

APORTES METODOLÓGICOS DE UN CORPUS ORAL DIAFÁSICO

Bernardo E. Pérez Álvarez

*Universidad Michoacana de San Nicolás de Hidalgo*

perezbernardo en hotmail com

Resumen

El artículo tiene como objetivo central mostrar algunas ventajas metodológicas de un corpus oral cuando se pueden establecer algunas comparaciones entre discursos de diversas situaciones de comunicación. El análisis está basado en el *Corpus Michoacano del Español*, que contiene datos sistemáticos orales de diversidad de contextos, y ahora cuenta ya con más de 40 hrs. de grabación y transcripción disponibles para el análisis lingüístico oral. Para la elaboración de este corpus se ha seguido la propuesta de Koch y Oesterreicher (1985, 2001, <sup>2</sup>2011), que permite integrar la variación lingüística desde los ejes de la inmediatez y la distancia comunicativa y con una distinción medial entre el canal fónico y el gráfico. Una ventaja de este corpus es que posibilita no sólo el análisis de textos orales de manera general con metodologías cualitativas, sino que permite también establecer comparaciones entre textos orales de diferentes registros discursivos, lo que posibilita el estudio diferenciado de ciertos fenómenos, incluso con metodologías cuantitativas.

Bernardo E. Pérez Álvarez. 2015.  
Aportes metodológicos de un corpus oral diafásico  
*Círculo de Lingüística Aplicada a la Comunicación* 64, 148-168.  
<http://www.ucm.es/info/circulo/no64/perez.pdf>  
<http://revistas.ucm.es/index.php/CLAC>  
[http://dx.doi.org/10.5209/rev\\_CLAC.2015.v64.51282](http://dx.doi.org/10.5209/rev_CLAC.2015.v64.51282)

© 2015 Bernardo E. Pérez Álvarez  
*Círculo de Lingüística Aplicada a la Comunicación (clac)*  
Universidad Complutense de Madrid. ISSN 1576-4737. <http://www.ucm.es/info/circulo>

Estas ventajas metodológicas se ejemplifican y discuten en la segunda parte. En particular, se trata de revisar algunos fenómenos como la deixis espacial, o de ciertas estructuras conectivas en el discurso oral. También se mostrarán formas que, siguiendo una estructura monológica formal que tiende a una concepción escrita del texto, entran en conflicto con la capacidad cognitiva de la memoria de trabajo de los hablantes, por lo que se recurre a usos paratácticos de construcciones comúnmente hipotácticas en la escritura.

La selección de diferentes discursos en el arco de tensión entre inmediatez y distancia comunicativas proporciona una visión integradora de la variación lingüística, bajo el presupuesto de que una lengua y su variación sólo pueden ser estudiados desde los discursos emitidos.

Palabras clave: Discurso oral, lingüística de corpus, diafasia, variación

#### Abstract

Methodological contributions of a diaphasic oral corpus. This paper shows some of the methodological advantages of an oral corpus which allows for a comparison of different communicative situations. Our analysis is based on the *Corpus Michoacano del Español*, which contains oral data of more than 40 hrs. of recording and transcription, systematically collected following Koch and Oesterreicher's (1985, 2001) model of language variation within the two poles of communicative immediacy and distance. An advantage of this corpus is that it facilitates a highly differentiated, qualitative as well as quantitative analysis of oral registers and "texts".

These and other methodological benefits are exemplified and discussed in the second part of the paper, in particular with respect to spatial deixis and some type of connective structures. Another point is to show that these and other structures typical of "written" monologic language, such as hypotaxis will be replaced by paratactic structures in oral language due to cognitive restrictions of the working memory, whenever there is a need to provide more information.

The selection of different types of oral discourse situated on the axis between the two poles of communicative immediacy and distance argues for an integrated view of linguistic variation under the premise that variation can only be studied in "real life discourse".

Key words: Oral language, corpus linguistics, diafasia, variation

## Índice

1. Introducción, 150
2. Posibilidades y limitaciones de un corpus oral del español, 151
  - 2.1. Algunos corpus con datos orales en español, 152
  - 2.2. La elaboración de un corpus con variación pragmático-discursiva del habla, 154
3. Variación en el discurso oral, 157
  - 3.1. Los géneros en el discurso oral, 157
  - 3.2. Tensiones entre diferentes planos de producción lingüística, 159
  - 3.3. Procesamiento cognitivo limitado, 162
  - 3.4. Interlocutores y el diseño para el receptor, 163
  - 3.5. La contextualización desde la variación discursiva, 164
4. Conclusiones, 165
5. Bibliografía, 166

### 1. Introducción

El presente artículo tiene como objetivo central mostrar algunas ventajas metodológicas de un corpus oral cuando se pueden establecer algunas comparaciones entre discursos de diversas situaciones de comunicación concebidos desde el medio oral.

El análisis está basado en el *Corpus Michoacano del Español*, que contiene datos sistemáticos orales de diversidad de contextos, y ahora cuenta ya con más de 40 hrs. de grabación y transcripción disponibles para el análisis lingüístico oral. Para la elaboración de este corpus se ha seguido la propuesta de Koch y Oesterreicher (1985, 2001), que permite integrar la variación lingüística desde los ejes de la inmediatez y la distancia comunicativa y con una distinción medial entre el canal fónico y el gráfico. De acuerdo con estos autores, todos los discursos realizables en una lengua específica podrían ser integrados en este doble eje de variación. El *Corpus Michoacano del Español* se centra en el canal fónico, sin olvidar que aún en este nivel, los discursos específicos tienen diferentes niveles de planificación y de determinación comunicativa, es decir, su organización está orientada por parámetros pragmadiscursivos variables según la situación comunicativa en la que se producen. Una ventaja de este corpus es

que posibilita no sólo el análisis de textos orales de manera general con metodologías cualitativas, sino que permite también establecer comparaciones entre textos orales de diferentes registros discursivos, lo que posibilita el estudio diferenciado de ciertos fenómenos, incluso con metodologías cuantitativas.

Estas ventajas metodológicas se mostrarán en la segunda parte del artículo. En particular, se trata de revisar algunos fenómenos como la deixis espacial, o de ciertas estructuras conectivas en el discurso oral. También se mostrarán formas que, siguiendo una estructura monológica formal que tiende a una concepción escrita del texto, entran en conflicto con la capacidad cognitiva de la memoria de trabajo de los hablantes, por lo que se recurre a usos paratácticos de construcciones comúnmente hipotácticas en la escritura.

La selección de diferentes discursos en el arco de tensión entre inmediatez y distancia comunicativas proporciona una visión integradora de la variación lingüística, bajo el presupuesto de que una lengua y su variación sólo pueden ser estudiados desde los discursos emitidos.

## 2. Posibilidades y limitaciones de un corpus oral del español

En la actualidad discutir sobre la necesidad de realizar estudios del discurso oral puede parecer ocioso. Sin embargo, estudiar el discurso oral como un conjunto sistemático con variación interna es un tema menos desarrollado. Apenas hace algunos años que empieza a reconocerse la necesidad de algo más que estudiar el discurso oral como otro registro dentro de la amplia gama de registros escritos. Sin embargo, no todos los materiales de discurso oral disponibles en la actualidad permiten el mismo tipo de obtención de datos y de análisis de la información con la que cuentan.

### 2.1. Algunos corpus con datos orales en español

Una carencia fundamental en muchos casos es la ausencia de la fuente oral de la cual se obtuvo la transcripción, aparte de que no se especifica el sistema mismo de transcripción. Si bien esta limitación imposibilita la realización de estudios por ejemplo de prosodia, no me centraré en este tipo de carencias, sino en otras que atañen a la variación lingüística en sus diferentes niveles. Para este fin, es necesario contar con

corpus orales diferenciados por nivel de registro que permitan estudiar la variación interna del discurso oral de manera sistemática.

El *CREA* cuenta con la posibilidad de seleccionar en el tipo de texto la opción oral, sin más detalles. Cuando se selecciona alguna entrada para realizar búsquedas en el registro oral, se puede entonces comenzar a revisar los discursos de donde provienen los ejemplos, al confrontar este eje en correlación con la zona geográfica, se puede observar que se tienen disponibles materiales muy dispares como para poder realizar un estudio más sistemático, es decir, si bien es posible reconocer diferentes tipos de textos orales registrados en dicho corpus, sólo es posible hacerlo al revisar caso por caso, y no como un filtro dentro del corpus. Por otro lado, este avance muestra diferencias notorias en las regiones tradicionales reconocidas por las Academias: mientras que para el español de España hay una mayor cantidad de registros, para México por ejemplo apenas se registran tres fuentes: versiones estenográficas de la cámara de senadores, entrevistas al expresidente Vicente Fox y en un mínimo porcentaje algunas entrevistas particulares, de las que ni siquiera es posible reconocer la fuente o el año de la entrevista. Esta carencia es aún mayor si se selecciona Argentina, que sólo tiene entrevistas y discurso del senado, o Venezuela, que sólo cuenta con entrevistas, por hablar de algunos de los países representativos de regiones lingüísticas y por el trabajo sobre discurso oral que ha sido realizado por investigadores de esos lugares.

En el caso del *CORPES XXI*, que aún se encuentra en etapa de prueba según se indica en la propia página de la Real Academia, sólo se distinguen algunas categorías bajo el rubro “medio”, entre ellas “oral”, pero en las búsquedas más específicas disponibles (tema y tipología) no aparece ningún rubro que permita distinguir entre tipos de texto oral. Tal parece que las carencias de este corpus serán las mismas del *CREA*.

El *Corpus del Español* de Mark Davies tiene también la posibilidad de elegir el rubro “oral” como criterio de búsqueda. Sin embargo, al igual que el *CREA*, los registros orales con los que se cuenta están basados en entrevistas, y exceptuando a España, país para el que se incluyen entrevistas del periódico *ABC* y algunas otras entrevistas, se trata de los registros del *Habla Culta* en ciudades hispanohablantes más algunas entrevistas a funcionarios públicos en turno, en ocasiones tomadas de las mismas

páginas de los partidos políticos. Es decir, no cuenta con registros suficientes que permitan realizar un estudio de variación diafásica.

Corpus orales formados con fines más específicos suelen tener carencias similares, en todo caso en ellos se muestra claramente cómo ha sido determinante la perspectiva teórica y metodológica para su formación. Aquellos que han seguido una metodología de carácter sociolingüístico toman como principio la selección de los hablantes, con la finalidad de controlar las variantes sociales a las que una persona está circunscrita. En estos casos resulta relevante su nivel de estudios o de ingresos, más que la diversidad de situaciones de comunicación en las que puede llegar a participar. Bajo esta metodología, resulta interesante el trabajo seguido por numerosos proyectos, entre ellos el *PRESEA*, el más reciente y abarcador de este tipo de trabajos. Sin embargo, si se utiliza este corpus con la finalidad de observar variación entre situaciones monológicas formales y conversaciones entre varios hablantes, o para realizar estudios sobre usos deícticos en situaciones cotidianas de comunicación, este tipo de corpus suele presentar limitaciones específicas.

Este parece ser el caso dominante de numerosos corpus orales, formados en su gran mayoría por entrevistas, o bien aquellos formados por discursos específicos de acuerdo con el objeto de estudio, como discurso político, escolar, o en menor medida de ambientes profesionales o laborales, o en todo caso con una selección de situaciones cotidianas de comunicación.

Si bien estos corpus que han permitido estudios específicos nos ayudan a tener una visión cada vez más clara de las diferencias entre el discurso oral parlamentario, o el discurso en el salón de clases y en ambientes académicos, o la conversación coloquial, aún es difícil poder realizar estudios más sistemáticos de conjunto que permitan poner en relación textos con diferentes niveles de variación, aún cuando han logrado acercarse a un ámbito de la lengua hasta antes poco estudiado (ver por ejemplo la discusión al respecto en Briz 2010). La procedencia de los materiales y de los mismos proyectos de investigación hace difícil una visión de conjunto sobre la variación dentro del discurso oral, y las posibilidades de análisis comparativos se dificultan ante la heterogeneidad de materiales y métodos de registro de los mismos.

En síntesis, en la actualidad en español es aún difícil encontrar corpora homogéneos en sus materiales y metodología de obtención que permitan un estudio diferenciado de la variación pragmático-discursiva en el nivel oral.

## 2.2. La elaboración de un corpus con variación pragmático-discursiva del habla

Los estudios sobre la lengua oral han recorrido un largo camino después de que Saussure declarara que el objeto de estudio de la lingüística debería ser la lengua y no el habla. Una demostración sistemática sobre la necesidad de replantear esta dicotomía saussureana lo constituye el trabajo de Eugenio Coseriu (por ejemplo Coseriu 1992, entre otros). Estudios más recientes han reconocido la necesidad de estudiar la lengua oral de manera más sistemática como recurso para la descripción lingüística, más allá de la discusión misma que se presenta en el análisis del discurso como ámbito de estudio ínter y transdisciplinario, desde la perspectiva específicamente lingüística se puede reconocer una serie de problemáticas, tanto a nivel gramatical (Cumming & Ono 2000, Hopper 2012) como discursivo contextual (Fox 2008), que permiten elaborar una serie de interrogantes susceptibles de ser respondidas en gran medida atendiendo a la variación de la lengua oral. Las posibilidades para estudiar una gramática emergente (Hopper 2012) surgida del discurso oral, o bien el análisis de factores pragmáticos centrales en la constitución de un texto, como son la influencia directa del emisor y el receptor en un discurso oral no planificado, requieren de corpus orales específicos que faciliten este tipo de acercamientos.

Uno de los problemas teóricos y metodológicos a resolver en el estudio de la variación pragmático-discursiva de la oralidad se presenta al momento de establecer las posibilidades de registro de datos que alimentarán a un corpus de trabajo, es decir, qué tipos de grabaciones y en qué situaciones comunicativas son obtenidas, son aspectos a considerar a fin de contar con un panorama amplio de la variación oral.

Koch y Oesterreicher (1985) elaboran una propuesta interesante que rompe con la tradicional dicotomía entre comunicación formal e informal para verla más bien como un continuo entre dos polos, lo cual les permite incorporar diferentes ejes de variación entre estos dos polos, como la dialogicidad o la selección léxica entre otros parámetros (Koch y Oesterreicher 2001). Esta perspectiva permite resolver la noción de género

discursivo como un fenómeno difuso constituido por haces de relaciones que se conjugan en cada discurso oral concreto, más que por características estructurales que lleven hacia una tipología de textos. Al considerar parámetros pragmáticos como una constelación comunicativa que se conjunta en cada discurso específico, el acento debe ponerse en el nivel de variación en el que se encuentran estos parámetros, por ejemplo un eje de relación que va de la dialogicidad a la monologicidad debe entenderse como un continuo específico, que se concreta en una entrevista como una monologicidad media alta, o en un sermón en la iglesia como una monologicidad alta, o en una conversación espontánea entre familiares como una dialogicidad alta. Estos parámetros específicos se combinan siempre en este tipo de ejes, de tal manera que lo que desde una perspectiva estructural podría resultar complejo, en el caso de un corpus oral constituye una ventaja metodológica puesto que permite analizar la variación de un fenómeno específico desde un haz de relaciones. Ahora no me detendré en la discusión teórica de los alcances de esta concepción de la variación lingüística, se puede seguir la ruta de esta discusión en trabajos como el de López Serena (2007), con una amplia bibliografía citada que refiere a los trabajos, en español y alemán, de los autores que elaboran esta perspectiva de trabajo.

El *Corpus Michoacano del Español*, en su segunda fase de recopilación de datos, contiene datos sistemáticos orales de diversidad de contextos, y ahora cuenta ya con más de 45 hrs. de grabación y transcripción disponibles para el análisis lingüístico oral. Para la elaboración del corpus se realizaron grabaciones de audio que luego fueron transcritas con el sistema de notación del Análisis Conversacional, en particular con la propuesta de Antaki (2002). Por su fácil acceso, se inició con grabaciones de programas radiofónicos de diversa índole, como noticieros, programas de micrófono abierto con llamadas de radioescuchas, consultas médicas vía telefónica, debates políticos, entrevistas de radio. Este material se ha complementado posteriormente con grabaciones en vivo, como sermones en la iglesia, grabaciones en situaciones coloquiales entre familiares o amigos en diversidad de contextos, como la escuela, la casa, algún café. De esta manera, se ha logrado establecer un primer conjunto de grabaciones que abarcan desde la inmediatez comunicativa, es decir, conversaciones entre varios interlocutores, no planificadas, con temáticas diversificadas, hasta grabaciones de la distancia comunicativa, monológicas como un sermón de iglesia, con un sólo hablante que se



dirige a una audiencia en una situación formal de una ceremonia religiosa, de tal manera que los parámetros que proponen Koch y Oesterreicher (1985) se abarcan en diversos momentos del continuo entre inmediatez y distancia comunicativa. Los parámetros mencionados por Koch y Oesterreicher (1985) son los siguientes:

|  |   |
|--|---|
| 1 Comunicación privada                       | Comunicación pública 1                    |
| 2 Interlocutor íntimo                        | Interlocutor desconocido 2                |
| 3 Emotividad fuerte                          | Emotividad débil 3                        |
| 4 Anclaje accional y situacional             | Desconexión accional y situacional 4      |
| 5 Anclaje referencial dentro de la situación | Desconexión referencial de la situación 5 |
| 6 Copresencia espacio-temporal               | Separación espacio-temporal 6             |
| 7 Cooperación comunicativa intensa           | Cooperación comunicativa mínima 7         |
| 8 Diálogo                                    | Monólogo 8                                |
| 9 Comunicación espontánea                    | Comunicación preparada 9                  |
| 10 Libertad temática                         | Fijación temática 10                      |
| Etc.   | Etc.                                      |

No se trata de una lista exhaustiva, de ahí que al final de la lista estos autores añadan “Etc.” Además, debe entenderse que en cada parámetro mencionado se incluyen los dos polos de oposición dentro del continuo funcional que proponen. Así, entre la comunicación privada y la pública existen casos de comunicación en un ámbito intermedio como puede ser la comunicación en un salón de clases, que no corresponde con la comunicación privada de una pareja, pero tampoco puede catalogarse como pública a la par de un discurso político transmitido por un medio de comunicación masivo. Lo mismo puede señalarse en el parámetro de la dialogicidad: una entrevista radiofónica a un artista plástico que se registra en el *Corpus Michoacano del Español* no puede catalogarse como dialógica en la misma medida que una conversación informal en una familia mientras preparan la comida, pero tampoco puede considerarse monológica en virtud de que participan dos interlocutores con turnos de habla largos.

Estos parámetros en combinación entre ellos constituyen un conjunto de relaciones específico de cada acto comunicativo, e incorporan diversas marcas lingüísticas que pueden ser analizadas en diferentes perspectivas.

### 3. Variación en el discurso oral

Es necesario aclarar que el concepto mismo de variación conlleva una concepción de continuo y traslape, a diferencia de las distinciones categoriales claras y distintas. De esta manera, es posible analizar el funcionamiento discursivo de una unidad lingüística de una manera diferenciada a partir de su comportamiento en diferentes estadios dentro del continuo, por ejemplo para comprender el nivel de integración o agregación sintáctica de una oración de relativo, o bien el nivel de integración semántica de un deíctico discursivo dentro de la continuidad temática del discurso.

Algunos planos de análisis que pueden ser considerados son los géneros dentro del discurso oral, la tensión que se genera entre diferentes planos de producción lingüística, la limitaciones del procesamiento cognitivo de acuerdo con la memoria de trabajo, la dinamicidad de la comunicación de acuerdo a la participación de los interlocutores, o bien la regulación entre la información dada en el discurso y las inferencias obtenidas de la situación de comunicación.

#### 3.1. Los géneros en el discurso oral

La producción de discursos orales está sujeta a diferentes tipos de requerimientos, que en ocasiones se ponen en tensión. Si partimos del principio de que “el discurso [...] es el ámbito apropiado para estudiar las gramáticas de las lenguas del mundo, por ser no sólo el lugar donde la gramática se pone en uso, sino también la fuente a partir de la cual se forma o ‘surge’ la gramática” (Cumming y Ono 2000: 171), es necesario contar con un panorama más completo de la producción de discursos orales, para entender que se trata de algo más que un género oral homogéneo que debe estudiarse en contraste con los géneros escritos. Si “la gramática se origina en los patrones recurrentes en el discurso, y estos patrones continuamente la configuran” (Cumming y Ono 2000: 171), es necesario aceptar que una gramática del discurso que considere a la oralidad como un género homogéneo está perdiendo la segunda parte de esta premisa, y no podrá dar cuenta adecuadamente de los patrones que configuran a la gramática.

El estudio de las marcas lingüísticas que aparecen en los discursos orales pueden seleccionarse como variables de análisis diferenciado en grupos distintos de textos

orales dentro de un corpus. Es posible dividir un corpus de manera genérica, cuando menos en dos tipos de textos orales, con la finalidad de observar internamente si hay tendencias claras para el aprovechamiento de recursos específicos, por ejemplo para el mantenimiento de turno en situaciones formales, selección léxica, oraciones hendidas, proformas, etc. Este tipo de estudio variacionista puede ayudar a entender mejor procesos de producción textual vinculados tanto a valores pragmáticos como cognitivos de procesamiento en línea.

Es importante considerar que, en la discusión en torno a los géneros discursivos, se ha señalado la necesidad de un proceso inductivo que permita encontrar los parámetros o ejes de relación que determinan a un género, y no como un proceso deductivo que determina de manera previa una tipología, para luego analizar casos específicos para ver cómo se ajustan a esta tipología. Charaudeau (2012: 22) lo plantea de manera clara cuando afirma: “Un género, o un tipo, es una categoría determinada luego de un procedimiento inductivo, según las propiedades internas que caracterizan a ciertos objetos, y cuyas similitudes y diferencias permiten establecer agrupamientos y diferenciaciones.”

Este proceso inductivo debe complementarse con un análisis de diferentes planos, que vayan más allá de las marcas formales del discurso, para poder incorporar tanto a los interlocutores como a la situación misma de comunicación.

### 3.2. Tensiones entre diferentes planos de producción lingüística

Un eje relevante para el estudio de la variación sintáctica (Serrano 2006) lo constituye la tensión existente entre las relaciones paratáticas e hipotáticas, es decir, el nivel de integración sintáctica existente en un discurso específico, cuando se pone en relación con otros parámetros que influyen directamente sobre la estructura textual. Es preciso señalar que la distinción entre parataxis e hipotaxis es considerada aquí como una distinción de grado (Raible 1992, Givón 2001, Givón y Shibatani 2009), que debe analizarse en relación con el tipo de discurso utilizado. En el trabajo de Givón (1979, 2001) se establece una vinculación directa entre las estructuras paratáticas y el discurso cara a cara, es decir, oral no planificado, mientras que la hipotaxis requiere de un proceso más complejo de integración sintáctica que se asocia a discursos escritos,

planificados y con posibilidades de revisión sobre sí mismos, capacidad propia de la escritura. De manera complementaria también ha sido estudiada la relación existente entre la capacidad de procesamiento lingüístico en el discurso oral no planificado y las estructuras sintácticas resultantes, que tienden a desarrollarse en un nivel paratáctico aún cuando la situación comunicativa, la formalidad entre los interlocutores y la temática tratada tiendan a un uso marcado por la distancia comunicativa, por lo que la mezcla entre formas paratácticas e hipotácticas prevalece pero con variaciones relevantes en relación con otros discursos, bien escritos o bien no planificados de la inmediatez comunicativa (Chafe 1987, 1994; Kotschi 1996). Así, se genera una tensión entre el discurso oral de la distancia comunicativa (Koch y Oesterreicher 1985, 2001) que, por una parte, en su concepción escritural exige una alta elaboración tanto en el nivel de integración de la estructura sintáctica como en su selección léxica, pero por otra parte se ve limitado en sus capacidades de procesamiento cognitivo en línea, donde prevalece la necesidad de ilación discursiva en el nivel informativo y disminuye el control sobre las posibilidades estructurales de la integración informativa en el nivel hipotáctico y en la selección de recursos léxicos.

Un ejemplo ilustrativo que refleja esta tensión entre la estructura sintáctica y la producción discursiva oral lo constituyen los demostrativos neutros del español *esto*, *eso* y *aquello* (Pérez Álvarez, en prensa). Al tratarse de formas neutras, no pueden concordar con ningún sustantivo femenino o masculino en español, por lo que su carácter pronominal no es el prototípico de la relación entre una unidad que sustituye a un nombre. Estas unidades tienen un valor de deíctico discursivo, tal como lo entienden Diessel (1999) o Levinson (1983), más que como pronombres, e incluso, en su uso como adjetivos, concuerdan con la definición de deixis discursiva que señala Fillmore (1997), o bien a las anáforas conceptuales (González Ruiz 2009). Esta área de estructuración semántica posibilita la referencialidad endofórica entre segmentos textuales. La referencialidad endofórica conlleva una carga cognitiva marcada por la correlación entre las unidades referidas (anafórica o catafóricamente), la concordancia sintáctica de género y número, la continuidad temática que permiten en el discurso, y finalmente la función sintáctica que realizan las unidades que recuperan información previa. Se trata de varios factores que unidos componen unidades con una alta densidad informativa a ser procesada por un enunciador. Cuando se utiliza el medio escrito, es

posible revisar cada uno de estos factores por separado y realizar correcciones, sin embargo, cuando el procesamiento del discurso se da en línea con la verbalización, como es el caso del discurso oral de la distancia comunicativa, se presentan problemas en el nivel de la estructura, como puede ser la falta de concordancia o la omisión de preposiciones en los complementos circunstanciales, o bien se opta por la elección de formas con menor densidad informativa como pueden ser los demostrativos neutros, que permiten la continuidad temática y la referencialidad endofórica, pero de manera genérica.

Bajo estos presupuestos es posible entender mejor algunos ejemplos presentes en el *Corpus Michoacano del Español*:

- (1) J: el primero en América Latina y que bueno mucho de ese dinero desafortunadamente no se queda en México que **eso** es lo dramático lo lo peor del caso yo creo que si las empresas deberán de pensar y definir precios más accesibles para una una comunidad como la nuestra que tiene un poder adquisitivo muy bajo y que: eh pues no puede desprenderse de doscientos pesos (.). (23RM03-11-07\_CHARLA)

En (1) se pueden reconocer varios elementos propios del procesamiento oral como el marcador *bueno*, la repetición de *lo* y de *de*, así como el alargamiento de *que*. Junto a estas marcas, encontramos el uso de *eso* en una estructura de oración hendida que establece una calificación sobre la temática previamente enunciada. El tópico discursivo de la conversación es la piratería, pero en lugar de utilizar alguna forma nominal como este asunto o esta situación, que contienen ya una valoración sobre el tema discutido, se opta por la forma neutra más simple *eso* (Pérez Alvarez 2011, González Ruiz 2009).

Este recurso también se relaciona con la capacidad de mantenimiento de tópico cuando existen varios nombres que pueden ser recuperados anafóricamente, en el procesamiento en línea se reconoce lo que Chafe (1987, 1994) llama un tópico intermedio o semiactivo. Un demostrativo neutro tiene la posibilidad de mantener la continuidad de tópico en un nivel episódico, sin requerir de una concordancia en un nivel de cohesión local.

Gundel (1988) habla de tópicos relacionales, que no son necesariamente de carácter referencial exofórico, para referirse también a este tipo de relación en el nivel episódico. En algunos casos, es posible encontrar los dos niveles de referencialidad en una zona de Hay casos en que los dos ámbitos de referencialidad se traslapan, pues refieren tanto a un segmento discursivo pero que también constituye un acto de habla, es decir,

establecen a la vez una relación semántica (de cohesión) y pragmática (de modalización de lo dicho). Tal es el caso del siguiente ejemplo:

- (2) P6: a ver quién me agarra la palabra porque es mi propuesta doctor yo no puedo hablar por todo mundo yo hablo por mí esta propuesta es sobre (.) doctor (.) este: sobre del voto doctor por ejemplo (.) m: los candidatos que son ahorita que ya salieron mire uno es senador -l otro dejó la presidencia municipal a **eso** me refiero doctor que les pongan un hasta aquí (.) que terminen doctor (.) (05RM19-10-07\_OPINIÓN).

O bien:

- (3) yo creo que el candidato que nos tiene que gobernar es el que tiene el medicamento el que sabe como curar todititas esas enfermedades la corrupción este la inseguridad todititas esto y nadie habla de eso hablan de que si dar un kilo de tortillas doctor hay que enseñar a pescar a esa gente hay que educarla para que no tenga tantos hijos hay que educarla y hay que crear fuentes de empleo **eso** quisiéramos oír de ellos pero nos están diciendo cositas así como que dar un kilo de tortillas que también nos va a costar a todos que pagamos impuestos (05RM19-10-07\_OPINIÓN)

Como sucede con estas formas encapsuladoras, su capacidad sintetizadora les permite utilizarse en locuciones casi fijas, como sucede con *según esto*, que permite insertar un distanciamiento respecto a lo que se dice como proveniente de otro enunciador sin garantía de verdad:

- (4) pues nada más le pareció que los pobres le estábamos **según esto** ayudando entonces quiere decir que (05RM19-10-07\_OPINIÓN)

Y también en su carácter conectivo al combinarse con la preposición *por*:

- (5) esos eran programas -tonces la política era si el otro candidato no está de acuerdo refutar decir que no está de acuerdo **por esto** por aquello y lo demás pero jamás emplearon insultos estaban a la altura de la política (16RM27-10-07\_COMENTARIOS)

Los últimos dos ejemplos nos muestran ya casos que han entrado en un proceso de fijación, y dentro de un nivel paratáctico y pragmático.

### 3.3. Procesamiento cognitivo limitado

Existe una estructura que podría ser considerada una construcción que favorece la planificación discursiva, como sucede también con los alargamientos, las pausas y las muletillas, sin una función sintáctica específica en la medida en que puede ser simplemente omitida sin alterar el contenido proposicional de la oración donde se inserta:

- (6) diecinueve horas siete con siete minutos temperatura ambiente aquí en la zona centro veintidós grados centígrados así que entramos de lleno a **lo que vienen siendo** los detalles de la información el presidente de la canacindra Héctor Magaña Álvarez dijo que patronos exigen que el Seguro Social mejore la calidad de los servicios médicos que ofrecen (30RU12-10-07\_NOTICIAS)

La aparición de esta fórmula en discursos orales monológicos, como lo son los de la distancia comunicativa, permite afirmar que se trata de una construcción que permite ganar tiempo en la planificación del discurso, y muestra cómo la carga cognitiva del procesamiento requiere de tiempo para continuar con el hilo discursivo. Este fenómeno es más patente en el siguiente ejemplo, pues se acompaña de reformulaciones en torno al nombre utilizado para la continuidad topical:

- (7) en realidad como esquema es bueno pero desafortunadamente no está funcionando ya no podemos seguir este: siendo eh:: paternalistas vedá↑ en un sistema donde eh: la base que está haciendo los pagos para **lo que viene siendo** las las cuotas este las pensiones se siga: disminuyendo y **lo que viene siendo** la parte de pensionados siga creciendo verdad↑ y a muy corta edad se sigan este jubilando no es posible de esa forma ahora los sueldos con los que se jubila un empleado del Seguro son muy superiores a su sueldo real verdad↑ cuando terminan y un empleado que toda su vida ha estado trabajando (30RU12-10-07\_NOTICIAS)

O bien con los cortes y reformulaciones que se muestran en (8):

- (8) eh bueno sí o sea en partes anteriores pero que no rebasen de de tres piezas esto siempre y cuando sea en lo que es empres- dos (.) el empres- dos tiene una dureza mucho más que que mayor que lo que viene siendo el empres- uno el empres- e- dos o uno se pueden usar (.) ya sea para hacer alguna: corona o como di- bien dicen por allí alguna fundita que va a ir recubriendo un solo diente (32RU15-10-07\_SALUD)

Un estudio más detallado de esta fórmula, tanto en su aparición oral como en la comprobación de su ausencia en el discurso escrito de la distancia comunicativa, permitirían confirmar la hipótesis de que se trata de una “muletilla compleja” utilizada para ganar tiempo de procesamiento del discurso.

Una combinación sintagmática adicional es la que se presenta con verbos epistémicos como *creer* o *considerar*, emotivos como *querer* o *preocupar*, o bien de comunicación como *decir* o *platicar*. El funcionamiento de estas construcciones puede perfectamente identificarse con lo que Fuentes Rodríguez (2003) llama operadores discursivos, pues su carácter modalizador del discurso es patente.

Un fenómeno adicional que aparece en la distancia comunicativa oral, directamente relacionado con el carácter monológico de esta forma de comunicación y con la

capacidad de procesamiento cognitivo es la aparición de segmentos con continuidad temática pero dentro de una estructura oracional incompleta. Esta característica del discurso oral es ya señalada por Chafe (1987), quien señala la prevalencia de la organización informativa con fines comunicativos sobre la estructura formal que aparece en los sonidos del habla. También se encuentra en relación directa con la capacidad limitada de la memoria de trabajo de un enunciador al momento de procesar el discurso oral.

#### 3.4. Interlocutores y el diseño para el receptor

El carácter dialógico y polifónico del discurso puede ser también estudiado de manera más detallada en un corpus oral diafásico, en la medida en que es posible percibir cómo desde el discurso oral cara a cara hasta el discurso formal monológico se puede analizar un continuo de organización donde cambian las estrategias de manifestación de la polifonía y la dialogicidad, que van desde los turnos de habla y las referencias directas a lo ya dicho hasta formas como la cita textual y la referencia a discursos previos. Fox (2008) recupera la noción de diseño para el receptor, propuesta desde el análisis conversacional, para incluir fenómenos como la coparticipación en el discurso, la selección léxica, la selección de tópico o el orden de secuencias:

With “recipient design” we intend to collect a multitude of respects in which the talk by a party in a conversation is constructed or designed in ways which displays an orientation and sensitivity to the particular other(s) who are the coparticipants. In our work, we have found recipient design to operate with regard to word selection, topic selection, the admissibility and ordering of sequences, the options and obligations for starting and terminating conversations and so on. [...] It is a major basis for that variability of actual conversation that is glossed by the notion “context-sensitive”. (Sacks y otros cit. en Fox 2008: 257)

Esta perspectiva de análisis puede enriquecer las nociones de modalización tal como se han estudiado hasta ahora, en particular en lo que se refiere a la relación entre discurso y hablante, así como la relación entre discurso y oyente, de tal manera que se pueda establecer un puente entre estos tres ámbitos de relación en la conformación del discurso, de tal manera que sea posible explicar la idea de Hopper (2012) respecto a la gramática:

La gramática de código fijo asume que la gramática precede lógicamente al discurso. La gramática emergente invierte esta premisa colocando el hecho de la comunicación



interactiva primero y observando la estructura como un producto secundario resultante de la interacción. (Hopper 2012: 304)

Los ejemplos de 3.2 y 3.3 ya presentados pueden también analizarse desde esta perspectiva. Por ejemplo la aparición de marcas fáticas como *vedá* o la recuperación temática con formas como *a eso me refiero*, que se introducen no únicamente con una finalidad semántico-estructural, sino también con una finalidad interaccional en el proceso de comunicación.

### 3.5. La contextualización desde la variación discursiva

En la actualidad la discusión en torno a cómo establecer relaciones entre el discurso y el contexto aún tiene terreno por recorrer. En trabajos recientes como el de Van Dijk (2012) o Charaudeau (2012) se plantea la necesidad de establecer un criterio más integral acerca de cómo establecer correlaciones válidas más allá de cada caso particular, es decir, de cada discurso producido en un contexto específico.

Desde esta perspectiva, las posibilidades de obtención de datos para la elaboración de un corpus pueden superar el problema central de cuánta información debe ser tomada en cuenta para dejar un registro del contexto cada que se realiza una grabación o se selecciona un texto escrito para fines investigativos. Con un corpus sistemático de carácter diafásico se pueden encontrar los elementos sistemáticos tomados en cuenta en cada situación de comunicación. Como lo explica Van Dijk (2012):

El contexto se basa en el conocimiento del mundo, pero no es lo mismo que éste último. De esta manera, elaborar una teoría del contexto significa, primeramente, seleccionar aquellos elementos de una situación comunicativa que son sistemáticamente relevantes para el discurso. Lo anterior significa que, primero, debemos examinar la manera en que se definen dichas situaciones en la lingüística, la sociolingüística, la psicología cognitiva y social, la sociología y la antropología, para luego definir los criterios de lo que se debe incluir o no en la teoría del contexto. (Van Dijk 2012: 13)

Incluso, muchos de estos elementos sistemáticos están codificados en cada texto particular, lo que posibilita su análisis en términos sistemáticos de recurrencias dentro de un número amplio de discursos dentro de un corpus.

Tal como lo presenta Charaudeau (2012), es posible entonces proceder por inducción a partir de un corpus amplio para desde las marcas lingüísticas que se encuentran en cada texto particular establecer los haces de relación que permiten entonces establecer los

diferentes niveles de la situación de comunicación, desde la situación específica de comunicación hasta el ámbito de práctica social, pasando por la situación global de comunicación. Estas situaciones, si bien en principio son únicas y específicas de cada acto comunicativo, pueden entenderse como comparables entre sí y, por tanto, como sistemáticas, de ahí entonces que las recurrencias discursivas puedan facilitar la interpretación de la situación de comunicación. Se trata en todo caso de una manera de ampliar la noción de contextualización (Auer y Di Luzio, 1992) a diversidad de discursos, y no únicamente como claves de contextualización dentro de un discurso oral particular.

La ventaja que se añade cuando se observan estos procesos desde un corpus oral es que se tiene una idea más específica de dinamismo con el que se ajusta y adapta cada caso particular de comunicación con los marcos sociales que circunscriben los intercambios lingüísticos, de ahí que la discusión en torno a los géneros discursivos adquiera una nueva arista en la perspectiva de la génesis y conformación de los diferentes haces de relación que se conjugan para conformar un discurso concreto.

#### 4. Conclusiones

La posibilidad de utilizar datos de un corpus oral que contenga material grabado y transcrito posibilita diversos tipos de estudio de manera sistemática que podrán ayudar a entender mejor al sistema de la lengua, no sólo basados en la escritura, sino en datos orales reales, pero que también permiten un estudio de la variación del discurso oral en diferentes planos y niveles. Tanto en el nivel de la estructura como en los diferentes planos de relación presentes en la situación de comunicación es posible establecer ejes de análisis desde diferentes discursos orales producidos en condiciones diferentes entre la inmediatez y la distancia comunicativa.

Una vez que se ha presentado el alcance que puede tener un corpus oral diafásico, es posible reconocer algunas posibilidades de análisis que se abren a partir de la disponibilidad de este tipo de material. Fenómenos como la gramaticalización, el análisis de la relación entre discurso e interlocutores como proceso dinámico e interactivo, usos deícticos que muestran la dinámica del texto como correlación entre el universo del discurso y el contexto, o bien algunas marcas de persona que muestran el

carácter dinámico de esta categoría, son algunas de las posibilidades de análisis que se pueden estudiar de manera más sistemática con un corpus oral de este tipo.

La lengua no se presenta en un corte sincrónico de manera estable, sino más bien como diferentes estados de lengua en diferentes estadios de cambio. Dicho de otro modo, un análisis de la lengua oral debe tener en cuenta el reflejo de la diacronía en la sincronía.

## 5. Bibliografía

Antaki, C. (2002): *An Introductory Tutorial in Conversation Analysis*. < <http://www-staff.lboro.ac.uk/~ssca1/sitemenu.htm> >. Última consulta: 04-11-2008.

Auer, P. y A. Di Luzio (1992): *The Contextualization of Language*. Amsterdam/Philadelphia: Benjamins.

Briz, A. (2010): El registro como centro de la variedad situacional. Esbozo de la propuesta del grupo Val.Es.Co. sobre las variedades diafásicas. En: Fonte Zarabozo, I. y L. Rodríguez Alfano (eds.), *Perspectivas dialógicas en estudios del lenguaje*. San Nicolás de los Garza: Universidad Autónoma de México-Iztapalapa, 21-56.

Chafe, W. (1987): Cognitive constraints on information flow. En: Tomlin, R., *Coherence and Grounding in Discourse*. Amsterdam/Philadelphia: Benjamins, 21-52.

Chafe, W. (1994): *Discourse, Consciousness and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press.

Charaudeau, P. (2012): Los géneros: Una perspectiva socio-comunicativa. En: Shiro, M. y otros (eds.), *Los géneros discursivos desde múltiples perspectivas: teorías y análisis*. Madrid: Iberoamericana Vervuert, 19-44.

CORPES XXI (= *Corpus del Español del Siglo XXI*). < <http://web.frl.es/CORPES/view/inicioExterno.view> > Última consulta: 25-05-2015.

Coseriu, E. (1992): *Competencia lingüística*. Madrid: Gredos.

- CREA (= *Corpus de Referencia del Español Actual*) < <http://corpus.rae.es/creanet.html> >  
Última consulta: 25-05-2015.
- Cumming, S. y T. Ono (2000): El discurso y la gramática. En Van Dijk, T. (ed.), *El discurso como estructura y proceso*. Barcelona: Gedisa, 171-205.
- Davies, Mark (2002): *Corpus del español*. < <http://www.corpusdelespanol.org/> > Última consulta: 25-05-2015.
- Diessel, H. (1999): *Demonstratives. Form, Function and Grammaticalization*, Amsterdam/Philadelphia: Benjamins.
- Fillmore, C. (1997): *Lectures on Deixis*. Stanford: CSLI Publications.
- Fox, B. (2008): Dynamics of discourse. En: Antos, G. y E. Ventola (eds.), *Handbook of Interpersonal Communication*. Berlín/Nueva York: Mouton de Gruyter, 255-284.
- Fuentes Rodríguez, C. (2003): Operador/Conector. Un criterio para la sintaxis discursiva. *RILCE* 19.1, 61-85.
- Givón, T. (1979): From discourse to syntax: grammar as a processing strategy. En: Givón, T. (ed.), *Discourse and Syntax*. Nueva York: Academic Press, 81-112.
- Givón, T. (2001): *Syntax*. Vol. II. Amsterdam/Philadelphia: Benjamins.
- Givón, T. y M. Shibatani (2009): *The Genesis of Syntactic Complexity*. Amsterdam/Philadelphia: Benjamins.
- González Ruiz, R. (2009): Algunas notas en torno a un mecanismo de cohesión textual: la anáfora conceptual. En: Penas Ibáñez, M<sup>a</sup> A. y R. González Pérez (eds.), *Estudios sobre el texto. Nuevos enfoques y propuestas*, Francfort: Lang, 247-278.
- Gundel, J. K. (1988): Universals of topic-comment structure. En: Hammond, M. y otros (eds.), *Studies in Syntactic Typology*. Amsterdam: Benjamins, 209-239.
- Hopper, P. J. (2012): Emergent grammar. En: Gee, J. P. y M. Handford (eds.), *The Routledge Handbook of Discourse Analysis*. Londres/Nueva York: Routledge, 301-314.
- Koch, P. y W. Oesterreicher (1985): Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36, 15-43.

- Koch, P. y W. Oesterreicher (2001): Langage parlé et langage écrit. En: Holtus, G. y otros (eds.), *Lexikon der Romanistischen Linguistik*. Vol. I, 2. Tübingen: Niemeyer, 584-628.
- Kotschi, T. (1996): Procedimientos de producción y estructura informacional en el lenguaje hablado. En: Kotschi, T. y otros (eds.), *El español hablado y la cultura oral en España e Hispanoamérica*, Madrid: Iberoamericana Vervuert, 185-207.
- Levinson, S. (1983): *Pragmatics*. Cambridge: CUP.
- López Serena, A. (2007): La importancia de la cadena variacional en la superación de la concepción de la modalidad coloquial como registro heterogéneo. *Revista Española de Lingüística* 37, 371-398.
- Pérez Álvarez, B. E. (2011): Función de los interpretadores en el desarrollo textual. *Estudios de Lingüística Aplicada* 51, 107-125.
- Pérez Álvarez, B. E. (en prensa): Funciones discursivas de *esto*, *eso* y *aquello* en un corpus oral. *Anuario de Letras. Lingüística y filología* II/1, 101-140.
- PRESEEA (= *Proyecto par el estudio socilingüístico del español de España y de América*). < <http://preseea.linguas.net/Corpus.aspx> > Última consulta: 25-05-2015.
- Raible, W. (1992): *Junktion. Eine Dimension der Sprache und ihre Realisierungsformen zwischen Aggregation und Integration*. Heidelberg: Carl Winter.
- Serrano, M. J. (2006): *Gramática del discurso*. Madrid: Akal.
- Van Dijk, T. (2012): *Discurso y contexto. Un enfoque sociocognitivo*. Barcelona: Gedisa.

Recibido: 4 de septiembre de 2014

Aceptado: 30 de octubre de 2014

Publicado: 30 de noviembre de 2015