus El Grial	sí	no	no	no	sí	
Corpus del español	sí	no	no	no	sí	
CALLHOME	sí	sí	sí	no	no	sí

CALLFRIEND	sí	sí	sí	no	no	sí
DIME	sí	sí + vídeo	sí	no	no	sí
Corpus Ferrovial	sí	sí	sí	no	no	sí
AKSAM (*Aktivitetstyper och samtalsstruktur hos L1-och L2-talare av spanska*	sí	sí	no	no	no	-
Corpus en **agencias de viajes**	sí	sí (en parte)	no	no	a disposición próximamente	sí
ClInt (***Clinical Interview***)	sí	sí	sí	no	no	sí
Interactive Systems Lab scheduling corpus	sí	sí	sí	no	no	sí
Pratid nelle lingue europee	sí	sí	no	sí	sí	
DiESPA	sí	sí	no	sí	sí	
Glissando	sí	sí	sí	no	no	sí
CIEMPIESS - UNAM	sí	sí	sí	no	sí	

Este elenco de corpus, que no tiene la intención de ser exhaustivo, nos da la posibilidad de trabajar en el ámbito del análisis del discurso siguiendo un método preciso y sistemático. Sin duda, la disponibilidad de datos en cierto sentido pre-analizados que nos proporcionan algunos de los corpus descritos permite un rápido análisis cuantitativo a través de la interrogación automática, ya sea prosódica, morfosintáctica o pragmática. A partir de una primera criba, que

facilita observaciones y la elaboración de descripciones de estrategias sobre bases estadísticas o datos cuantitativos, se pueden llevar a cabo análisis más detallados de tipo cualitativo.

Referencias

Alfano, Iolanda, Savy, Renata, Sbranna, Simona & Schettino, Loredana 2018. Strategie discorsive in spagnolo L1 ed L2 a confronto: un'indagine su corpora dialogici. *Chimera* 2018.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., *et al.* 1991. The hrc map task corpus. *Language and Speech* 24: 351–366.

Azorín Fernández, M. D. 2002. *Corpus Alcore (Alicante Corpus oral de español)*. Alicante: Universidad de Alicante, 1 disco (CD-ROM).

Barros García, P., COGILA. 2012. *Español oral conversacional. Corpus y guía didáctica*. Granada: Editorial Universidad de Granada.

Briz Gómez, A. & Grupo Val.Es.Co. 2002. *Corpus de conversaciones coloquiales*. Madrid: Arco Libros.

Briz Gómez, A. & Albelda Marco, M. 2009. Estado actual de los corpus de lengua española hablada y escrita: I+D, *Anuario 2009*, Istituto Cervantes.

(https://cvc.cervantes.es/lengua/anuario/anuario_09/briz_albeida/p06.htm – Fecha de consulta 2.09.2018).

Canavan, A. & Zipperlen G. 1996 a. *CALLHOME Spanish Speech*. Philadelphia: Linguistic Data Consortium.

Canavan, A. & Zipperlen G. 1996 b. *CALLHOME Spanish Non-Caribbean Dialect*. Philadelphia: Linguistic Data Consortium.

CLIPS (Corpora e Lessici dell'Italiano Parlato e Scritto) (www.clips.unina.it – Fecha de consulta 2.09.2018).

Contreras, J. 2005. *El uso de la cortesía y las sobreposiciones en las conversaciones. Un análisis contrastivo alemán-español*. Valencia: Universitat de València, CD-Rom.

Cresti, E. & Moneglia, M. 2005. *C-ORAL-ROM integrated reference corpora for spoken romance languages.* Amsterdam/Philadelphia: John Benjamins Studies in Corpus Linguistics 15. CD-Rom.

Fant L., *et al*. 1996. *Korpus AKSAM (Aktivitetstyper och samtalsstruktur hos L1-och L2-talare av spanska)*, Estocolmo: Departamento de Español y Portugués, Universidad de Estocolmo.

Garrido, J.M., Escudero, D., Aguilar, L. *et al*. 2013. Glissando: a corpus for multidisciplinary prosodic studies in Spanish and Catalan. *Lang Resources & Evaluation*, 47: 945–971.

Hernández Alonso, C. 2009. *Estudios lingüísticos del español hablado en América*, 4 vol. Madrid: Visor.

Llisterri, J. 2018. *La lingüística de corpus*. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona. (Fecha de consulta 10.09.2018).

http://liceu.uab.cat/~joaquim/language_resources/lang_res/linguistica_corpus.html

Llisterri, J., Machuca, M. J., de la Mota, C. *et. al*. 2005. Corpus orales para el desarrollo de las tecnologías del habla en español. *Oralia*, 8: 289-325.

Nicolás Martínez, C. 2012. *C-Or-DiAL (Corpus Oral Didáctico Anotado Lingüísticamente)*. Madrid: Liceus.

Péan V., Williams S. & Eskénazi M. 1993. The design and recording of ICY, a corpus for the study of intraspeaker variability and the characterisation of speaking styles. *Eurospeech'93. 3rd European Conference on Speech Communication and Technology*. Berlin, Germany, 21-23 September 1993. CD: Vol. 1: 627-630.

Pineda Cortés, L.A., Massé Márquez, A., Meza, I. *et al.*. 2002. The DIME Project. *Proceeding MICAI '02 Proceedings of the Second Mexican International Conference on Artificial Intelligence: Advances in Artificial Intelligence*, April 22 - 26, 2002, London: Springer-Verlag: 166-175.

Rojo, G. Los corpus textuales del español. 2016. En Gutiérrez-Rexach, J. (ed.). *Enciclopedia lingüística hispánica*. Oxon: Routledge: 285-296.

Taboada, M. 2001. *Collaborating through Talk: The Interactive Construction of Task-Oriented Dialogue in English and in Spanish*, Ph.D. dissertation. Madrid: Universidad Complutense.

Vila Pujol, Mª R. 2001. *Corpus del español conversacional de Barcelona y su área metropolitana*. Grupo GRIESBA. Barcelona: Edicions Universitat de Barcelona.

Vila, M., González, S., Martí, M. A. *et al.*. 2010. ClInt: a Bilingual Spanish-Catalan Spoken Corpus of Clinical. *Procesamiento del Lenguaje Natural*, 45, septiembre 2010: 105-111.