

Johannes Kabatek

## ¿Es posible una lingüística histórica basada en un corpus representativo?

**Resumen:** La cuestión de los tipos de corpus y los “géneros discursivos” se ha planteado, en los últimos años y en una serie de trabajos, como una de las variables que pueden influir en el cambio lingüístico. Sin embargo, me parece que el estatus teórico de las tradiciones discursivas se ha de plantear de otra manera: no es “una variable” como otras, sino condición profunda de la producción y recepción de los textos, y cuando investigamos los corpus, estamos mirando textos y no lengua, aunque ambos se suelen confundir. Otra cuestión teórico-metodológica que hay que plantear en este contexto es la del papel de la frecuencia: ¿en qué sentido la frecuencia de los elementos en los textos nos indica algo acerca de la cuestión de desarrollo de un fenómeno y, además, acerca de la cuestión de las tradiciones de los textos? Nuestro propósito es, pues, presentar estos problemas teórico-metodológicos en general y en concreto, basándonos en una serie de cuestiones tipológicas y de determinación nominal.

---

**Johannes Kabatek:** Universität Tübingen, Romanisches Seminar, Wilhelmstr. 50, D-72074 Tübingen, E-Mail: kabatek@uni-tuebingen.de

### 1. Introducción

Hace algún tiempo, un colega me mandó un mensaje preguntando si le podía recomendar un pequeño corpus representativo, basado en ediciones fidedignas, para trabajar con sus alumnos sobre temas de la historia del español. Lo que buscaba era una colección de textos de tamaño manejable y que al mismo tiempo incluyera una cierta gama de tradiciones discursivas, permitiendo así ver la evolución de los hechos lingüísticos a lo largo de los siglos.

El pedido no pudo haber sido formulado de manera más clara, y por ello sentí aún más gravemente el apuro del que sabe que debería tener una respuesta inmediata a la pregunta por el fundamento del propio trabajo. Es como si al albañil le preguntaran cómo prepara el cemento y se quedara sin respuesta. Tardé en contestar ... Daré la respuesta a continuación, pero violando tal vez la máxima de calidad de Grice: la respuesta no contesta a la pregunta, y no por incapacidad, sino por imposibilidad. Suena bien lo del corpus pequeño, variado y representativo, basado en ediciones fidedignas. Pero pensándolo bien, hay que admitir que

ese “cemento” de nuestra disciplina, la base incuestionable de nuestro trabajo, no existe. Y si no existe es en parte porque no puede existir, en parte también porque no se ha hecho. Me explico:

- un corpus representativo para la historia de una lengua es una construcción teóricamente imposible ya que la lengua, aunque solo se manifieste en textos, no es la suma de los textos sino algo distinto.
- un corpus representativo es, además, una construcción empíricamente imposible ya que la producción de textos que se ha archivado no corresponde a más que una mínima parte de la producción lingüística total, y esto no solo en lo que se refiere a los textos concretos, los *tokens* de los textos, sino incluso con referencia a los *tipos* de textos producidos.
- con todo, hay que plantearse la cuestión de la representatividad de los textos (representatividad si ya no de la historia de la lengua, quizá por lo menos representatividad menor o mayor para la reconstrucción de una evolución particular).

En lo que sigue trataré, pues, algunos de los problemas relacionados con la reconstrucción histórica por medio de corpus, y me limitaré a algunos aspectos de las ediciones que sientan la base de los corpus y a la selección de textos para un corpus representativo, sin ignorar que los corpus plantean otros problemas importantes más allá de esos dos.

## 2. Un corpus basado en ediciones fidedignas

La base de todo corpus son los textos, pero, como es obvio, entre la mera existencia de un manuscrito, una carta, una grabación oral e incluso un libro y un texto digitalizado para el análisis lingüístico, hay varios pasos a dar. En el caso de la Edad Media, de la cultura del manuscrito, es particularmente espinoso el trabajo de preparación. Es lógico que la lingüística de corpus no pudo empezar desde cero y elegir libremente entre el amplio mundo de los manuscritos. Para los grandes corpus históricos se solían aprovechar textos ya editados, seleccionando o bien – ateniéndose al criterio de calidad – aquellos de los que se suponía que tenían algún valor particular en la historia de la lengua (textos literarios, sobre todo, como, para el español, el *Cid*, *El Conde Lucanor*, *La Celestina* o el *Quijote*, por ejemplo) o bien – ateniéndose al criterio de cantidad – simplemente textos ya disponibles en alguna versión digitalizada. Los problemas que esto traía consigo son numerosos, como el que a veces precisamente los textos literarios famosos no son los más “típicos” de su época, o que en alguna época tenemos mucho material de un determinado tipo (p. ej. los fueros medievales) y poco de otros; o

que a veces se confunde la fecha de composición de un existente o supuesto original con la fecha del manuscrito o de la edición, en ocasiones posterior en siglos<sup>1</sup>, o en otras ocasiones se piensa que originales y copias no se distinguen, por lo menos no en los aspectos sintácticos<sup>2</sup>.

Podría haber un corpus basado en ediciones fidedignas, claramente fechadas y bien preparadas, pero mientras sigamos con los problemas mencionados, en realidad estamos todavía lejos de la “fidelidad” como criterio, y en parte nuestras ideas de la historia del español no son siquiera insuficientes sino simplemente falsas. Y habría que añadir que tampoco es garantía absoluta saber que un manuscrito es realmente de la época que tratamos; hay por lo menos dos factores más que interfieren:

- Por un lado, los criterios de transcripción. No todas las ediciones que manejamos corresponden al riguroso criterio filológico de la fidelidad al manuscrito, encontramos enmiendas a veces injustificadas, correcciones que se guían por intuiciones dudosas que los editores piensan tener de la lengua de una época lejana y que nos pueden llevar a la circularidad del estudio de la lengua de una época creada en parte por los investigadores mismos.
- Existe, por otro lado, un problema más inmediato, el de la calidad del texto de los manuscritos. Hay, evidentemente, manuscritos altamente elaborados que nos permiten transcripciones casi directas. Pero en la cultura del manuscrito hay también muchos fenómenos que aparecen debido a los errores y los descuidos de los copistas. Al mismo tiempo hay fenómenos que a primera vista podrían parecer errores o descuidos ya que en épocas más tardías desaparecen de la lengua escrita, siguiendo en parte vivos en la lengua oral, como por ejemplo ciertos anacolutos aún presentes en la lengua antigua y después eliminados (Palacios Alcaine 1996). La diferencia entre ambos fenómenos es enorme: el error del copista es un mero fenómeno del texto escrito y puede no tener ninguna relación con la lengua de la época mientras que los anacolutos bien pueden representar fenómenos de la época. Pero ¿cómo distinguirlo?

No existe ninguna receta perfecta, pero sí podemos partir de un principio fundamental: los textos – tanto escritos como hablados – siempre tienen una intención

---

1 Así, nos hallamos todavía ante situaciones en las que el *Cid* es considerado, sin comentario alguno, un texto de mediados del siglo XII, o que la edición del *Calila e Dimna* de Cacho Blecua y de Lacarra basada en dos manuscritos del siglo XV se considera texto del siglo XIII, etc. (cf. ahora la edición crítica de Hans-Jörg Döhla (2009): *Calila e Dimna. Nueva edición y estudio de los dos manuscritos castellanos*. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo).

2 Para una visión crítica, véase, entre otros, Morala (2002).

“gramatical”, es decir que lo que encontramos en un manuscrito o bien tiene que ver con la voluntad de escribir algo gramatical o es un error de escritura, del proceso de escribir. Los errores de escritura tienen su sistematicidad, con omisiones, saltos de línea, grafías equivocadas, etc., y una vez identificados, nos quedan textos que suelen ser muy gramaticales también en la época de los manuscritos, aunque la gramaticalidad a la que corresponden sea otra que la de los tiempos de la imprenta. Podemos partir del axioma – y esto hay que subrayarlo – de que más allá de las llamadas máximas de Grice, como el principio de calidad o de la cantidad<sup>3</sup>, existe otro principio axiomático válido para la lengua, y es que ningún discurso, sea hablado o escrito, es agramatical.

Se dice a veces lo contrario, sobre todo de la lengua hablada, pero lo de la agramaticalidad de lo hablado es un mito inventado por los que nunca han analizado profundamente el habla o que se niegan a ver los principios gramaticales propios de la lengua hablada<sup>4</sup>. Con la intuición, con la empatía de los hablantes somos casi siempre capaces de reconstruir el sentido de un enunciado oral actual, y podemos también establecer o reconstruir su gramaticalidad. Y lo que es válido para la lengua hablada, vale también en cierto sentido para lo escrito. En parte, los errores de la escritura son fácilmente reconocibles, sobre todo cuando tenemos distintas versiones de un texto y podemos, por ejemplo, reconstruir claramente incoherencias debidas a errores de los escribanos. Pero a veces parece imposible la reconstrucción de los errores del pasado. Uno de los temas candentes de la filología es, pues, la cuestión del estatus de un fenómeno suelto, de un hápax sintáctico, morfológico o léxico, que tanto puede ser error o descuido como indicador de algo, un fenómeno, una tendencia real de la lengua.

---

3 Me parece un profundo error hablar de *máximas* o de “pautas de orientación” en el caso de algo que en realidad forma parte de la esencia del lenguaje. Los hablantes no se orientan en las máximas como si fueran normas de conducta: no tienen otra opción. Partimos siempre del principio de que lo que nos dice alguien corresponde a algún sentido, y generalmente no decimos “estás loco” si no lo entendemos, sino que solemos pedir aclaración o decir “no te he entendido”. E incluso si decimos “estás loco”, el principio sigue en pie ya que la locura nos sirve precisamente como explicación del sinsentido.

4 Así, se suele citar a Chomsky y su idea de los datos “degenerados” del *input* de los niños en el proceso de aprendizaje: Chomsky (1965, 58) habla de “degenerate quality and narrowly limited extent of the available data”. Sin entrar en la larga discusión que en los últimos decenios ha producido ese problema, hay que insistir – abusando de un término chomskiano – en que el discurso oral no es *degenerado* sino *generado*, y que los oyentes saben cómo reproducir la gramaticalidad que subyace a los enunciados orales, distinguiendo entre elementos de organización textual (marcadores del discurso etc.), elementos universales de lo oral (anacolutos, construcciones *ad sensum* etc.), elementos de una gramática de lo oral y un fondo de reglas oracionales que más o menos corresponden – evidentemente de manera descriptiva y no normativa – a lo que tradicionalmente se considera gramática.

A los problemas señalados se añade hoy en día otro más, evitable pero desafortunadamente no siempre evitado: el problema de la transmisión de los datos a los corpus electrónicos. José Ramón Morala (2006) señaló, entre otros casos, el de *quienes* en los grandes corpus históricos, un elemento innovador tardío en la lengua pero que en los grandes corpus aparece ya desde orígenes principalmente debido a malas lecturas de *quien es* (o de lecturas correspondientes a una *scriptio continua* mal segmentada). Otro ejemplo frecuentemente aducido es el del italianismo *mafia* que aparece en castellano antes que en italiano. Realmente las ocurrencias tempranas se deben a textos escaneados en los que la palabra *maña* fue reconocida como *mafia*. Se suelen contar ejemplos de este tipo con valor anecdótico, pero las anécdotas señalan fuentes reales de “ruido” en nuestros datos, y ponen en duda la cuestión de la fidelidad de los textos que usamos en nuestros análisis.

### 3. La representatividad

Parece que, a pesar de todo lo logrado, a pesar de una enorme productividad, la lingüística histórica – y, más aún la lingüística en general – está en un momento problemático. No quiero hablar de crisis, sino de un discurso “progresista” entre los lingüistas que opone lo avanzado a lo anacrónico y que suele consistir básicamente en la defensa de la penetración de nuevos métodos en la investigación, métodos “objetivos”, datos intersubjetivamente comprobables, experimentos, datos cuantitativos. No sé en cuántos lugares últimamente parece que un estudio solo es serio si trabaja con datos llamados psico- o neurolingüísticos, que a veces de lingüísticos tienen bien poco. Cuántas veces se mide la seriedad de un estudio en si ha empleado o no un *eye-tracker*, un EMG o datos de FMRI. Existen estudios que descubrieron las ondas eléctricas cerebrales correspondientes a la semántica y las distinguen de otras, correspondientes a la pragmática; y hay estudios que creen que saben demostrar si un elemento pertenece al léxico mental o a la sintaxis mediante el seguimiento de los ojos. Pero el seguimiento de los ojos, aunque nos pueda indicar muchas cosas, nunca va a “solucionar” cuestiones teóricas de esta índole, que, además, no son objetivas sino conceptuales. Pero se hace esto porque estamos en un momento de “objetivación” de la ciencia, se mide y cuantifica todo, el impacto de las revistas como el impacto de la investigación, y se mide con métodos adoptados desde las ciencias naturales sin plantearse la cuestión de si esto es realmente posible y si vale del mismo modo para las humanidades (cf. López Serena 2003).

Al mismo tiempo, la lingüística se está volviendo una de las disciplinas de las humanidades de más éxito precisamente por su cercanía a las ciencias naturales,

por el simple hecho de que el lenguaje humano es un fenómeno también natural. Pero lo que tiene éxito es sobre todo la lingüística que imita o adopta los modelos de las ciencias naturales, y hay que preguntarse si esto corresponde realmente al objeto de estudio en cuestión. No vamos a entrar más en detalles, pero podríamos citar numerosos ejemplos de trabajos en los que se demuestra con experimentos que una construcción compleja es más compleja de procesar cognitivamente que una construcción más simple<sup>5</sup>. Lo que sí quiero acentuar es que la misma tendencia se da también en la lingüística histórica, y es una tendencia que en el fondo es antigua y deriva del innecesario complejo de inferioridad de los lingüistas frente a las ciencias llamadas “exactas” (que tampoco lo son).

Lo que en la lingüística que se ocupa del presente es la tendencia hacia el biologismo, los experimentos y los cálculos<sup>6</sup>, en lingüística diacrónica es el afán de buscar la objetividad en datos cuantificables y en los corpus. Y también aquí se da el mismo efecto que en la lingüística experimental: se piensa a menudo que la innovación técnica ya es un logro de por sí, y que la mera técnica ya es innovadora. Pero esto no es así. El progreso solo es progreso si viene acompañado de reflexión sobre lo pasado, y la aplicación de una técnica solo tiene sentido si es adecuada al objeto.

Los lingüistas hemos adoptado el término *representatividad* de la sociología y partimos en general del supuesto de que el lenguaje es un fenómeno social como otros; que los elementos que varían son investigables como otras variables sociológicas. Son muy frecuentes los juicios acerca de los datos que hablan de representatividad del corpus, como cuando en la descripción del *CORDE* en la página de la Academia se dice:

Se pretende recoger todas las variedades geográficas, históricas y genéricas para que el conjunto sea suficientemente representativo<sup>7</sup>.

O cuando en trabajos sobre aspectos particulares se habla de la representatividad de los datos, como en el siguiente ejemplo:

consideramos que, en lo que respecta a la prosa escrita, literaria y no literaria, al menos, los resultados del cuadro 1 son representativos del empleo del orden en la FN en los periodos históricos atendidos. (Martínez 2009: 1241)

---

5 Para más detalles, véase Kabatek (2012).

6 Para el caso de la modelización de contactos lingüísticos, véase Kabatek (2012).

7 <http://www.rae.es/rae/gestores/gespub000019.nsf/voTodosporId/B4E26FC2520104D8C125716400455C06?OpenDocument>.

¿Qué quiere realmente decir *representatividad*? Me parece que en lingüística de corpus esta cuestión normalmente no se plantea desde sus principios, sino desde una serie de presupuestos. Decimos por ejemplo que la representatividad está dada cuando mezclamos un porcentaje más o menos parecido de textos jurídicos con textos literarios y prosa no literaria. Y la mezcla de estas tres partes nos dará la lengua de la época. Si antes señalamos el problema de los posibles errores de copistas, aquí estamos no solo delante de un sinfín de posibles errores, estamos, en mi opinión, ante una cuestión mal planteada.

La representatividad sociológica deriva de la identificación de parámetros correlacionables que condicionan el comportamiento humano. Tales parámetros permiten reducir la investigación de un fenómeno supuestamente relevante para una población exhaustiva a una porción reducida de la población, la cual refleja porcentualmente los parámetros relevantes. El ejemplo más conocido es el de la investigación electoral: poniendo un caso simplificado, supongamos que en un pueblo de mil personas resulta que las mujeres suelen votar por un partido y los hombres por otro. En tal caso, en principio la investigación de una sola pareja sería representativa para la población entera: la muestra de dos sería entonces extrapolable a toda la población. Pero esto no lo sabe el investigador de antemano, lo sabrá después de haber preguntado a unos cuantos y no haber encontrado variación. Con cada pareja preguntada, la seguridad aumenta. Y si acertamos con la representatividad, la repetición del estudio o con la misma o con otra porción de la población no cambiará el resultado. La extrapolación de un resultado, evidentemente, siempre contiene un margen de posible error, margen que se relativiza comparando datos de varias muestras y conociendo los factores que suelen tener alguna importancia en el caso concreto investigado. En lingüística, podemos decir que es sobre todo en sociolingüística y sociología del lenguaje donde los métodos estocásticos son aplicados sin mayores inconvenientes. Si por ejemplo observamos que hay una correlación entre hábitat, nivel social, edad y procedencia y el uso de una u otra lengua en comunidades bilingües como Valencia, Mallorca, Galicia o el País Vasco, podemos tipificar los factores y, mediante una muestra “representativa”, extrapolar nuestra información a la población entera.

Pero hay que preguntarse en qué medida es comparable la cuestión de la representatividad en el caso de un corpus histórico.

Los votantes de un país forman un conjunto cerrado de número más o menos conocido. En la mayoría de las elecciones, el voto relevante es uno, y los partidos o los candidatos que se presentan, aunque suelen ser varios, se reducen generalmente a dos o quizá tres principales, tal vez algunos más, según el caso. Todo es bastante simple, las variables se pueden manejar de manera relativamente clara. Aun así, existe todo un mundo de estudios sobre cómo calcular bien la probabili-

dad de los resultados de unas elecciones: incluso con parámetros que parecen tan simples, la cosa es todo menos fácil.

Si ahora queremos hablar de “representatividad” en materia de historia de la lengua, a primera vista esto solo parece posible desde una visión muy simplificada de la lengua y su historia. Esta visión – no solo simplificada, sino realmente falsa – podría describirse más o menos así:

- “una lengua es un sistema para la producción de enunciados. La lengua es un hecho social, y esto lleva a que la lengua de los hablantes de una comunidad sea más o menos idéntica. La lengua varía a lo largo del tiempo (dejemos de lado aquí el porqué). Como la lengua es más o menos unitaria y la única variación es diacrónica, hay que cotejar textos perfectamente fechados, claramente ubicados en el tiempo, y se podrá reconstruir la diacronía.”

Esto es claramente exagerado, pero me parece que la exageración está mucho más cerca de la práctica de lo que parece. Y es curioso que mantengamos la tendencia a este tipo de simplificaciones a pesar de las críticas epistemológicas en la propia disciplina. La idea de la representatividad de un texto para una época es parecida – e históricamente está relacionada – a la idea gilliéroniana de una geografía lingüística que, entrevistando en cada lugar a un hablante representativo, nos dará la verdadera imagen de la variación dialectal de una lengua. Pero ya por lo menos desde Gauchat (1905) sabemos que esto no funciona así, y que incluso dentro de comunidades pequeñas la lengua no suele ser homogénea. Y lo mismo vale para la historia de la lengua: hay varios factores de variación, y tenemos que tomarlos en consideración.

Para plantear coherentemente la cuestión de la representatividad de un corpus tenemos que preguntarnos por dos cosas esenciales: por un lado, por el objeto mismo que llamamos historia de la lengua y los datos correspondientes, y por otro lado, por las posibilidades que nos ofrecen los corpus.

Hay una cierta unanimidad sobre el hecho de que la única posibilidad de reconstruir la historia de una lengua como la española es a través de datos de corpus (aunque esto no es toda la verdad: el latín vulgar también se reconstruye sin corpus, se reconstruyen las lenguas perdidas y las historias de lenguas que parecen sin historia). Los corpus nos ofrecen un instrumento más poderoso que la reconstrucción indirecta, pero necesitamos introducir algunas distinciones para saber lo que realmente hacemos cuando trabajamos con ellos.

La primera distinción es que los corpus contienen textos, pero la lengua no es lo mismo que el texto. La lengua histórica no es una entidad homogénea, la idea de que se trata de un sistema capaz de producir un sinfín de enunciados mediante una serie de reglas limitadas es una idea simplificadora y hasta cierto punto ingenua. Si fuera así, una lingüística histórica inmediatamente basada en un



corpus con textos bien fechados sería posible. Parece que, en algunas escuelas, la poca preocupación por los datos de los corpus deriva precisamente de esta idea. Pero la lengua histórica no es un sistema sino un conjunto de sistemas – tanto objetivamente como en la propia competencia de los hablantes. Las variedades diatópicas, diastráticas y diafásicas son dimensiones del hablar junto a la dimensión diacrónica. Los hablantes disponemos de sistemas, y disponemos, además, de conocimientos parciales de reglas de otros sistemas, reglas que para nosotros vienen normalmente asociadas con lugares, grupos, situaciones, y también con textos. Hasta podemos decir que podríamos reducir todas esas dimensiones a los textos, los textos enunciados en lugares, con grupos, en situaciones, por escrito y hablado.

La segunda distinción es la que tenemos que establecer entre sistemas lingüísticos, bases para la generación de textos, y textos ya producidos, con sus aspectos formales y tradicionales. Y no es que exista primero la lengua y después, a manera de añadidura, una variable “género” que en algunos casos, cuando el mero sistema no nos basta para explicar un fenómeno, sirve como término elástico para todo lo que no podemos explicar de otra manera. No podemos decir: esta es la evolución de la lengua, y si hay desvíos o fenómenos con los que no contábamos, será por la variable género. El problema de esa variable es algo más profundo. Es lo que Mikhail Bakhtin ya intuía en su crítica de la lingüística sistémica y su insistencia en la tradicionalidad de los discursos: todo texto, del saludo a la novela, está evocando, repitiéndolos o no, textos anteriores. Y esos textos se caracterizan también por marcas gramaticales, no solo por un determinado léxico o determinadas fórmulas concretas. Así, el poeta que a finales del siglo XX habla de “una perdida estrella”, de “imprevista languidez”, del “acabado poema” o de la “fría y serena corola” evoca una tradición, igual que lo hace el dictador que en su discurso de 1931 habla del “esplendoroso sol”, del “espinoso camino” o de la “noble hidalguía”, o del autor de una guía de viajes que evoca la “madrileña plaza mayor” o el filósofo que habla de la “moderna filosofía”. Ejemplos diversos de un mismo fenómeno que evoca tradiciones muy diversas, pero obviamente las evoca a través de su carácter gramaticalmente marcado.

Los hablantes, dicen Peter Koch y Wulf Oesterreicher, hablan una lengua y al mismo tiempo evocan tradiciones; *tradiciones del hablar*, como decía Brigitte Schlieben-Lange (1993), *tradiciones discursivas*, en la terminología de Peter Koch (1997). La idea de la tradicionalidad de los textos lleva a Koch y Oesterreicher a postular la duplicación del nivel histórico de la concepción coseriana de los tres niveles del hablar, el universal, el histórico y el individual. Sin embargo, me parece que esta duplicación no es realmente necesaria. Sin negar la importancia esencial de las tradiciones discursivas, me parece preferible considerarlas de manera distinta, como tradiciones del producto. Nos servimos de las tradiciones

pero asumimos el sistema: los sistemas lingüísticos son evidentemente sistemas históricamente dados, pero mediante el aprendizaje de la primera lengua el sistema se autonomiza dentro del hablante, que se libera de la historia ya que la lleva dentro. Las tradiciones discursivas, sin embargo, no son asumidas de la misma manera, nos servimos de ellas, y evocamos a través de ellas la tradición. Es decir: la *perdida estrella* evoca los textos del siglo XIX a los que explícitamente hace referencia; *árbol verde*, sin embargo, evoca en primer lugar un árbol verde. Diría, pues, que no hay dos niveles históricos, sigue habiendo uno solo, el asumido y apropiado, pero a su lado hay otra historicidad, no de *enérgeia* histórica, sino de *ergon* tradicional, un nivel de historia de los textos y de su tradicionalidad.

La tradicionalidad deriva de los productos repetidos, y puede ser una tradicionalidad de forma o de contenido, de textos completos o de elementos textuales.

La tradicionalidad de los textos es una característica del lenguaje que lo aproxima de otros hechos culturales. En la tradición de los textos, el lenguaje es un producto cultural como cualquier otro, como en la tradición de vestir, la tradición de construir casas, etc. Para realmente distinguir la “tradición discursiva”, la tradición de los textos, de la lengua histórica, hemos adoptado y elaborado una distinción de Eugenio Coseriu que identifica tres tipos distintos de historicidad. En una discusión entre Harald Weydt, Brigitte Schlieben Lange y Eugenio Coseriu sobre la cuestión de la historicidad de los actos de habla, este último introduce una diferenciación que me parece altamente pertinente: Coseriu niega, en primer lugar, la historicidad de los actos de habla. Los actos, en cuanto actos, son universales, derivan de la semiótica fundamental de la comunicación y hacen posible que la actuación del hablante y del oyente varíe según la relación entre ellos y lo referido. La universalidad de los actos no impide que éstos adquieran distintas formas históricas; así, un acto directivo que ordena algo puede tener una forma gramatical particular en una lengua, puede expresarse mediante una fórmula tradicional y ser un texto que se repite y puede incluso ser no-verbal y consistir en un simple gesto. El grado de gramaticalización de los actos varía de lengua en lengua, y puede que la lengua incorpore, por ejemplo, plenamente los actos polifónicos de anticipación de la reacción del oyente mediante partículas modales (como el neerlandés o el alemán) así como es posible que exprese la misma función mediante fórmulas tradicionales o que ni siquiera tenga medios explícitos para expresarlos.

Los actos pertenecen, pues, al nivel del hablar en general, el nivel de la pragmática – lo cual subraya evidentemente la universalidad de la pragmática y el problema que conllevan nociones como “pragmática del español” (habría que hablar más bien de pragmática ejemplificada mediante el español). Los actos

pueden llegar a gramaticalizarse en una lengua determinada, y cuando esto ocurre, entonces sí pertenecen a la “pragmática de la lengua x” ya que una lengua suele disponer de elementos más o menos gramaticalizados con función básicamente pragmática. Y los actos pueden también expresarse tradicionalmente en textos más o menos fijados o mediante formas textuales más o menos establecidas.

Lo importante es aquí la diferenciación de niveles: en el nivel universal encontramos los actos como abstracciones, en el nivel histórico, las lenguas como sistemas funcionales. Los sistemas funcionales son condición imprescindible para el acceso a la universalidad; los seres humanos llegan al mundo mediante el lenguaje, y el lenguaje les es dado en forma de una lengua histórica. Con esta base de conocimiento, desde esta historicidad primaria, los hablantes producen textos, y estos textos forman una segunda historicidad o tradición, y la tradición abarca por un lado más, y por otro menos que los signos lingüísticos: la tradición no es una tradición de signos abstractos sino de signos y referentes o marcos referenciales; es una tradición de eventos históricos concretos, eventos únicos e irrepetibles que a su vez corresponden a un tercer tipo de historicidad, el de la ubicación en el tiempo. De este modo, tenemos cuatro niveles copresentes en los actos lingüísticos: lo que tú estás leyendo en este momento corresponde a ciertos actos referenciales, corresponde asimismo a la gramática y al léxico de una lengua, a la tradición de un ensayo lingüístico y es acto único producido solo una vez en la historia en un lugar determinado. Lo que complica el análisis de esta copresencia es precisamente que el material es el mismo: los signos lingüísticos realizan actos, son gramaticales, son tradicionales y cada actualización es única. La copresencia hace a veces difícil separar los niveles y distinguir lo tradicional de lo gramatical. Las tradiciones discursivas no son más que otras tradiciones culturales, pero a diferencia de la tradición de la arquitectura, por ejemplo, que es material y formal pero que se describe mediante el lenguaje, en el caso de las tradiciones discursivas hay simultaneidad de gramática y tradición.

Lo que producimos cuando hablamos no es, pues, ni realización de un único sistema ni solo gramática y léxico de un estado de lengua; es, además, tradición; y la tradición puede también quebrar – hasta cierto punto – las reglas del sistema, o, mejor dicho, permitir que aparezcan elementos correspondientes a reglas de sistemas distintos.

La tradición puede ser fórmula o forma, puede ser texto o construcción, puede añadir elementos a lo esencial y así insertar un texto en una tradición como cuando digo “érase una vez”, o puede restar elementos ya que evoca textos enteros como cuando digo “buenas”. La tradición es signo, y puede ser nombrada: “buenas” es signo de saludo y “érase una vez” de cuento. Y la tradición puede ser género discursivo de tipo “tradición jurídica”, “literaria”, “oral”, etc. Pero preferimos hablar de *tradición discursiva* y no de género discursivo, primero porque la

noción de género es muy ambigua y suele evocar una fuerte connotación literaria, segundo porque la noción de género discursivo suele evocar a Bakhtin o, mejor dicho, a la recepción de Bakhtin, la cual es parcial, por un lado, y exagerada, por otro, ya que ve una “teoría de los géneros discursivos” donde en realidad no hay más que algunas nociones generales, aunque muy sugerentes. Además, si hoy en día en muchos trabajos de historia de la lengua aparece la noción de “género discursivo” para dar cuenta de la variación textual, podemos, por un lado, destacar como evolución positiva el abandono de la concepción monolítica de la historia de la lengua y el respeto por la variación textual, pero tenemos que criticar al mismo tiempo la limitación con la que suele aparecer esta noción. Se habla de la “lengua oral” como género, del “género periodístico”, de la “prosa jurídica”, y se piensa que así se ha pagado el tributo merecido a la variación textual. Pero lo que se llama “oral” por ejemplo en un macrocorpus como *CREA* son textos muy variados, y encima en gran parte textos transcritos con omisión de los elementos fundamentales de la oralidad (cf. López Serena 2006); lo que se llama “género periodístico” es solo un denominador común de numerosos fenómenos textuales (basta con la consulta de cualquier periódico para darse cuenta), e incluso lo que al lego le parece ser una denominación bastante exacta, la de “prosa jurídica”, denomina un conjunto altamente complejo de textos. Antes de hablar de una “variable género”, deberíamos, pues, precisar cuáles son los factores de variación según la tradición textual. Y para hacerlo, deberíamos librarnos primero de categorizaciones dadas de antemano y preguntarnos sin prejuicio por todos los posibles elementos “tradicionales” en un texto y así entender la noción de Tradición discursiva en toda su amplitud (Kabatek 2011). Esto no quiere decir que las categorizaciones usuales en la lingüística de corpus, en la lingüística del texto o en la tipología textual no puedan ser relevantes a la hora de analizar la historia de una lengua, sobre todo aquellas categorizaciones que son propias de los hablantes y que les sirven como pautas de orientación formal y de contenido cuando escriben textos o cuando hablan. Pero tal como la tipología textual está lejos de ser una disciplina en la que reine la unanimidad, la realidad del objeto texto nos enseña que permite múltiples maneras de clasificación, global o interna, y que todas ellas corresponden a tradiciones discursivas. Es decir que, si clasificamos ciertos textos bajo el denominador “prosa jurídica”, puede haber debajo de esta denominación textos de muy diversa índole, textos relacionados con otras tradiciones, con diferenciación interna en parte enorme, textos homogéneos o heterogéneos, largos, breves, elaborados, poco elaborados; textos formulaicos o libres, arcaizantes o innovadores. Todas estas subtradiciones son también tradiciones discursivas, y lo son o lo pueden ser, además, las diferentes secciones de un texto, que pueden tener una “vida tradicional propia”, como el título, el prólogo, una fórmula determinada o cualquier subsección del texto, que también es siempre un

“Teilganzen” (un ‘todo-parcial’), para emplear este término de la fenomenología introducido por Wolfgang Raible (1980: 327) en la lingüística del texto.

## 4. La reconstrucción de la lengua histórica (I): Filología

¿Cómo sabemos todo esto, cómo podemos identificar el entramado tradicional de un texto? Pues seguramente, para dar cuenta de manera exhaustiva de las distintas tradicionalidades representadas en un texto, la única vía es el análisis filológico minucioso y detallado. La identificación de las posibles alusiones históricas de cada palabra, el acercamiento a la más completa intertextualidad (siempre aproximativo), todo esto no es cuestión de frecuencia o de datos cuantitativos: es cuestión de un trabajo de reconstrucción histórica pormenorizada, en la línea de lo que Wulf Oesterreicher (2011) llamó la *recontextualización* de los textos de la historia de una lengua: la reconstrucción de los entornos de los textos, de sus condiciones de producción, lo cual incluye – afán utópico, desde luego – la reconstrucción no solo de la tradición de alguna que otra expresión o frase, sino del acto de producción mismo; incluye no solo la situación concreta, sino el repertorio lingüístico y tradicional del que disponía en el momento dado el productor del texto, un repertorio amplio, de variedades y de textos, del que eligió precisamente lo que después nos es dado como resultado textual. Todo este repertorio está – aunque sea negativamente – presente en la producción del texto, y no deberíamos ignorarlo y pensar que el mero resultado de este proceso describe el proceso entero.

Abrimos aquí un pequeño paréntesis: uno de los principios de la lingüística cognitiva que se ocupa de las cuestiones de frecuencia trata de la relación entre frecuencia y anclaje cognitivo o “entrenchment”. Se supone que los elementos lingüísticos, cuanto más frecuentemente producidos, más anclados en la competencia estarán. Esta idea tiene algo de verdad y nos ayuda a explicar la larga supervivencia de las formas irregulares altamente frecuentes y ancladas. Pero la idea tiene también un defecto ya observado en la lingüística histórica cuando se equipara el producto con la lengua y no se distingue entre *ergon* y *energeia*. Hay por ejemplo palabras tabú fuertemente presentes en la mente de los hablantes que muy pocas veces se pronuncian y que se evitan en los textos escritos<sup>8</sup>. La estrategia

---

<sup>8</sup> Un ejemplo que me gusta citar en este contexto es el de la redacción del texto de las llamadas *Normes de Castelló*, de 1932, en las que se propone una ortografía unificada para las variedades

de evitar ciertas palabras o construcciones puede hasta estar acompañada de un aumento del anclaje cognitivo de las mismas – por lo menos en una primera fase.

## 5. La reconstrucción (II): corpus multidimensionales

Hemos llegado a un punto de destructor o casi destructivo de esta exposición en el que podríamos concluir que la lingüística histórica es simplemente imposible, que la lingüística histórica ha muerto, ya que su objeto o bien no existe, o no es observable o reconstruible. Pero evidentemente, lo que vamos a intentar presentar en esta última sección será algo más positivo. Es una propuesta que complementa la vía tradicional de la lingüística histórica, la de la reconstrucción “filológica” y detallada, la cual será siempre la base de cualquier aproximación “innovadora”.

Una aproximación diferenciada a una historia de la lengua parece una tarea circular si pensamos que queremos describir la organización de la variación en un corpus en el que lo único que tenemos son los elementos que varían. Sin embargo, tal circularidad se resuelve si tomamos en consideración la cantidad de variables que hay en el juego y la posibilidad – o incluso las múltiples posibilidades – de identificar variables dependientes de otras variables. Las variables son en parte conocidas: si en un texto de los últimos años encuentro la construcción “una perdida estrella”, puedo estar seguro, por un lado, de que no se trata de un texto cotidiano, por ejemplo, y de que además esta forma aparecerá en copresencia con otras formas características de una tradición discursiva poética particular. Esta idea fundamental de la *organización* lingüística de los textos y su diferenciación relativamente sistemática mediante co-ocurrencias de ciertos elementos es la que Douglas Biber aprovechó para sus análisis de diferenciación de lo que él llama *registros* en corpus<sup>9</sup>: la idea de fondo es muy simple, aunque después los aspectos técnicos sean a veces complejos. Se mide automáticamente una serie de elementos en un corpus, elementos que en primer lugar deben ser fácilmente identificables y que, además, presenten una cierta variación. Después se calcula el peso relativo de los elementos en cuestión, su apariencia por cada mil palabras; y en un tercer

---

del catalán, incluyendo el valenciano. Seguramente la palabra más importante de las normas es *catalán* o *lengua catalana*. Pero esa palabra, de obvia presencia en la mente de los productores y receptores del texto, no aparece en él ni una sola vez.

<sup>9</sup> Cf. entre otros Biber (1993); Biber/Conrad/Reppen (1998); Biber/Davies/Jones/Tracy-Ventura (2006).

paso se calculan las co-ocurrencias y se factorizan. El resultado es sorprendente en dos maneras: por un lado, es sorprendente la posibilidad de identificar, por ejemplo, las características de la prosa académica frente a las cartas personales y llegar así a una diferenciación “objetiva” de las características de estas tradiciones discursivas. Por otro lado, lo que llama la atención es que, aplicado a un corpus exhaustivo, este método ayuda a identificar una serie de factores que parecen condicionar la *producción* de los textos, y si proyectamos todas las variables sobre una sola dimensión de variación, la más destacada parece corresponder en primer lugar a la diferenciación entre textos formales e informales, y cuando hay material oral en los corpus, entre textos orales informales, en un polo del continuo, y textos escritos formales, en el otro polo. Ya hemos señalado más de una vez que esta observación empírica se parece mucho al continuo establecido por Peter Koch y Wulf Oesterreicher entre lenguaje de inmediatez y lenguaje de distancia, al que Koch y Oesterreicher (2007) llegaron por una serie de reflexiones teóricas.

Me parece que la diferenciación propuesta por Biber nos acerca a una visión más diferenciada y más realista de la historia de la lengua. En vez de preguntarnos por el estado de la lengua *x* en una época determinada, lo que tenemos que hacer es reconstruir la gama de posibilidades entre inmediatez y distancia que ofrece esta lengua en la época en cuestión.

Aun así, la metodología de Biber, por muy potente que sea, no nos ofrece más que una aproximación a la complejidad de los datos contenidos en un corpus y que trata los textos como unidades. Hemos señalado que puede haber, sin embargo, una heterogeneidad horizontal del texto, es decir que distintas secciones de un mismo texto pueden presentar rasgos lingüísticos bien distintos, pertenecientes a tradiciones diferentes, y por lo tanto hemos propuesto una serie de herramientas para la diferenciación interna de los textos.

En nuestros propios trabajos (Kabatek 2004, 2005a, 2005c, 2005d), hemos adoptado en parte las ideas de Biber, pero en vez de trabajar con una lista exhaustiva de más de 60 elementos, difícilmente medibles sobre todo en corpus históricos con datos gráficamente muy heterogéneos y que presupondrían un enorme trabajo previo de etiquetaje, nuestra propuesta es trabajar con una lista muy reducida de aquellos elementos que consideramos más “sintomáticos” para la variación de las tradiciones discursivas. Esta idea derivó de nuestros trabajos sobre las tradiciones jurídicas en el siglo XIII y aplicó el esquema elaborado por Wolfgang Raible para clasificar los elementos de conexión interoracional, un esquema de juntores y de elementos de junción, a los corpus históricos (cf. Kabatek, Obrist & Vincis 2010). Sabemos desde hace mucho tiempo que la conexión interoracional es un indicador altamente relevante para el grado de elaboración de un texto. Proponemos una primera fase de identificación de los juntores en un corpus, una segunda fase de anotación semiautomática y una tercera fase

de visualización de los resultados en tres distintas maneras: por un lado, identificamos la cantidad relativa y la complejidad sintáctica y semántica de los nexos interoracionales y llegamos a una clasificación global de los textos. En una segunda aproximación mediante un algoritmo que procura dar cuenta del factor de complejidad de los elementos, calculamos lo que llamamos el “índice de complejidad” de un texto. Este índice permite una primera clasificación global de los textos a lo largo de un continuo de dos polos claramente relacionados con los criterios de inmediatez y distancia. En tercer lugar, visualizamos lo que llamamos el “juntograma”, es decir una representación horizontal de los juntores presentes en un texto, lo cual permite identificar diferentes zonas dentro del texto, a veces correspondientes a tradiciones discursivas distintas, como pasajes argumentales versus pasajes narrativos. La herramienta del “juntograma” ha sido recientemente ampliada para medir otros elementos, según la relevancia relativa en distintos textos<sup>10</sup>.

Los análisis de los textos nos llevan a los siguientes resultados, que no son nuevos y en parte retoman las ideas antiguas formuladas ya hace décadas por Badía i Margarit (1960) y otros, pero que hoy en día podemos incluir en una lingüística histórica basada en corpus:

- en cada época, no existe (solo) un “estado de lengua” y por lo tanto tampoco tiene sentido intentar reconstruirlo. Existe una gama de posibilidades de la lengua, y la variación de estas posibilidades corresponde en gran medida a la variación de los textos entre inmediatez y distancia. Estos textos corresponden a *tradiciones*, y en este sentido es necesario considerar la dimensión histórica del continuo entre inmediatez y distancia. Las tradiciones no están fijas, puede haber influencia mutua entre ellas, interferencia textual que produce movilidad de los elementos lingüísticos a lo ancho del continuo. En vez de reconstruir la cantidad relativa de ocurrencias de un cierto elemento en la historia de la lengua (que es nada más que una proyección abstracta), habría, pues, que reconstruir su correlación con un punto o una zona del continuo. Habrá elementos que abarquen el continuo entero junto a otros que pueden cambiar su posición a lo largo de la historia. Pero hay que tener

---

**10** Célia Lopes adaptó el juntograma para una representación horizontal de formas de tratamiento. El correspondiente análisis permite observar claramente cómo las formas innovadoras conviven con las formas antiguas en el interior del mismo texto y cómo a través del tiempo las formas innovadoras van extendiendo sus posiciones en el texto hasta alcanzar la casi totalidad de las ocurrencias.

Los análisis están hechos mediante la ayuda de una herramienta informática elaborada en el seno del proyecto B 14 del SFB 441 *Estructuras de datos lingüísticos* en la Universidad de Tübinga. Véase <http://www.sfb441.uni-tuebingen.de/b14/tools.html>.



cuidado con no confundir el continuo universalmente posible con el continuo de realizaciones concretas en una lengua histórica determinada. Por lo menos cuando nos referimos a la tradición de los textos escritos, la evolución histórica de la lengua muchas veces corresponde más a una evolución de la extensión del continuo que a una diacronía de existencia o cambio de elementos. Así, el continuo realizado y presente en un corpus puede estar limitado a una sección reducida del continuo posible. Lo que se ha llamado *Ausbau* o elaboración del idioma se ha confundido en parte con la evolución diacrónica ya que corresponde a una faceta de la diacronía y es un proceso que se desarrolla en el tiempo. Lo que hace el castellano en el siglo XIII, por ejemplo, más que evolucionar en las zonas donde ya está presente, es sobre todo expandir sus posibilidades hacia el polo de la distancia comunicativa. La “elaboración” evidentemente tiene consecuencias para la lengua entera, pero es un proceso que solo indirectamente afecta a otras tradiciones. Por otra parte, resulta claro que el continuo queda en gran parte perdido en la historia ya que no está documentado. Y así surgen efectos supuestamente diacrónicos que no son más que resultados de cambios de la frontera entre lo conocido y lo desconocido. Un famoso ejemplo es el latín vulgar, en gran medida no sucesor del latín clásico sino apariencia de algo ya dado. Y lo mismo ocurre con numerosísimos fenómenos que en los últimos años se están descubriendo en la evolución del español en los siglos XVIII y XIX, por ejemplo (cf. Guzmán & Sáez en prensa). Se trata de siglos de supuesta estabilidad de un “primer español moderno” (Octavio de Toledo y Huerta 2007). Pero si miramos más hacia los detalles, esos siglos presencian numerosos cambios, y si por ejemplo hay extensión del uso de ciertos marcadores de discurso en el siglo XVIII es quizá también porque el límite de lo que antes se consideraba oral y no apto para la escritura había cambiado o porque nuevos tipos textuales entraron en juego (cf. Pons Rodríguez 2010a). De manera semejante, podemos ver en la actualidad cómo numerosos fenómenos lingüísticos aparecen masivamente en los textos de *facebook* o en otros medios recientes, fenómenos que en realidad no son nuevos, pero nuevos en forma escrita.

Evidentemente, a ambos lados del continuo (y quizá a lo largo del continuo entero) hay también posibilidades de innovación, y las posibilidades pueden extenderse hacia el otro polo del continuo, como suele ser el caso de los fenómenos de gramaticalización que se extienden del polo de la inmediatez hacia el de la distancia, o ciertos fenómenos de elaboración que caminan en la dirección opuesta. O hasta fenómenos que parecen ser unitarios y que, viéndolos desde más cerca, presentan una variación interna llamativa.

## Conclusiones

Ha sido nuestro intento presentar algunas reflexiones generales acerca de una lingüística histórica basada en corpus diferenciados y respetando la importancia de la variable género o, mejor dicho, de las tradiciones discursivas entre inmediatez y distancia comunicativas. Las observaciones más destacadas han sido las siguientes:

- no existe representatividad cuantitativa de un corpus,
- lo que (teóricamente) podría existir es una cierta representatividad cualitativa: un corpus con (casi) toda la gama de posibilidades de una lengua en un momento dado,
- las posibilidades de una lengua varían de acuerdo con las tradiciones discursivas,
- las tradiciones discursivas abarcan mucho más que la “variable género”, pero también incluyen esta variable,
- las tradiciones discursivas se pueden ordenar a lo largo del eje de inmediatez y distancia,
- las innovaciones siempre son innovaciones en una tradición discursiva y pueden desplazarse a otras,
- las “innovaciones” en los corpus históricos (escritos) son a veces cambios del límite entre lo que se escribe y lo que no se escribe,
- los corpus solo pueden mostrar que un fenómeno existe, no permiten la confirmación de la inexistencia de un elemento.

Con los aspectos críticos mencionados, no quiero ofrecer un panorama destructor o desesperanzador, todo lo contrario: me parece que estamos viviendo una época fascinante en la que la lingüística histórica se presenta como disciplina en ebullición. Los grandes corpus históricos han abierto nuevos horizontes, pero también han dejado ver nuevas lagunas y han creado nuevas necesidades. El siglo XXI será el siglo de los corpus diferenciados, de la variación, de la diversidad multidimensional, pero también de la reconstrucción de grandes líneas evolutivas que están detrás de los detalles. La diferenciación de variedades, tradiciones discursivas y líneas de influencia no solo diacrónicas, sino también entre textos permitirá reconstruir una diacronía más adecuada, con lo cual tendremos también una base para estudios más generales, de correlaciones tipológicas o de grandes regularidades o irregularidades más allá de los meros detalles.

## Bibliografía

- Badía Margarit, Antonio [i.e. Antoni Badia i Margarit] (1960): «Dos tipos de lengua cara a cara», en: *Studia Philologica. Homenaje a Dámaso Alonso*, I, Madrid, pp. 115–139.
- Biber, Douglas (1993): «Using register-diversified corpora for general language studies», en: *Computational Linguistics* 19.2, pp. 219–241.
- Biber, Douglas; Conrad, Susan; Reppen, Randi (1998): *Corpus linguistics: Investigating language structure and use*, Cambridge, UK: Cambridge University Press.
- Biber, Douglas; Davies, Mark; Jones, James K.; Tracy-Ventura, Nicole (2006): «Spoken and written register variation in Spanish: A multi-dimensional analysis», *Corpora* 1, pp. 7–38.
- Ciapuscio, Guiomar; Jungbluth, Konstanze; Kaiser, Dorothee; Lopes, Célia (eds.) (2006): *Sincronía y diacronía de tradiciones discursivas en Latinoamérica*, Frankfurt am Main/Madrid: Vervuert/Iberoamericana.
- Company Company, Concepción (ed.) (2009a): *Sintaxis histórica de la lengua española*, Vol. II: *La frase nominal*, 2 vols., México: UNAM/FCE.
- (2009b): «Introducción», en: Company Company (ed.) (2009), I, pp. XI–XXII.
- Company Company, Concepción; Moreno de Alba, José G. (eds.) (2008): *Actas del VII Congreso Internacional de Historia de La lengua española*, Madrid: Arco.
- Davies, Mark (2009): «Creating Useful Historical Corpora: A comparison of CORDE, the *Corpus del español*, and the *Corpus do Português*», en: Enrique-Arias (ed.) (2009), pp. 137–166.
- Enrique-Arias, Andrés (ed.) (2009): *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*, Madrid/Frankfurt am Main: Iberoamericana/Vervuert.
- Fernández Alcaide, Marta; López Serena, Araceli (eds.) (2007): *Cuatrocientos años de la lengua del Quijote*, Sevilla: USE.
- Gauchat, Louis (1905): «L'unité phonétique dans le patois d'une commune», en: *Aus romanischen Sprachen und Literaturen. Festschrift Heinrich Morf*, Halle an d. Saale, pp. 175–232.
- Guzmán Riverón, Martha (2007): «Tradiciones discursivas e historia de la lengua española en América», en: Fernández Alcaide/López Serena (eds.) (2007), pp. 79–88.
- Guzmán Riverón, Martha; Sáez Rivera, Daniel (ed.) (en prensa): *Márgenes y centros en el español del siglo XVIII*, Frankfurt am Main: Peter Lang.
- Jacob, Daniel; Kabatek, Johannes (2001): *Lengua medieval y tradiciones discursivas en la Península Ibérica: descripción gramatical – pragmática histórica – metodología*, Frankfurt am Main/Madrid: Vervuert/Iberoamericana.
- Kabatek, Johannes (2003a): «La lingüística románica histórica: tradición e innovación en una disciplina viva», *La Corónica* 31.2, pp. 35–40.
- (2003b): «Unidad del significado, Designado y *Lingüística Integral*», *Odisea* (Almería) 3, pp. 87–99.
- (2004): «Tradiciones discursivas jurídicas y elaboración lingüística en la España medieval», *Cahiers de Linguistique Hispanique Médiévale* 27, pp. 249–261.
- (2005a): *Die Bolognesische Renaissance und der Ausbau romanischer Sprachen. Juristische Diskurstraditionen und Sprachentwicklung in Südfrankreich und Spanien im 12. und 13. Jahrhundert*, Tübingen: Niemeyer.
- (2005b): «Sobre a historicidade de textos», *Linha d'água* (São Paulo) 17, pp. 159–170.
- (2005c): «Tradiciones discursivas y cambio lingüístico», *Lexis* 29/2, pp. 151–177.
- (2005d): «Las tradiciones discursivas del español medieval: historia de textos e historia de la lengua», *Iberoromania* 62, pp. 28–43.

- (ed.) (2008): *Sintaxis histórica del español y cambio lingüístico: Nuevas perspectivas desde las Tradiciones Discursivas*, Frankfurt am Main/Madrid: Vervuert/Iberoamericana.
  - (2011): «Diskurstraditionen und Genres», en: Dessì Schmid, Sarah; Detges, Ulrich; Gévaudan, Paul; Mihatsch, Wiltrud; Waltereit, Richard (eds.): *Rahmen des Sprechens. Beiträge zu Valenztheorie, Varietätenlinguistik, Kreolistik, Kognitiver und Historischer Semantik. Peter Koch zum 60. Geburtstag*, Tübingen: Narr, pp. 89–100.
  - (2012): «Modelos matemáticos e substitución lingüística», *Estudos de Lingüística Galega* 4, pp. 27–43.
  - (2012): «Intuición y empirismo», en: Martínez del Castillo (ed.): *Eugenio Coseriu (1921–2002) en los comienzos del siglo XXI, Analecta Malacitana, Anejo LXXXVI*, pp. 99–115.
- Kabatek, Johannes; Obrist, Philipp; Vincis, Valentina (2010): «Clause-linkage techniques as a symptom of discourse traditions: methodological issues and evidence from Romance languages», en: Dorgelo, Heidrun; Wanner, Anja (eds.): *Syntactic Variation and Genres*, Berlin/ New York: Mouton De Gruyter, pp. 247–275.
- Kabatek, Johannes; Pusch, Claus D. (2009): *Spanische Sprachwissenschaft*, Tübingen: Narr.
- Koch, Peter (1997): «Diskurstraditionen: zu ihrem sprachtheoretischen Status und ihrer Dynamik», en: Frank, Barbara; Haye, Thomas; Tophinke, Doris (eds.), *Gattungen mittelalterlicher Schriftlichkeit*, Tübingen: Narr, pp. 43–79.
- (2008): «Tradiciones discursivas y cambio lingüístico: el ejemplo del tratamiento *vuestra merced* en español», en: Kabatek (ed.) 2008, pp. 53–88.
- Koch, Peter; Oesterreicher, Wulf (2007): *Lengua hablada en la Romania: español, francés, italiano*, trad. de Araceli López Serena, Madrid: Gredos.
- Lopes, Célia Regina dos Santos (Ms.): «Tradições Discursivas e Mudanças no Sistema Pronominal de Tratamento no Português Brasileiro: Aplicação de Metodologia para Corpora Diacrônicos».
- López Serena, Araceli (2003): «Algunos aspectos epistemológicos de la lingüística contemporánea», *Res Diachronicae* 2, pp. 212–220.
- (2006): «La edición como construcción del objeto de estudio: el ejemplo de los corpus orales», en: Pons, Lola (ed.): *Historia de la lengua y crítica textual*, Frankfurt am Main/ Madrid, Vervuert/Iberoamericana, pp. 301–334.
  - (2007): «Las tradiciones discursivas en la historiografía lingüística y en la historia de la lengua española», en: Fernández Alcaide/López Serena (eds.) (2007), pp. 49–54.
- Martínez, Angelita (2009): «La frase adjetiva – el orden del sustantivo y del adjetivo», en: Company Company (ed.) (2009a), *Segunda parte: la frase nominal*, vol. 2, pp. 1223–1320.
- Morala, José R. (2002): «Originales y copias. El proceso de castellanización en el área leonesa», en: Echenique Elizondo, María Teresa; Sánchez Méndez, Juan (eds.): *Actas del Quinto Congreso Internacional de Historia de la Lengua Española (Valencia, 31.1–4.2 2000)*, Madrid: Gredos, vol. 2, pp. 1335–1345.
- (2006): «Datos para la cronología del plural *quienes*», en Bustos Tovar, José Jesús de; Girón Alconchel, José Luis (eds.): *Actas del VI Congreso Internacional de Historia de la Lengua Española*, Madrid: Arco Libros, vol. 1, pp. 923–936.
- Oesterreicher, Wulf (1997): «Zur Fundierung von Diskurstraditionen», en: Haye, Thomas; Tophinke, Doris (ed.): *Gattungen mittelalterlicher Schriftlichkeit*, Tübingen: Narr, pp. 19–41.
- (2001): «La «recontextualización» de los géneros medievales como tarea hermenéutica», en: Jacob/Kabatek (2001), pp. 199–232.

- Obrist, Philipp (2008): «La categoría sintáctico-semántica de la concesividad en el español medieval. Para una interpretación discursiva de los datos diacrónicos», en: Company Company/Moreno de Alba (eds.) (2008), vol. 2, pp. 2033–2042.
- Octavio de Toledo y Huerta, Álvaro (2007): «Un rasgo sintáctico del primer español moderno (ca. 1675–1825): las relaciones interoracionales con *interin (que)*», en: Fernández Alcaide/López Serena (eds.) (2007), pp. 421–442.
- (2010): «Direktionalität und Diffusion an den Rändern nominaler Determinierersysteme», Ms. Tübingen.
- Palacios Alcaine, Azucena (1996): «Sobre un tipo de anacoluto en la lengua medieval», *Anacleta Malacitana* 19.1, pp. 89–99.
- Pons Rodríguez, Lola (2007): «Retórica y tradiciones discursivas», en: Fernández Alcaide/López Serena (eds.) (2007), pp. 67–78.
- (2008): «El peso de la tradición discursiva en un proceso de textualización: un ejemplo en la Edad Media castellana», en: Kabatek (ed.) (2008), 197–224.
  - (2010a): «Los marcadores del discurso en la historia del español», en: Loureda Lamas, Óscar; Acín Villa, Esperanza: *Los estudios de los marcadores del discurso, hoy*, Madrid: Arco, pp. 523–615.
  - (2010b): *La lengua de ayer*, Madrid: Arco.
- Pusch, Claus; Kabatek, Johannes; Raible, Wolfgang (eds.) (2005): *Romance Corpus Linguistics II: Corpora and Diachronic Linguistics*, Tübingen: Narr.
- Raible, Wolfgang (1980): «Was sind Gattungen? Eine Antwort aus semiotischer und textlinguistischer Sicht», *Poetica* 12, pp. 320–349.
- Rojó, Guillermo (2008): «Lingüística de corpus y lingüística del español», XV Congreso de la ALFAL [en línea: [gramatica.usc.es/~grojo/Publicaciones/Lgca\\_corpus\\_lgca\\_espanol.pdf](http://gramatica.usc.es/~grojo/Publicaciones/Lgca_corpus_lgca_espanol.pdf), 01/02/2013].
- Sáez Rivera, Daniel (2007): «Tradiciones discursivas, historiografía lingüística y e historia de la lengua», en: Fernández Alcaide; López Serena (eds.) (2007), pp. 89–112.
- Schlieben-Lange, Brigitte (1983): *Traditionen des Sprechens. Elemente einer pragmatischen Sprachgeschichtsschreibung*, Stuttgart: Kohlhammer.
- Schlieben-Lange, Brigitte; Weydt, Harald [mit Beiträgen von Eugenio Coseriu und Hans-Ulrich Gumbrecht] (1979): «Streitgespräch zur Historizität von Sprechakten», *Linguistische Berichte* 60, pp. 65–78.
- Vincis, Valentina (2008): «Técnicas de junción y tradiciones discursivas medievales: el testimonio de los textos», en: Company Company/Moreno de Alba (2008) (eds.), vol. 2, pp. 2201–2222.
- Wilhelm, Raymund (2001): «Diskurstraditionen», en: Haspelmath, Martin; König, Ekkehard; Oesterreicher, Wulf; Raible, Wolfgang (eds.): *Language Typology and Language Universals. An International Handbook*, Berlin/New York: de Gruyter, vol. 1, pp. 467–477.

**Recursos electrónicos relacionados con este tema:**

[www.sfb441.uni-tuebingen.de/b14/tools.html](http://www.sfb441.uni-tuebingen.de/b14/tools.html)  
[www.kabatek.de/discursowww.rae.es](http://www.kabatek.de/discursowww.rae.es)  
[www.corpusdelespanol.org](http://www.corpusdelespanol.org)