

Tercera etapa en los estudios sobre español coloquial: objetivos y requisitos*

Third Stage in Colloquial Spanish Studies: Goals and Requirements for Future Research

Salvador PONS BORDERÍA

Grupo Val.Es.Co./Universidad de Valencia

salvador.pons@uv.es

<https://orcid.org/0000-0001-5788-5506>

Recibido: 17/09/2024. Aceptado: 15/10/2024

A Johannes Kabatek, maestro de hispanistas

Resumen: El presente artículo propone una clasificación de los estudios sobre español coloquial en tres fases: dos retrospectivas y una prospectiva, y estipula una serie de criterios que debería cumplir toda explicación sobre la estructura del español coloquial que se postule como candidata para entrar en dicha tercera fase: trabajo sobre corpus de conversaciones, teoría de segmentación del discurso, anotación de un corpus, procesamiento estadístico de los datos y, finalmente, hipótesis validables –de bajo nivel y de alto nivel– sobre la estructura del lenguaje. A la luz de estos parámetros,

* Este trabajo ha sido posible gracias al proyecto CIPROM/2021/038 «Hacia la caracterización diacrónica del siglo XX» (DIA20), de la Generalitat Valenciana, y al proyecto de I+D+I PID2021-125222NB-I00 «Aportaciones para una caracterización diacrónica del siglo XX», financiado por MCIN/ AEI /10.13039/501100011033/ y por FEDER *Una manera de hacer Europa*.

evalúa el modelo desarrollado por el grupo Val.Es.Co. para concluir que cumple todos los criterios enumerados.

Palabras clave: Análisis conversacional, pragmática, español hablado, segmentación del discurso, Grupo de Investigación Val.Es.Co.

Abstract: *This article proposes a classification of studies on colloquial Spanish into three phases: two retrospective and one prospective. It establishes a series of criteria that any explanation of the structure of colloquial Spanish must meet to be considered for inclusion in this third phase. These criteria include work on a corpus of conversations, a theory of discourse segmentation, corpus annotation, statistical data processing, and, finally, testable hypotheses—both low-level and high-level—regarding the structure of language. In light of these parameters, the article evaluates the model developed by the Val.Es.Co. group and concludes that it meets all the specified criterion.*

Keywords: *Conversational Analysis, Pragmatics, spoken Spanish, discourse segmentation, Val.Es.Co. Research Group.*

1. INTRODUCCIÓN

Los estudios sobre español coloquial son una línea de trabajo original de la lingüística hispánica que constituye una aportación propia al campo de la pragmática y el análisis del discurso. A punto de llegar a los primeros cien años de su inicio (Beinhauer, 1929), resulta conveniente realizar una reflexión retro- y prospectiva sobre esta (sub) disciplina pragmática. Retrospectiva, porque el periodo transcurrido permite hacer una aproximación a su historiografía lingüística y valorar sus principales aportaciones a la luz de la (micro)diacronía transcurrida. Prospectiva porque, ante el estancamiento actual en los estudios sobre el coloquio, los responsables de su desarrollo debemos establecer unos estándares que permitan, tanto a nuestra generación como a las generaciones que se inician en la investigación, marcar unas reglas del juego claras y un horizonte de objetivos tangible.

2. ESTUDIOS SOBRE ESPAÑOL HABLADO. PERSPECTIVA HISTORIOGRÁFICA

Desde un punto de vista historiográfico, y de forma programática, pueden distinguirse tres etapas en el estudio del español coloquial¹:

¹ Para una versión historiográfica alternativa –y ajustada a la época de su publicación– ver Cortés Rodríguez (1999).

2.1. Primera etapa: orígenes

La primera, que se podría denominar *fundacional*, se inicia con Beinhauer (1929), libro de carácter descriptivo que, al ser traducido al español en 1963, se convierte en un estudio teórico del español hablado que inicia una línea de estudios intuitiva, cualitativa y estilística en la que sitúan tanto trabajos realizados sobre la oralidad (Vigara Tauste, 1980, 1992) o mimesis de la oralidad –por ejemplo, los de Seco (1970) o Hernando Cuadrado (1988)– como las primeras propuestas empíricas de grabaciones y transcripciones coloquiales (Criado de Val, 1966) o el proyecto PILEI y sus publicaciones derivadas (Lope Blanch, 1971), que inician la fértil tradición de diseño y recogida de corpus orales en la tradición hispánica, continuada después en proyectos como PRESEEA o Ameresco.

Esta primera etapa se cierra con los estudios de Narbona (en especial, 1989a, 1989b y 1990), que plantean la inadecuación de la sintaxis tradicional para explicar las estructuras del hablado. En efecto, enunciados como

- (1) Es bruto pero bruto.
- (2) Para que lo haga él lo hago yo.
- (3) Como lo agarre, lo mato.

(de)muestran que los valores asociados a las conjunciones de coordinación o de subordinación no se agotan en los descritos por la gramática (ejemplo 1), que la nómima de oraciones subordinadas no da cuenta de todas las estructuras que se pueden dar en el lenguaje hablado (ejemplo 2), o que se pueden identificar oraciones subordinadas bien descritas por la tradición que no cumplen, sin embargo, el criterio de solidaridad entre tipo de oración subordinada y marca de subordinación (ejemplo 3).

La explicación de estos ejemplos, según Narbona, no debería asumir su carácter defectivo, erróneo o simplificado con respecto a las formaciones canónicas de la sintaxis escritural, sino que debería que basarse en criterios propios, que trataran la *sintaxis del coloquio* de forma inmanente sobre una base descriptiva propia. Si la sintaxis no basta para constituirse como base explicativa de dicha descripción, se plantea entonces la cuestión de las bases sobre las que se podría construir esta *nueva sintaxis*. Llamaremos a este dilema *la pregunta de Narbona*, que se puede sintetizar del siguiente modo:

Si la sintaxis no, entonces, ¿qué?

2.2. Segunda etapa: a la búsqueda de una respuesta

La pregunta de Narbona se erige como programa de investigación para una generación que, a principios de los años noventa, inicia su formación como lingüistas

nativos pragmatistas o deriva a los estudios de lengua hablada desde distintas disciplinas (Dialectología –Antonio Briz–, Sociolingüística –Luis Cortés– o Sintaxis –Antonio Narbona–). Se produce, además, la confluencia de los estudios sobre español coloquial con la llegada de las corrientes de estudio pragmático y cognitivista que aterrizan en nuestro dominio en esas mismas fechas. Con estos mimbres, se tejerá la segunda etapa en los estudios del español coloquial.

La obra que mejor representa esta etapa –y esta hibridación– es Briz (1998), que señala de forma explícita el camino que deben tomar los estudios sobre lengua hablada. La conclusión de este trabajo es clara: para responder a la pregunta de Narbona, la respuesta ha de ser de base pragmática.

Esta respuesta se demuestra necesaria pero no suficiente, ya que no existe, en la época, una base descriptiva que permita una comprensión general de la pragmática del español. No es casual que la década de los noventa sea la década de los estudios particulares sobre los marcadores del discurso, el hilo rojo que recorre el desarrollo de la pragmática hispánica y la piedra Rosetta para esta nueva descripción pragmática de la *andadura sintáctica* del coloquio. Las obras de conjunto de Portolés (1998) y de Martín Zorraquino y Portolés (1999), así como la presencia de tres diccionarios sobre marcadores (Santos, 2003; Briz, Pons y Portolés, 2008; Fuentes, 2009) solo se pueden entender sobre la extensa base descriptiva desarrollada de forma continuada durante más de una década. La amplitud de esta base descriptiva se puede apreciar considerando que la nómina de autores que ha colaborado en el *DPDE* supera los cuarenta investigadores, una cantidad enorme si se tiene en cuenta el número de especialistas en pragmática que había en la época en el dominio hispánico.

Cubierta la sincronía, la diacronía toma el relevo en la década de los dos mil y, al igual que sucede con el capítulo de la *NGLÉ*, no se entiende la extensa compilación de Pons Rodríguez (2010) sobre los estudios históricos de marcadores discursivos sin la base descriptiva previa que marca dicha década. La asimilación de los estudios sobre gramaticalización y la llegada de una generación de jóvenes gramáticos históricos con formación pragmática y filológica (Araceli López, Lola Pons, Margarita Borreguero, Javier Rodríguez, Álvaro Octavio de Toledo, entre otros) hizo posible un salto adelante en los estudios de sintaxis histórica que no se habría podido prever siquiera diez años antes.

Pero la estructura del coloquio es mucho más que los omnipresentes marcadores del discurso. Desde principios de los dos mil, los estudios sobre cortesía y conceptos afines (descortesía, anticortesía, cortesía codificada/interpretada) (Bravo, Briz (eds.), 2004) llaman la atención sobre estructuras y estrategias de la oralidad dirigidas a mantener la relación entre los interlocutores. La dimensión social del discurso oral comienza a desarrollarse, favorecida también por un detalle externo: la imbricación de la formación lingüística con la antropológica y la sociológica en muchos investigadores (especialmente los formados en Hispanoamérica) casa bien con este tipo de estudios, híbridos por naturaleza y alejados de la teorización lingüística que caracteriza los trabajos sobre

marcadores del discurso. Se abre así el camino hacia la pragmática sociocultural, que es la dimensión del discurso predominante en la última década, simbolizada en el dominio hispánico por los estudios sobre atenuación e intensificación (Briz y Albelda, 2013), evidencialidad (Albelda, 2016) o sobre géneros discursivos digitales (Yus, 2020).

Por último, los estudios sobre el lenguaje hablado se han incorporado a las corrientes romanísticas de estudios sobre las unidades de lo hablado (Roulet *et al.*, 1985, 2001; Blanche-Benveniste y Jeanjean, 1987) con varios modelos teóricos que abordan la construcción del coloquio (Briz y Grupo Val.Es.Co., 2003; Cortés y Camacho, 2005; Fuentes, 2019; Garrido Medina y Rodríguez Ramalle, 2015; Pons Bordería, 2022). Estos modelos constituyen una alternativa no sintáctica para explicar la organización del material lingüístico en el lenguaje hablado y se perfilan como el camino más prometedor para responder a *la pregunta de Narbona*, pero no pueden desarrollarse en el vacío; en este sentido, las numerosas descripciones particulares sobre unidades, funciones y géneros del español hablado son una precondition para la construcción de una teoría sólida sobre unidades discursivas.

Es destacable, además, que el único trabajo que agrupa todas las perspectivas románicas parta de la iniciativa española (Pons Bordería (ed.), 2014), así como que la extensión de dichos modelos al discurso escrito se haya planteado, también de forma global, desde la lingüística hispánica (Salameh y Pardo, 2024).

Vista en conjunto, esta segunda etapa puede considerarse un gran logro de nuestra lingüística que, si bien llegó con retraso a los estudios pragmáticos, tras treinta años de fértil trabajo investigador ha convertido al español en una de las lenguas más y mejor descritas a nivel pragmático. En un artículo programático, Narbona (1989b) advertía de que el avance en los estudios sintácticos solo se podría realizar mediante la incorporación de dos dominios o, en palabras del propio autor, *apellidos*: la sintaxis debería ser *histórica* y *coloquial*. Puede decirse que esta segunda etapa ha contribuido, con un enorme esfuerzo descriptivo de la comunidad investigadora, a la genealogía de la sintaxis que planteaba el autor sevillano.

2.3. Necesidad de una tercera etapa

Dicho esto, creemos que se hace necesario abordar una nueva y tercera etapa en los estudios sobre el coloquio. Parafraseando a Sperber y Wilson (1986), se podría afirmar que la teoría sobre el español hablado se arriesga a la vaguedad si es incapaz de ir más allá de los estudios particulares –necesarios, pero atomísticos– y si no puede ofrecer una caracterización sobre la sintaxis del coloquio que no sea puramente negativa (la sintaxis del coloquio *no* es sintaxis oracional).

Las descripciones realizadas hasta ahora son un requisito *necesario* para el avance de la disciplina; sin ellas, el conocimiento sobre la estructura del español hablado sería

mucho menor del que el que poseemos hoy en día. Sin embargo, no son un requisito *suficiente*. En primer lugar porque, como resultado de las acciones de investigaciones particulares, muestran una sobreabundancia de publicaciones en ciertos temas (entre ellos, los marcadores del discurso, así como la caracterización de géneros discursivos cercanos al coloquio o coloquializados, como debates, tertulias y entrevistas) y una evidente carencia en aspectos no menos importantes, como la variación dialectal, la estructura prosódica, la organización de las unidades discursivas, la estructura interactiva de la conversación, los elementos subestructurales, etc.

En segundo lugar, porque abordan la descripción y explicación de las categorías y funciones del español coloquial desde métodos y perspectivas heterogéneos, en ocasiones difícilmente comparables: de los estudios puramente lingüísticos a los multimodales; de la pragmática experimental (Loureda *et al.*, 2021) a la lingüística con corpus, o de las descripciones teóricas a las encuestas socio- y sicolingüísticas (como en los estudios derivados de los proyectos PRESEEA y Ameresco). Estos estudios, además, se fundamentan en bases teóricas tan dispares como el análisis de la conversación y la teoría de la relevancia, los estudios sobre cortesía, epistemicidad o sobre atenuación e intensificación; la teoría de la argumentación, la gramática de construcciones, las teorías sobre el humor o los estudios de gramaticalización. Esta multiplicidad de perspectivas no es de extrañar, dada la confluencia histórica (ver sección 2) de los estudios de pragmática con la tradición coloquial española.

En tercer lugar porque, de una forma sorpresiva, se ha producido un salto atrás en los estudios sobre el coloquio. Vuelven a aparecer publicaciones que se preguntan por la naturaleza y definición de los marcadores del discurso, obviando los límites de la clase y tomándola por una categoría irrestricta, en la que todo cabe. Estas publicaciones además, frecuentemente prescinden de referencias básicas de la bibliografía sobre el tema; se redescubren marcadores ya descritos desde una misma óptica teórica. Asimismo, se analiza como válido cualquier acontecimiento comunicativo, como el «discurso de las declaraciones de los futbolistas», lo que amplía de forma fractal el campo de investigación, ya que dichas investigaciones se podrían replicar en el discurso de cualquier rol (el discurso de las declaraciones de los actores, de los bomberos, de los modelos o, en términos más generales, *el discurso de las declaraciones de los X*).

Asimismo, se presentan estudios de corpus que se basan en muestras ínfimas (llamadas, en línea con el AC, *colecciones*, consistentes en cinco, ocho o diez ejemplos, de uno o de dos hablantes). Se llega incluso a publicar resultados negativos, obviando una de las líneas rojas del quehacer científico: retorciendo la cita atribuida a Edison, en el discurso científico solo se debe publicar la forma correcta de hacer una bombilla, no las mil formas incorrectas de no hacerla (recuérdese, a este respecto, la crítica que hacían Sperber y Wilson (1986, pp. 95-102) a las reglas de introducción en lógica formal y a sus consecuencias cognitivas sobre la memoria humana; aplicado a

la investigación sobre el coloquio, esto equivaldría a llenar las publicaciones científicas de artículos-ruido, que no añaden nada al estado de la cuestión).

Para superar esta fase, y para responder a la pregunta de Narbona, creemos que resulta una buena estrategia teórica plantear una serie de requisitos mínimos que deba cumplir toda teoría sobre la estructura del coloquio para ser considerada como un acercamiento válido, que permita avanzar hacia esta tercera fase. Tales requisitos pueden funcionar como hitos mensurables, objetivos y objetivables en el camino de la construcción de esta sintaxis del habla; la conquista del territorio, ya explorado pero con grandes áreas vírgenes, de la lengua saussureana.

Y, dado que las teorías sobre unidades discursivas son las únicas que permiten, hoy en día, vislumbrar una respuesta a la pregunta de la organización de la materia hablada, dichos requisitos hacen posible contrastar las distintas teorías de unidades discursivas con unas unidades de medida independientes contra las que comparar sus logros.

3. REQUISITOS MÍNIMOS DE UNA TEORÍA DE UNIDADES PARA EL ESTUDIO DEL ESPAÑOL COLOQUIAL

Enumeramos a continuación los requisitos anunciados arriba, que se dividen en seis criterios principales y tres secundarios. Los requisitos principales funcionan como precondiciones para considerar validada una teoría sobre las unidades del español coloquial, mientras que los secundarios añaden valor a sus resultados. Son los siguientes:

1. Trabajar sobre un corpus de conversaciones coloquiales
 - a. Desarrollar un corpus de conversaciones coloquiales propio
2. Trabajar con un modelo de unidades de lo oral
 - a. Desarrollar un modelo de unidades de lo escrito.
3. Anotar un corpus de conversaciones coloquiales
4. Validar estadísticamente los resultados de la anotación
5. Desarrollar un conjunto de resultados contrastables sobre la estructura del español coloquial conversacional
6. Desarrollar hipótesis de alto nivel sobre la estructura del español hablado
 - a. Desarrollar hipótesis de alto nivel, a partir de la estructura del español hablado, sobre problemas de la Lingüística General.

3.1. Trabajo sobre corpus de conversaciones coloquiales

La primera de estas condiciones es la que se nos antoja más básica de todas. Pedir que una teoría sobre la estructura del coloquio trabaje sobre un corpus de

conversaciones coloquiales puede parecer un requisito evidente y, sin embargo, no lo es. La conversación coloquial es la base de pruebas más dura que existe, dado que refleja las estructuras de la lengua hablada tal y como se manifiestan en las conversaciones coloquiales prototípicas (Briz *et al.*, 1995). El coloquio, con sus estructuras inacabadas, elementos subestructurales, cambios de tópico y toma de turno no predeterminada, plantea problemas descriptivos en todos los niveles de su estudio. Resulta tentador, a este respecto, aplicar la teoría para el registro coloquial a géneros discursivos cercanos a la conversación pero más amables, como las entrevistas (semidirigidas), los debates o las tertulias. Y cierto es que presentan muchos rasgos en común con la conversación coloquial, especialmente en el caso de los más coloquializados, pero también lo es que es esta última, como género prototípico, presenta una mayor variedad descriptiva tanto desde el punto de vista social (número de FTAs en una conversación coloquial ~ entrevista semidirigida –Albelda, 2008–), interactivo (tipo de figuras conversacionales en una conversación coloquial ~ entrevista –Espinosa Guerri, 2024–) o estructural (número de elementos subestructurales –Pascual, 2020–).

3.1'. Desarrollo de un corpus propio

La investigación resulta mucho más completa si los autores de la teoría han desarrollado su propio corpus. En efecto, el número de horas necesario para transcribir, corregir o preparar una conversación hasta su puesta a disposición de la comunidad investigadora desarrolla el mismo tipo de saber idiomático que desarrolla el investigador de historia de la lengua al preparar la edición crítica de un texto antiguo. Este saber se refleja, a su vez, en la teoría que se produce.

En este sentido, resulta necesario hacer una salvedad, y es que la unión de un conjunto de muestras procedentes de distintos géneros de la oralidad o de la escrituralidad no constituye un corpus, sino un archivo de textos (Sinclair, 1996). Para que un corpus sea considerado como tal tiene que cumplir las condiciones de ser procesable electrónicamente, de haberse producido en entornos comunicativos naturales, y de haber sido compilado con la intención de estar equilibrado y de ser analizado lingüísticamente (Gries, 2009, p. 7). Con estos criterios, algunos de los corpus del dominio hispánico deberían pasar a llamarse archivos de textos.

3.2. Uso de un modelo de unidades de lo oral

La respuesta a la pregunta de Narbona implica un programa de investigación que cumpla un requisito paradójico: debe producir una sintaxis sin recurrir a la sintaxis, aunque sería más exacto decir que debe producir una *sintaxis* sin recurrir a la *sintaxis oracional*. En efecto, la incapacidad de la sintaxis oracional para producir una

explicación sin residuo de las producciones lingüísticas del español oral conversacional tiene que ver con la esencia misma de la disciplina: etimológicamente, este término significa, según el *DRAE*, «disponer conjuntamente, ordenar». La sintaxis estudia las relaciones que se crean cuando dos o más elementos coaparecen y dichas relaciones (de interdependencia, determinación y constelación, según Hjelmslev (1984:1943, pp. 55-64) se reconocen por medio de métodos probatorios (Rabanales, 1987).

Las relaciones sintácticas tienen un límite superior, que se encuentra en un lugar de límites imprecisos denominado, en la tradición hispánica, *oración*, *proposición*, *cláusula* o, de una forma más abstracta, *proyección sintagmática superior*. Por encima de dicho límite (más allá de la atmósfera del planeta Sintaxis), dejan de distinguirse sintagmas, relaciones de núcleo a complemento o funciones sintácticas, y deja de aplicarse la mayor parte de los métodos probatorios utilizados en el campo.

Además, la gramática es también una disciplina construida históricamente mediante procesos culturales; es una tradición discursiva (Kabatek (ed.), 2008) o, quizás, un conjunto de tradiciones discursivas cuyas clases y funciones remiten a una historicidad que se puede trazar historiográficamente (Gómez Asencio, 1981, 1985; Calero Vaquera, 1986, entre otros).

Por todas estas razones, la organización de la materia hablada en el español oral conversacional no puede ser sintáctica (o únicamente sintáctica), pero ha de dar cuenta del orden (de la *syn*), y de su *taxis*, de la ordenación entre los elementos que coaparecen. Y para esto se hace necesario un modelo.

Los modelos de unidades discursivas desarrollados en la lingüística románica son la respuesta a este problema y, de su análisis (Borreguero, Athayan y Grosse, 2018), se deriva la presencia de una serie de ingredientes comunes: la incorporación de relaciones pragmáticas y la valoración de la información prosódica, las relaciones semánticas, e incluso la presencia de relaciones sintácticas. Cada modelo combinará estos ingredientes de formas distintas. Así, la prosodia es determinante en el modelo L-AcT (Cresti, 2000), es el cincuenta por ciento de la información en el modelo de Degand y Simon (2009), y tiene una importancia no cuantificada en el modelo Val.Es.Co. (Pons Bordería, 2022).

A pesar de las grandes diferencias entre los distintos modelos de segmentación discursiva, se puede enunciar una serie de requisitos exigibles a toda teoría de segmentación de lo oral. Tales requisitos se pueden cumplir de formas diversas, lo que no quita que no se puedan establecer unos mínimos deseables:

- a) Debe ser capaz de segmentar una conversación coloquial sin residuo. Esto quiere decir que su aplicación no debe limitarse a un conjunto de ejemplos particularmente favorables a la teoría, sino que tiene que ofrecer un análisis omnicompreensivo, en los planos monológico y dialógico, de todos los elementos de una conversación, de forma semejante a como se procede en un análisis sintáctico.

- b) Debe ofrecer una segmentación coherente con los datos sintácticos y prosódicos. La segmentación sintáctica oracional y la conversacional de base pragmática tienen puntos de contacto; fronteras en las que el segundo análisis entra en el dominio del primero o en las que el primero se expande al territorio del segundo. En estas fronteras (piénsese, por ejemplo, en el resbaladizo territorio de las llamadas *conjunciones ilativas* –RAE, 1771–) ambos análisis deberían ser traducibles; de lo contrario, estaríamos ante fenómenos completamente distintos, y no ante dos formas diferentes de analizar el mismo fenómeno. Lo mismo vale para el análisis prosódico. Aunque la prosodia, para las necesidades de un modelo de segmentación, sobregenera información (Cabedo, 2014), los parámetros prosódicos deberían estar vinculados al modelo de segmentación en una proporción que cada modelo deberá determinar.
- c) Debe ofrecer un equilibrio entre los distintos criterios escogidos. Ya se base el modelo en criterios sintácticos y prosódicos (Crible y Degand, Morel), o semánticos, pragmáticos y prosódicos (Val.Es.Co.), la relación entre todos ellos tiene que estar motivada y hacerse explícita en la medida de lo posible; es decir, en la medida del avance de sus investigaciones.
- d) Debe poder aplicarse en estudios micro y macro. Si el objeto de estudio es la conversación coloquial, esta debe poder explicarse tanto en su nivel interactivo (relaciones entre los participantes cuyo ámbito mínimo es la intervención) como en su nivel monológico (relaciones entre constituyentes cuyo ámbito máximo es la intervención). Explicar uno solo de estos niveles es legítimo, pero produce un resultado incompleto.

3.2'. Uso de un modelo propio de unidades de lo escrito

Con la ayuda de los modelos de unidades de lo oral se hace posible la descripción del uso de los marcadores del discurso en las conversaciones coloquiales, un hito descriptivo imposible para la sintaxis oracional ya que estos elementos, desde su propia definición (Portolés, 1998; Fraser, 1990, 1999), se sitúan fuera del ámbito sintáctico, a pesar de los intentos para expandirlo (Rizzi, 1997 y toda la bibliografía generativista sobre la periferia izquierda de la oración).

Sin embargo, los mismos marcadores del discurso plantean un problema a dichos modelos, ya que la mayor parte de ellos aparece con mayor frecuencia en géneros discursivos propios de la escrituralidad (Pons, Pardo y Alemany, 2023), especialmente en relaciones textuales de tipo extraoracional vinculadas a unidades como el párrafo (Ferrari *et al.*, 2008; Ferrari y Borreguero, 2015). Estos usos plantean la necesidad de una teoría de unidades discursivas de la escrituralidad para conseguir una explicación unitaria del funcionamiento de los marcadores discursivos. Aunque este objetivo es secundario para los fines que se plantean en este trabajo, una teoría de unidades de lo

oral que posea, además, una extensión a las unidades de lo escrito será descriptivamente superior a una teoría de unidades sin ella, ya que producirá una respuesta más completa al problema de los marcadores del discurso.

3.3. Segmentación de un corpus de conversaciones coloquiales

Todo modelo de unidades de lo oral que desee superar la fase inicial deberá testarse contra la base de pruebas más dura que existe: las estructuras de la lengua hablada, tal y como se manifiestan en las conversaciones coloquiales prototípicas (Briz *et al.*, 1995). En este sentido, la segmentación sin residuo de un corpus de conversaciones coloquiales nos parece el criterio básico para determinar la mayor o menor idoneidad de un modelo de segmentación.

Al confrontarse contra un corpus conversacional extenso, las deficiencias, inexactitudes e incoherencias de un modelo saltan a la vista y obligan a cuestionar la teoría y a adaptarla para obtener una descripción más ajustada a la realidad. Este proceso necesariamente producirá un número alto de segmentaciones incorrectas: si los análisis sintácticos siguen suscitando debates teóricos después de siglos de teoría gramatical, ingenuo sería suponer algo distinto para la segmentación de la oralidad; pero, por muy equivocado que sea el análisis, permitirá, sobre dicha base, acercarse a una mejor solución.

La comprobación empírica de la adecuación de un modelo de unidades discursivas constituye, a nuestro entender, el tajamar que divide los modelos en dos: los que han contrastado su adecuación mediante un estudio empírico y los que todavía no. Los primeros modelos pueden considerarse *validables*; los segundos, *especulativos*.

3.4. Validación estadística de la segmentación

Al hilo del punto 3, la validación estadística de los resultados es un criterio necesario hoy en día en la presentación de resultados y, en especial, en la lingüística con corpus. Esto no quiere decir que la aplicación de una prueba inferencial convierta el estudio que lo contiene en correcto (los lingüistas tenemos mucho que aprender todavía sobre los *caveat* que lanzan los estadísticos a la aplicación de determinadas técnicas sin conocer sus fundamentos), pero sí que le aporta una capa extra de validación, lo que es especialmente importante en el caso de modelos que compiten en un mismo nicho teórico.

Para que la validación sea útil, primero se ha de establecer un número fijo de hipótesis concretas, cuya confirmación estadística aporte solidez a las mismas. Las hipótesis preexisten a las pruebas estadísticas (en el sentido de que las seleccionan) y,

al mismo tiempo, se ven afectadas por dichas pruebas (en el sentido de que pueden ser validadas, validadas parcialmente o desconfirmadas como resultado del análisis).

3.5. Conjunto de resultados contrastables

Como consecuencia del proceso descrito en (3), y de su validación en (4), los modelos validables dispondrán de un conjunto de hipótesis de bajo nivel sobre estructuras concretas de la oralidad, ya sean estructuras suspendidas, relaciones entre elementos conceptuales y procedimentales o relaciones interactivas entre los participantes. Dichas hipótesis permiten su discusión detallada e individualizada, lo que permite la *ampliación por concreción* de los estudios sobre lengua hablada. Esto aporta, a la vez, precisión y profundidad al campo de estudio. Por ejemplo, en lugar de señalar de forma vaga la importancia de la prosodia en la estructura del coloquio, se puede discutir en qué medida la prosodia aporta información para la delimitación de unidades procedimentales y si esta medida es diferente en función del carácter textual, modal o interactivo de dichas unidades; se pueden diseñar estudios confirmatorios para determinar el valor desambiguador de la prosodia en un entorno dado y cuantificar dichos resultados mediante pruebas estadísticas inferenciales, las cuales, a su vez, establecen un punto de partida para confirmar o desconfirmar dichos resultados iniciales en nuevos estudios, que pueden ser, de carácter introspectivo o experimental.

3.6. Hipótesis de alto nivel sobre el español coloquial

Los modelos de segmentación son un medio para conseguir un fin. Y este fin no es otro que la respuesta a la pregunta de Narbona, de modo que el resultado de un modelo de segmentación debería producir hipótesis fuertes acerca de la organización de la materia hablada. No basta, en efecto, con etiquetar una serie de construcciones; las etiquetas proporcionadas por el modelo deberían reflejar una idea falsable sobre qué principios subyacen a dichas construcciones. Si, tras la aplicación de una segmentación, nuestro conocimiento sobre la estructura del coloquio no ha aumentado, la segmentación no resulta productiva (describe pero no explica).

3.6'. Hipótesis de alto nivel sobre problemas generales

Las hipótesis derivadas de la aplicación de una segmentación pueden referirse únicamente al objeto estudiado o pueden ser de aplicación a otras lenguas. Desde la aparición del generativismo, la Lingüística General vive bajo el dominio de la estructura del inglés; lo descrito para dicha lengua se eleva automáticamente a rasgo general de

todo lenguaje natural. Si, desde la Lingüística Románica, se desarrolla una teoría sobre la organización de la materia hablada, es objetivo secundario –pero deseable– de dicha teoría tener vocación generalista y buscar, por ello, su aplicación a otras lenguas. De este modo, se reivindica la capacidad de nuestras lingüísticas para crear lingüística teórica.

4. EL MODELO VAL.ES.CO. A LA LUZ DE LOS REQUISITOS MÍNIMOS

Dados los requisitos que permiten medir la adecuación y la eficacia de los distintos modelos descritos en la Sección 3, pasamos ahora a cerner el modelo Val.Es.Co. por el tamiz que hemos descrito previamente, para determinar, de una forma objetiva, cuántos de dichos parámetros cumple.

4.1. Trabajo sobre corpus de conversaciones coloquiales

El centro de la actividad del grupo Val.Es.Co. durante la mayor parte de sus treinta años de actividad ha sido la conversación coloquial, objeto de estudio de la mayoría de las tesis doctorales defendidas a lo largo de dos generaciones de lingüistas.

En el caso del corpus Val.Es.Co., la adopción de nuestro sistema de transcripción por otros investigadores, su extensión a géneros distintos a la conversación coloquial y su aplicación a macrocorpus coloquiales como Ameresco son indicios de la validación por pares de nuestro sistema. Desde un punto de vista interno, los resultados del acuerdo entre anotadores (Pons Bordería y Pascual, 2020) (ver Sección 4) informan de la alta coherencia interna de nuestro sistema de segmentación.

4.1'. Desarrollo de un corpus propio

El grupo Val.Es.Co. dispone, desde 1995, de un corpus de conversaciones coloquiales que, en su versión actual (Corpus Val.Es.Co. 3.0, www.valesco.es), consta de 72 conversaciones y unas doscientas cincuenta mil palabras. Es de acceso abierto, está etiquetado y anonimizado, incluye audios y se puede consultar en modo karaoke (Figura 1):

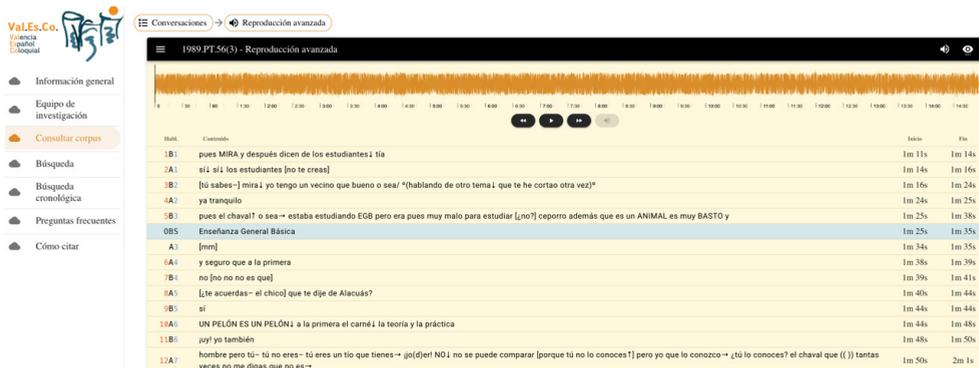


FIGURA 1. Conversación en modo karaoke

Permite búsquedas (Figura 2) por metadatos y los resultados de la búsqueda se pueden exportar en formato.docx y .xml.

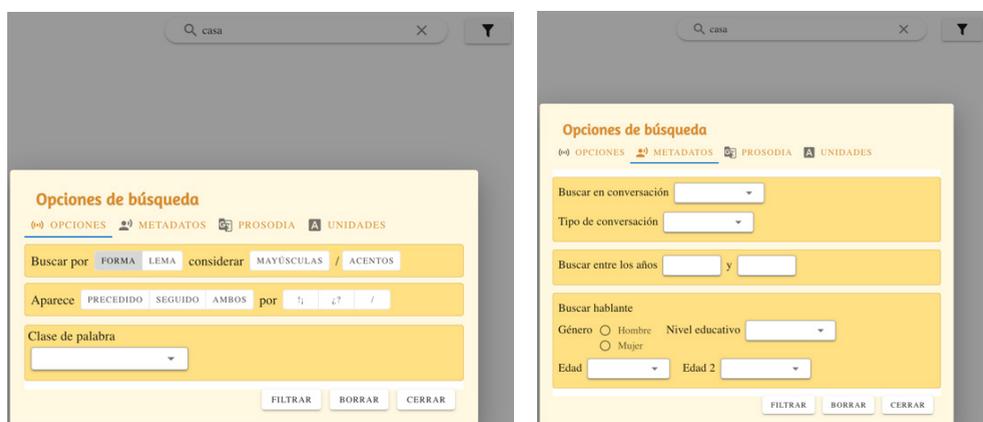


FIGURA 2. Diferentes tipos de búsqueda en el corpus Val.Es.Co. 3.0.

El proceso para el tratamiento de las conversaciones, desde que se registran hasta que se incorporan al corpus, está descrito en Pons Bordería (2022) y comprende veintidós operaciones consecutivas diferentes (ver Figura 3).

Todos los miembros del grupo han pasado, en algún momento de su carrera, por la fase de grabación y de preparación de conversaciones, una tarea propedéutica que estimula su sensibilidad hacia los fenómenos conversacionales y hacia la estructura de la conversación. En este sentido, el proceso de incorporación de los nuevos miembros ha pasado por las etapas de redacción de trabajos de fin de grado y de fin de máster sobre temas relacionados con la conversación coloquial, a partir de los cuales se les ha introducido en cuestiones metodológicas, de análisis de ejemplos, de tratamiento de los datos y de evaluación de resultados propias de la lingüística con corpus. Al hilo

de este trabajo, han adquirido la base teórica necesaria para abordar los problemas objeto de estudio de sus investigaciones.

APÉNDICE VII: PROTOCOLO DE REVISIÓN DE CONVERSACIONES (CORPUS VALES.CO. 3.0)

ID CONVERSACIÓN	PERMISOS	FICHA TÉCNICA	ELAN/WORD	WAV/MP3	REVISIÓN TRANSCRIPCIÓN					
					IDENTIFICACIÓN PARTICIPANTES	ALINEACIÓN	ETIQUETAS	ANONIMIZACIÓN	DOCUMENTO ANONIMIZACIONES	ORTOGRAFÍA
1	SÍ		ELAN	MP3	SÍ	SÍ	SÍ	SÍ	SÍ	SÍ
2			ELAN	MP3	-	-	-	-	-	-
3	NO	SÍ	ELAN	WAV	SÍ	SÍ	-	SÍ	-	-
-	-	-	-	-	-	-	-	-	-	-

CAMBIAR ALMOHADILLA POR ETIQUETA	TOKE NIZAR	ANONIMIZAR AUDIO					DRIVE	DTD (revisión etiquetas)	WEB		OBSERVACIONES REVISORES
		.TXT	EXCEL	AUDIO ANONIMIZADO	AUDIO VINCULADO A ELAN	ARCHIVOS SUBIDOS (s)			SUBIDAS TRAS REVISIÓN	COMPROBAR DATOS	
SÍ		SÍ	SÍ	SÍ	SÍ	SÍ					
-		-	-	-	-	-					
-		-	-	-	-	-					
-		-	-	-	-	-					

FIGURA 3. Protocolo para el tratamiento de las conversaciones (Pons Bordería 2022)

Esta base formativa se ha continuado en la etapa correspondiente a la tesis doctoral. Se puede decir que nuestros miembros se han formado en el trabajo sobre conversaciones coloquiales, trabajo que se ha reforzado en las numerosas reuniones grupales para abordar temas comunes, como la preparación de trabajos conjuntos o el desarrollo de la teoría de unidades, que no por casualidad lleva una firma colectiva (casi un hápax en la lingüística española).

4.2. Uso de un modelo de unidades de lo oral

El modelo Val.Es.Co. se basa en los trabajos descritos en (1) sobre la conversación coloquial, que ha sido su objeto de estudio principal desde principios de los años noventa. En este sentido, la estructura de la teoría está radicada en este proceso de observación y estudio.

Dicho modelo, en su última versión, está descrito en Pons Bordería (2022) y sus características principales son las siguientes²:

Consta de ocho unidades que van, desde la mínima monológica hasta la máxima dialógica: subacto, acto, intervención, turno, intercambio, alternancia de turno, diálogo

2 Tras esta fecha, se han añadido algunas modificaciones debidas a las últimas investigaciones realizadas en el grupo como, por ejemplo, la introducción de la posición simultánea (Cabanés, 2023) para dar cuenta del paralenguaje que se produce a la vez que la producción lingüística.

y discurso³; cinco posiciones (inicial, inicial relativa, media, final e independiente) y tres órdenes (estructural, social e informativo). Cada unidad se define de forma independiente, se caracteriza por un rasgo distintivo y se reconoce por una serie de marcas formales. Es un modelo de base fundamentalmente pragmática, definido y pensado para el análisis de la conversación coloquial.

En este sentido, su antecedente más directo es el modelo ginebrino de Eddy Roulet en su primera versión (Roulet *et al.*, 1985), ya que dicho modelo se proponía como válido para el análisis tanto de conversaciones como de textos escritos, pero sus ejemplificaciones se basaban en acontecimientos comunicativos transaccionales, como diálogos en librerías. Desde la observación de las conversaciones coloquiales, la teoría del grupo de Ginebra parecía poco adecuada para este objeto de estudio, lo que estimuló el desarrollo de un modelo de segmentación específico, que se ha ido desarrollando desde su primera formulación, en 2003, y que se ha aplicado a diferentes géneros orales, además de la conversación coloquial.

4.2'. Uso de un modelo propio de unidades de lo escrito

Este modelo posee una extensión de unidades aplicables a la escrituralidad en el que las unidades dependientes de la estructura interactiva de la conversación (intervención, intercambio, y sus contrapartidas sociales) se sustituyen por unidades paralelas de carácter estructural que ofrecen una organización jerárquica de los diferentes géneros escritos:

Unidades de lo oral	Id. de lo escrito
Subacto	
Acto	
Intervención/turno	Conjunto de actos
Diálogo	Movimiento
Discurso	

FIGURA 4. Unidades de lo oral y unidades de lo escrito en el modelo Val.Es.Co.

Tal extensión se justifica como necesaria para el tratamiento de los marcadores discursivos, que presentan especializaciones en los centros categoriales del continuo diafásico, y para cuyo tratamiento unitario se ha concebido, especialmente a partir de funciones como la llamada digresión (Estellés y Pons Bordería, 2008) o de la aplicación

³ A estas hay que añadir el *residuo*, que no es una unidad propiamente dicha, sino un elemento identificable y etiquetable.

El resultado de esta segmentación muestra que es posible someter el habla saussureana a un proceso de análisis lingüístico que reproduzca, desde principios pragmáticos, un proceso similar al que se somete la oración mediante el análisis sintáctico. De este modo, el modelo del grupo Val.Es.Co. ofrece una primera respuesta a la pregunta de Narbona: sí, es posible ofrecer una explicación alternativa a la estructuración del habla coloquial que no haga referencia, ni a esquemas sintácticos expandidos, ni a nociones como *desvío*, *falta* o *defectividad*. En este sentido, y siguiendo el criterio expuesto en (2), el modelo Val.Es.Co. sería, a día de hoy, un modelo sobre la estructura de la conversación coloquial al que se puede aplicar el adjetivo *validable*, lo que lo diferencia de los modelos que no hayan pasado por este filtro que, aplicando el este mismo criterio, pueden ser denominados *especulativos*.

4.4. Validación estadística de la segmentación

Una segmentación como la presentada en (4.3) puede parecer altamente subjetiva; de hecho, pasarán años antes de que se disponga un cuerpo de segmentaciones ampliamente aceptado e incorporado al arsenal de herramientas de los analistas del lenguaje hablado. Por esta razón, la segmentación de nuestro grupo se ha validado con el procedimiento estándar en este tipo de casos, que es el acuerdo entre anotadores (*inter-annotator agreement*). Este procedimiento, apenas conocido en la lingüística hispánica (con escasas y muy recientes aplicaciones, como De Latte (2024), busca medir la validez de un análisis (normalmente, el etiquetado de un corpus) mediante su réplica en varios investigadores (normalmente, dos) que, de forma independiente y sobre la base de unos criterios comunes, realizan una misma tarea y comparan después sus resultados. El contraste se valida estadísticamente y se comprueban los índices de homogeneidad en el análisis, lo que permite extrapolar la estabilidad del etiquetado si fuera replicado por un nuevo grupo de investigadores.

Un ejemplo de este proceso se puede ver en Crible y Degand (2019): dos anotadores etiquetaron 423 muestras de treinta y tres tipos de marcadores del discurso a partir de dos variables: cuatro dominios (ideacional, retórico, secuencial e interpersonal) y una nómina de quince funciones.

La prueba correspondiente para el etiquetado del modelo Val.Es.Co. (Pons Bordería y Pascual Aliaga, 2020) posee una complejidad mucho mayor, ya que no se limita a etiquetar una clase de palabras, sino que se extiende a las etiquetas de los constituyentes de toda una conversación de diecinueve minutos (4352 palabras), ha sido realizada por tres investigadores, frente a los dos habituales, con el objetivo de comprobar la identificación de subactos y su etiquetado en tipos subactos.

Para esta última tarea, se utilizaron diez etiquetas distintas (SSD, SSS, SStop, SAM, SAT, SAI, SAX, SSX, SXX y R[esiduo]). A diferencia de análisis previos, que se limitan a

etiquetar un elemento bien definido, en este caso la segmentación implica: a) la identificación de unidades en el eje temporal; b) la identificación de los límites de cada unidad (por ejemplo, que los elementos fragmentarios se identifiquen como R o se incluyan en un subacto que lo hospede); c) la coincidencia en el etiquetado de subactos conceptuales y procedimentales, así como la de los tipos de subactos.

Estos resultados se miden, no con un estadístico, sino con la familia de estadísticos denominada *alfa de Krippendorf*, que implica cuatro medidas diferentes (ua, |ua, cua y (k)ua). Los resultados arrojan una media de coincidencia cercana al 0.8 en casi todas las mediciones. Esta es una validación estadística muy precisa a partir de unos estándares de exigencia superiores a los utilizados en el campo hasta ahora, lo que se debe tomar como una muestra de alta homogeneidad en el análisis. Modelos de segmentación alternativos deberían someterse a una validación de este tipo para garantizar que sus análisis pueden replicarse (algo que solo se puede comprobar, dicho sea de paso, si se ha cumplido el tercer requisito enumerado arriba).

4.5. Conjunto de resultados contrastables

La teoría sobre las unidades del discurso desarrollada y el proceso de segmentación aplicado producen una serie de hipótesis de alto y de bajo nivel sobre la estructura del español hablado y sobre la organización de la conversación en general. Estas hipótesis se pueden contrastar y, consecuentemente, confirmar o desconfirmar, lo que ofrece unas bases sobre las que continuar la tarea investigadora y afinar los resultados. Enumeramos a continuación algunas de las de bajo nivel (en el caso de datos, las hipótesis se refieren siempre al corpus analizado):

- a) Entre la unidad mínima monológica y los residuos existe una gradación de constituyentes denominados *subestructuras*: *subactos integrables*, *residuos integrables* y *residuos* (Pascual, 2020). Este constituye el límite inferior de la segmentación lingüística en la conversación coloquial.
- b) La introducción del paralenguaje en el análisis implica la introducción de una nueva posición, denominada *simultánea*, en la que un gesto se produce al mismo tiempo que el mensaje lingüístico (Cabanés, 2023).
- c) El rango tonal del tonema suspendido se da en el intervalo $[\pm 3]$ ST.
- d) La velocidad media de la conversación coloquial se sitúa en el entorno de 3,58 palabras por subacto y segundo (Pons, Salameh y Badía, 2025).
- e) *Decir* introduciendo estilo directo se analiza como un SAT que introduce tres tipos de unidades: un discurso (*decir₁*), una intervención (*decir₂*) o un acto/subacto (*decir₃*) (Benavent, 2023).

- f) La teoría de unidades y posiciones se puede validar experimentalmente mediante estudios de *eye-tracking*: para el marcador *o sea*, las distintas configuraciones estructurales se correlacionan con distintas medidas en el procesamiento de la lectura (Salameh, 2021).
- g) La introducción de la multimodalidad en la transcripción añade un 18% más de intervenciones a la conversación. De esas, el 58% son intervenciones que se realizan de manera únicamente gestual.
- h) La introducción de la multimodalidad en la transcripción modifica un 9% de las intervenciones existentes (264 de 3037) (Cabanés 2023).
- i) Las posiciones más frecuentes en las que se sitúan los marcadores discursivos (para ambos registros) son [I, A], [Irel, A], [M, A], [F, A] y [M, SSD] e [I, li]. Estas seis configuraciones concentran el 54.15% de las apariciones de marcadores (Pons, Pardo y Alemany, 2023).
- j) Las posiciones más frecuentes en la combinación de marcadores discursivos (registro formal) son [I,SA], [I, A], [I, li]. Estas tres casillas concentran aproximadamente el 90% de las combinaciones (Pons Bordería, 2018b).

A estos resultados teóricos se añaden los desarrollos prácticos, que son técnicas y procedimientos implementados como resultado del trabajo sobre corpus y del procesamiento de las conversaciones. Entre estos, figura el desarrollo de un procedimiento estandarizado para el procesamiento de las conversaciones (4.1) (Pons Bordería, 2022).

Por su parte, la transcripción genera una hoja de ELAN muy compleja en la que el habla de cada participante se descompone en una serie de líneas (*tiers*) que ofrecen información desagregada sobre su producción lingüística: en el caso de que la transcripción no incluya el paralinguaje, comprende quince capas de información (ocho para el estilo principal, seis para el estilo directo y una línea de observaciones –Badía, 2024–); en el caso de que se añada la información paralingüística, se añaden siete más (una para el tipo de gesto, dos para su orientación y cuatro para las partes del cuerpo implicadas –Cabanés, 2023–). El número total de líneas se multiplica por el número de hablantes.

Además del ELAN, para el tratamiento de las figuras conversacionales (Espinosa, 2024) se ha desarrollado un programa informático, denominado *Conversations*, que sirve para dibujar automáticamente dichas figuras y para contar el número y tipo de intervenciones que las componen (Badía y Espinosa, 2024).

4.6. Hipótesis de alto nivel sobre el español coloquial

Del análisis descrito arriba se derivan las siguientes hipótesis de alto nivel:

- a) La conversación coloquial, como género prototípico del registro no marcado, se caracteriza por ser el ámbito en el que se puede encontrar una mayor diversidad, ya sea en el ámbito interactivo (presencia de figuras conversacionales –Espinosa, 2016, 2024–), en la variedad de configuraciones en los marcadores (Pons, Pardo y Alemany, 2023) y en la estructura compositiva de la unidad *acto* (Pons, Salameh y Badía, 2025).
- b) El modelo Val.Es.Co. puede utilizarse para estudiar otros géneros discursivos orales, como los videojuegos en línea (Badía, 2024).
- c) El modelo Val.Es.Co. puede utilizarse para ofrecer una explicación sencilla del funcionamiento de los marcadores del discurso en el campo de E/LE (Pardo, 2020, 2021; Salameh 2021).
- d) El discurso hablado se puede dividir en un número finito de casillas (entendidas como pares ordenados [posición * unidad] en las que se sitúan las funciones discursivas –Pons Bordería, 2008–):

FIGURA 6: universo de unidades*posiciones en el discurso

POSICIÓN	UNIDAD						
	SUBACTO	ACTO	INTERVENCIÓN		INTERCAMBIO	DIALOGO	DISCURSO
			REACTIVA	INICIATIVA			
INICIAL	[I, SA]	[I, A]	[I, Ir]	[I, Ii]	[I, Int]	[I, Di]	[I, Disc]
INICIAL RELATIVA	[Irel, SA]	[Irel, A]	[Irel, Ir]	[Irel, Ii]	[Irel, Int]	[Irel, Di]	[Irel, Disc]
MEDIA	[M, SA]	[M, A]	[M, Ir]	[M, Ii]	[M, Int]	[M, Di]	[M, Disc]
FINAL	[F, SA]	[F, SA]	[F, Ir]	[F, Ii]	[F, Int]	[F, Di]	[F, Disc]
INDEPENDIENTE			[Ind, I]				

Estas rejillas ofrecen el universo de posibilidades que puede adoptar un elemento en la conversación, desde las posiciones de ámbito inferior, monologal, a las de ámbito superior, dialogal.

- e) La unidad en la que aparezca un marcador, la posición de dicho marcador dentro de cada unidad y su significado base determinan la función de una determinada ocurrencia de un marcador (*unidad*posición*sdo = función*). Esta hipótesis ha sido validada mediante el análisis de los 111 marcadores del DPDE

(análisis que se encuentra en el campo «Posición avanzada»), y cuyos resultados se han publicado en Pons, Pardo y Alemany (2023).

- f) Los elementos procedimentales poseen especializaciones funcionales en un reducido número de casillas. Esto significa que, aunque el universo de posibilidades teóricas del modelo sea de treinta y cinco, en la práctica un marcador polifuncional se puede explicar a partir de siete, ocho o nueve casillas, lo que permite una descripción muy económica (Pons, Pardo y Alemany, 2023).
- g) En el plano interactivo, las intervenciones se agrupan en un número limitado de *figuras conversacionales*, que reflejan patrones interactivos (Briz, 2009; Espinosa, 2016). La conversación coloquial, como género no marcado de la oralidad, muestra la mayor variedad de figuras conversacionales, mientras que otros géneros orales (entrevista, debate, etc.) presentan una nómina más restrictiva (Espinosa, 2024).
- h) La escrituralidad puede segmentarse mediante una extensión del modelo de unidades en la que las unidades claramente interactivas (intervención e intercambio) se sustituyen por nuevas unidades de tipo estructural (conjunto de actos y movimiento) (Pons y Borreguero, 2024; Salameh y Pardo, 2024).

4.6'. Hipótesis de alto nivel sobre problemas generales

De las hipótesis de alto nivel se derivan las siguientes ideas relacionadas con problemas de Lingüística General:

- a) El análisis de los marcadores en términos de unidad*posición es extensible a otros idiomas (Pons Bordería y Fisher (2021) para *well*, Robles (2023) para *doch*).
- b) La hipótesis de la gramaticalización de marcadores en términos de periferias derecha e izquierda (Beeching, Degand *et al.*, 2009) es incorrecta (Salameh, Estellés y Pons Bordería, 2018). En su lugar,
- c) no existe un único proceso de gramaticalización; existen diversas gramaticalizaciones, resultado de los cambios en la rejilla de unidades*posición sufridos por un marcador a lo largo de la historia y que pueden desarrollarse en distintas direcciones (Pons Bordería, 2018).
- d) La gramática del español hablado no es una gramática tética (Kaltenböck, Heine, Kuteva, 2011) que se superpone a la organización oracional y la recubre; la gramática oracional es una especialización, compuesta por estructuras sintácticas contrastables mediante métodos probatorios, a partir de una base común, rastreable en el discurso, cuyo desarrollo no se entiende sin el concurso de las diferentes tradiciones discursivas y teorías filosóficas que han ido conformando el pensamiento gramatical occidental. La sintaxis es, a la vez, un desarrollo natural y cultural.

5. CONCLUSIÓN

El establecimiento de una serie de requisitos exigibles a toda teoría que intente responder a la pregunta de Narbona sobre la estructura del español coloquial permite proponer un estándar con el que abordar la tercera fase de estudios que hemos delineado en este trabajo. Las diferentes propuestas teóricas sobre la organización del español coloquial que, de forma competitiva, aspiren a responderla se podrán contrastar de forma objetiva, ya se pueden medir contra un mismo conjunto de parámetros en todos los casos.

En el caso del modelo del grupo Val.Es.Co., puede decirse que el trabajo realizado para la creación de la versión actual de su corpus, que incluye la segmentación sin residuo de un subcorpus de conversaciones, y las investigaciones desarrolladas a partir del mismo, permiten superar los requisitos de entrada a la tercera fase en el estudio del español coloquial; una fase que se puede caracterizar como *de análisis de estructuras y contraste de hipótesis*. Estos dos objetivos, además, se pueden desarrollar desde una doble perspectiva: cualitativa (en el plano teórico) y cuantitativa (sobre una base empírica). Los estudios sobre español coloquial siguen adelante.

REFERENCIAS BIBLIOGRÁFICAS

- Albelda Marco, Marta (2008). Influence of situational factors on the codification and interpretation of impoliteness. *Pragmatics*, 18(4), 751-773.
- Albelda, M. (2016). Estableciendo límites entre la evidencialidad y la atenuación en español. En R. González, D. Izquierdo y O. Loureda (Eds.), *La evidencialidad en español. Teoría y descripción* (pp. 75-102). Iberoamericana/Vervuert.
- Badia, Sara (2024). *Comunicación en la arena digital: un análisis de la interacción en los videojuegos MOBA*. Peter Lang.
- Badia, Sara y Espinosa, Guadalupe (2024). Conversations: un software para el análisis semiautomático de la estructura conversacional. *Revista de Lingüística Aplicada*, 62(1).
- Beinhauer, Werner (1929/1963). *El español coloquial*. Gredos.
- Benavent Payá, Elisa (2023). *Decir y discurso directo en los relatos de la conversación coloquial*. Peter Lang.
- Borreguero, Margarita, Athayan, Vahran y Grosse, Sybille (Eds.). (2018). *Models of Discourse Units in Romance Languages. Revue Romane*, 53(1).
- Bravo, Diana y Briz, Antonio (Eds.). (2004). *Pragmática sociocultural: estudios sobre el discurso de cortesía en español*. Ariel.
- Briz, Antonio (1998). *El español coloquial en la conversación. Esbozo de pragmagramática*. Ariel.
- Briz, Antonio et al. (1995). *La conversación coloquial. Materiales para su estudio* (Vol. xv). Universidad de Valencia.
- Briz, Antonio, Pons, Salvador y Portolés, José (Eds.). (2008). *Diccionario de partículas discursivas del español*. Recuperado de www.dpde.es

- Briz Gómez, Antonio y Albelda Marco, Marta (2013). Una propuesta teórica y metodológica para el análisis de la atenuación en español y portugués. La base de un proyecto en común (ES:VAR.ATENUACIÓN). *Onomázein*, 28, 288-319.
- Briz, Antonio y Grupo Val.Es.Co. (2003). Un sistema de unidades para el estudio del español coloquial. *Oralia*, 6, 7-63.
- Cabanes P., Sandra (2023). *Análisis multimodal en la distinción entre intervención y turno: efectos en la segmentación de la conversación desde el modelo Val.Es.Co.* [Tesis doctoral, Universidad de Valencia].
- Calero Vaquera, M. L. (1986). *Historia de la gramática española (1847-1920)*. Gredos. Corpus Val.Es.Co. 3.0.
- Cresti, Emanuela (2000). *Corpus di italiano parlato*. Accademia della Crusca.
- Cortés, Luis y Camacho, M.ª Matilde (2005). *Unidades de segmentación y marcadores del discurso*. Arco Libros.
- Criado de Val, Manuel (1966). Transcripciones coloquiales. *Yelmo*, 5-9.
- Crible, Ludivine y Degand, Liesbeth (2019). Domains and functions: A two-dimensional account of discourse markers. *Discours. Revue de linguistique, psycholinguistique et informatique*, 24.
- De Latte, Fien (2024). *El vocativo en el español coloquial actual: variación pragmática, socio-indexical e interindividual* [Tesis doctoral, Universidad de Gante].
- Degand, Liesbeth y Simon, Anne-Catherine (2009). Minimal discourse units in spoken French: On the role of syntactic and prosodic units in discourse segmentation. *Discourse*, 4. Recuperado de <http://discours.revues.org/5852>
- Espinosa Guerri, Guadalupe (2016). Hacia una visualización sistemática de la conversación coloquial. En *Nuevas investigaciones lingüísticas: XXX Congreso internacional de la Asociación de Jóvenes Lingüistas*. Universidad de Cantabria.
- Espinosa Guerri, Guadalupe (2024). *Estructura visual de la conversación coloquial española* [Tesis doctoral, Universidad de Valencia].
- Ferrari, Angela, Cignetti, Luca, de Cesare, Anna-Maria, Lala, Letizia, Mandelli, Magda, Ricci, Claudia y Roggia, Enrico (2008). *L'interfaccia lingua-testo. Natura e funzioni dell'articolazione informativa dell'enunciato*. Edizioni dell'Orso.
- Ferrari, Angela y Borreguero Zuloaga, Margarita (2015). *La interfaz lengua-texto. Un modelo de estructura informativa*. Biblioteca Nueva.
- Fraser, Bruce (1990). An approach to discourse markers. *Journal of Pragmatics*, 14(3), 383-398.
- Fraser, Bruce (1999). What are discourse markers? *Journal of Pragmatics*, 31, 931-952.
- Fuentes Rodríguez, Catalina (2009). *Diccionario de conectores y operadores del español*. Arco Libros.
- Fuentes Rodríguez, Catalina (2019). Categorías discursivas y segmentación en macrosintaxis. En C. Fuentes Rodríguez y S. Gutiérrez Ordóñez (Eds.), *Avances en macrosintaxis* (pp. 15-65). Arco Libros.
- Garrido Medina, Joaquín y Rodríguez Ramalle, Teresa (2015). Constituyentes y relaciones en la oración y en el discurso. *CLAC*, 62, 199-225.
- Gómez Asencio, J. J. (1981). *Gramática y categorías verbales en la tradición española 1771-1847*. Universidad de Salamanca.
- Gómez Asencio, J. J. (1985). *Subclases de palabras en la tradición española: 1771-1847*. Universidad de Salamanca.
- Gries, Stefan (2009). *Statistics for Linguistics with R: A Practical Introduction*. Routledge.
- Gutiérrez Ordóñez, Salvador (1997-1998). La interdependencia en sintaxis. *Cauce*, 20-21, 703-725.
- Hernando Cuadrado, L. A. (1988). *El español coloquial en El Jarama*. Playor.

- Hjelmslev, L. (1943/1984). *Prolegómenos a una teoría del lenguaje*. Gredos.
- Kabatek, Johannes (Ed.). (2008). *Sintaxis histórica del español y cambio lingüístico: nuevas perspectivas desde las tradiciones discursivas*. Vervuert.
- Kaltenböck, Gunther, Heine, Bernd y Kuteva, Tania (2011). On thetical grammar. *Studies in Language*, 35(4), 848-893.
- Lope Blanch, Juan M. (1971). *El habla de la ciudad de México. Materiales para su estudio*. Centro de Lingüística Hispánica.
- Loureda, Óscar et al. (2021). *Comunicación, partículas discursivas y pragmática experimental*. Arco Libros.
- Marqu ez-Reiter, Rosina (2022). Mobility and stasis: Migrant portraits from a Madrid market. En Rosina Marqu ez-Reiter y Adriana Pati o-Santos (Eds.), *Language practices and processes among Latin Americans in Europe* (pp. 25-49). Routledge.
- Mart n Zorraquino, Mar a Antonia y Portol s, Jos  (1999). Los marcadores del discurso. En Ignacio Bosque y V. Demonte (Eds.), *NGLE*. Espasa-Calpe.
- Narbona, Antonio (1989a). *Las subordinadas adverbiales impropias en espa ol (I)*.  gora.
- Narbona Jim nez, Antonio (1989b). *Sintaxis espa ola: nuevos y viejos enfoques*. Ariel.
- Narbona, Antonio (1990). *Las subordinadas adverbiales impropias en espa ol (II)*.  gora.
- Pardo Llibrer, Adri  (2020). La polifuncionalidad de los marcadores discursivos en E/LE seg n unidad y posici n. *Foro de profesores de E/LE*, 16, 275-286.
- Pardo Llibrer, Adri  (2021). El problema de la ense anza de los marcadores discursivos y su visualizaci n: sobre *bueno* en la conversaci n coloquial espa ola. *Foro de profesores de E/LE*, 17, 25-40.
- Pascual Aliaga, Elena (2020). *Los truncamientos en la conversaci n coloquial: estudio de las huellas de formulaci n discursiva desde un modelo de unidades de lo oral* (Tesis doctoral). Universidad de Valencia.
- Pons Border a, Salvador (2008). La combinaci n de marcadores del discurso en la conversaci n coloquial: Interacciones entre posici n y funci n. *Estudios ling sticos*, 2, 141-160.
- Pons Border a, Salvador (2018). Paths of grammaticalization: Beyond the LP/RP debate. En S. Pons Border a y  . Loureda Lamas (Eds.), *Beyond grammaticalization and discourse markers: New issues in the study of language change* (pp. 334-383). Brill.
- Pons Border a, Salvador (2018). The combination of discourse markers in spontaneous conversations: Keys to untying a Gordian knot. *Revue Romane*, 121-158.
- Pons Border a, Salvador (2022). *Creaci n y an lisis de corpus orales. Saberes pr cticos y reflexiones te ricas*. Peter Lang.
- Pons Border a, Salvador (Ed.). (2014). *Discourse segmentation in Romance languages*. John Benjamins.
- Pons Border a, Salvador y Estell s Arguedas, Mar a (2008). Do digressive markers exist? *Journal of Pragmatics*, 41(5), 921-936.
- Pons Border a, Salvador y Fischer, Kerstin (2021). Using discourse segmentation to account for the polyfunctionality of discourse markers: The case of *well*. *Journal of Pragmatics*, 173, 101-118.
- Pons Border a, Salvador y Pascual, Elena (2020). Inter-annotator agreement in spoken language annotation: Applying *uc*-family coefficients to discourse segmentation. *Russian Journal of Linguistics*, 478-506.
- Pons Border a, Salvador y Borreguero, Margarita (2024). Unidades discursivas del texto escrito: Revisi n cr tica del estado de la cuesti n y directrices para una nueva propuesta. En S. Salameh Jim nez y A. Pardo Llibrer (Eds.), *Del p rrafo al texto: Nuevos enfoques en torno al an lisis de las unidades superiores del discurso escrito* (CLAC 99, pp. 7-21).

- Pons Bordería, Salvador, Pardo Llibrer, Adrià y Alemany, Amparo (2023). *La marcación discursiva en español: descripción y análisis conjunto desde el DPDE*. Editorial Universidad de Sevilla.
- Pons Bordería, Salvador, Salameh Jiménez, Shima y Badía, Sara (2025). *El español coloquial, en cifras: estudio cuantitativo de la estructura de lenguaje hablado*.
- Pons Rodríguez, Lola (2010). Los marcadores del discurso en la historia del español. En Óscar Loureda Lamas y Esperanza Acín Villa (Coords.), *Los estudios sobre marcadores del discurso en español, hoy* (pp. 523-616). Arco/Libros.
- Portolés, José (1998). *Marcadores del discurso*. Ariel.
- Real Academia Española. (1771). *Gramática de la lengua castellana*. Imprenta de Joaquín de Ibarra.
- Rizzi, Luigi y Haegeman, Lilian (1997). The fine structure of the left periphery. En *Elements of Grammar* (pp. 281-337).
- Robles Sabater, Ferrán y Carbonell Saiz, María (2024). La polifuncionalidad de *doch* en los diálogos literarios: aproximación estructural-informativa mediante un modelo de unidades discursivas. *CLAC 99*, 79-92.
- Roulet, Eddy et al. (1985). *L'articulation du discours en français contemporain*. Peter Lang.
- Roulet, Eddy et al. (2001). Un modèle et un instrument d'analyse de l'organisation du discours. Peter Lang.
- Salameh Jiménez, Shima (2021). Más allá de los huecos. El tratamiento de los marcadores discursivos en español peninsular mediante las nociones de ámbito y posición. *Foro de profesores de E/LE, 17*, 63-84.
- Salameh Jiménez, Shima (2021). *Reframing reformulation: A theoretical-experimental approach: Evidence from the Spanish discourse marker o sea*. Peter Lang.
- Salameh Jiménez, Shima y Pardo Llibrer, Adrià (Eds.). (2024). *Del párrafo al texto: nuevos enfoques en torno al análisis de las unidades superiores del discurso escrito* (CLAC 99).
- Salameh Jiménez, Shima y Pardo Llibrer, Adrià (2024). Más allá del párrafo en la segmentación de lo escrito: Hacia una caracterización de las unidades conjunto de actos {AA} y movimiento {M}. En Shima Salameh Jiménez y Adrià Pardo Llibrer (Eds.), *Del párrafo al texto: Nuevos enfoques en torno al análisis de las unidades superiores del discurso escrito* (CLAC 99, pp. 27-37).
- Salameh Jiménez, Shima, Estellés, María y Pons Bordería, Salvador (2018). Beyond the notion of periphery: An account of polyfunctional discourse markers within the Val.Es.Co. model of discourse segmentation. En Kate Beeching, Chiara Ghezzi y Piera Molinelli (Eds.), *Positioning the Self and Others: Linguistic perspectives* (pp. 105-125).
- Santos Río, Luis (2003). *Diccionario de partículas*. Luso-Española Ediciones.
- Seco, Rafael (1970). *Arniches y el habla de Madrid*. Alfaguara.
- Sinclair, John (1996). Preliminary recommendations on corpus typology. Recuperado de <http://www.ilc.cnr.it/EAGLES/corpusyp/corpusyp.html>
- Sperber, Dan y Wilson, Deirdre (1986). *Relevance: Communication and cognition*. Harvard University Press.
- Vigara Tauste, Ana María (1980). *Aspectos del español hablado*. SGEL.
- Vigara Tauste, Ana María (1992). *Morfosintaxis del español coloquial: Esbozo estilístico*. Gredos.
- Yus, Francisco (2020). La comunicación en la era digital. En M.^a Victoria Escandell Vidal et al. (Eds.), *Pragmática* (pp. 608-623). Akal.