

# Descripción y análisis de un corpus de español oral de la comunidad de habla LGTBI

CARLES NAVARRO-CARRASCOSA  
*Universidad de Valladolid*  
carles.navarro@uva.es

---

**Resumen:** Los corpus orales son herramientas esenciales para la investigación lingüística, ya que proporcionan datos auténticos del uso del lenguaje en contextos reales. En este trabajo se presenta el Corpus Oral de la Comunidad de Habla LGTBI (Navarro-Carrascosa 2023a), que recopila muestras de habla de esta comunidad, facilitando el estudio de sus códigos lingüísticos y prácticas comunicativas. Para mejorar este corpus se propone incluir muestras de habla de miembros del colectivo que no fueron acopiadas en los primeros registros, así como ampliar las muestras a más ciudades y áreas rurales, además de las variantes hispanohablantes de otros países para enriquecer el corpus con una mayor diversidad lingüística y cultural. Estas mejoras permitirán que el corpus beneficie investigaciones detalladas sobre variaciones lingüísticas, influencias socioculturales y la evolución del lenguaje en la comunidad LGTBI. Con una expansión y desarrollo continuos, el corpus seguirá proporcionando datos valiosos para una comprensión más completa y diversa del lenguaje dentro de esta comunidad.

**Palabras clave:** lingüística de corpus, lingüística *queer*, corpus orales, comunidad de habla LGTBI, sociolingüística.

## Description and analysis of an oral Spanish corpus of the LGTBI speaking community

**Abstract:** Oral corpora are essential tools for linguistic research, as they provide authentic data on language use in real contexts. This work presents the Oral Corpus of the LGTBI Speaking Community (Navarro-Carrascosa 2023a), which collects speech samples from this community, facilitating the study of its linguistic codes and communicative practices. To improve this corpus, it is proposed to include samples from community members that were not collected in the initial records, as well as to expand the recordings to more cities and rural areas, and to incorporate Spanish-speaking variants from other

countries to enrich the corpus with greater linguistic and cultural diversity. These improvements will allow the corpus to benefit detailed investigations into linguistic variations, sociocultural influences, and the evolution of language in the LGBTBI community. With continuous expansion and development, the corpus will continue to provide valuable data for a more comprehensive and diverse understanding of language within this community.

**Keywords:** corpus linguistics, queer linguistics, oral corpora, LGBTBI speaking community, sociolinguistics.

## 1. Introducción

Los corpus orales son herramientas indispensables para la investigación lingüística, ya que proporcionan datos auténticos del uso del lenguaje en contextos reales. Estas colecciones estructuradas de grabaciones de habla permiten a los investigadores analizar una amplia gama de fenómenos lingüísticos, desde variaciones dialectales hasta la pragmática y la sociolingüística. La importancia de los corpus orales radica en su capacidad para capturar la riqueza y diversidad del habla, ofreciendo una base empírica sólida para estudios detallados y precisos.

En este contexto, se presenta el Corpus Oral de la Comunidad de Habla LGBTBI (Navarro-Carrascosa 2023a), un recurso pionero que recopila muestras de habla reales de esta comunidad. Este corpus ofrece numerosos beneficios, como la posibilidad de estudiar los códigos lingüísticos propios del colectivo LGBTBI y las intenciones pragmalingüísticas en la comunicación interna entre sus miembros. Sin embargo, también enfrenta ciertos desafíos, especialmente en términos de la representación de subgrupos menos visibilizados y la obtención de grabaciones en modalidad secreta para captar el habla más espontánea posible.

Este trabajo no solo describe y explica el corpus actual, sino que también indica pautas para su ampliación y desarrollo futuros. Se explorarán estrategias para incluir a más personas transexuales binarias, intergénero, bisexuales y de otras categorías no registradas hasta ahora, como asexuales y demisexuales. Además, se discutirán métodos para registrar más grabaciones con personas de mayor edad, permitiendo así estudios diacrónicos sobre la evolución del argot a través de generaciones y la expansión geográfica del corpus a más ciudades y áreas rurales, así como a otras variantes hispanohablantes en países de habla española.

Finalmente, se presentarán las perspectivas de futuro del corpus y las posibles investigaciones que podría beneficiar. Desde estudios fonético-fonológicos y léxicos hasta análisis morfológicos y pragmáticos, este corpus tiene el potencial de enriquecer significativamente la investigación lingüística. Su desarrollo y expansión continuarán proporcionando datos valiosos para una comprensión más completa y diversa del lenguaje dentro de la comunidad de habla LGTBI.

## 2. Los corpus orales del español

Los corpus orales del español son colecciones estructuradas de grabaciones de habla en lengua española que se utilizan para el estudio y análisis lingüístico. Estos corpus contienen datos orales recolectados en diversas situaciones comunicativas y contextos, incluyendo conversaciones espontáneas, discursos formales, entrevistas, debates, y otros tipos de interacción verbal. El objetivo principal de estos corpus es proporcionar material auténtico y representativo del uso del idioma en situaciones reales.

### 2.1. *Corpus orales generales*

Los corpus orales generales son aquellos que recogen muestras de habla de los miembros pertenecientes a una sociedad. Este tipo de corpus busca capturar una representación amplia y diversa del uso del lenguaje, incluyendo a personas de diferentes edades, géneros, niveles educativos, ocupaciones y orígenes geográficos. Un ejemplo destacado de corpus general es el de conversaciones coloquiales del grupo Val. Es.Co.<sup>1</sup>, el cual, aunque tiene un objeto de estudio muy delimitado (el español coloquial), no discrimina a informantes, ya que cualquier miembro de una sociedad es susceptible de utilizar un registro familiar en determinados contextos.

Estos corpus son fundamentales para el estudio sociolingüístico, pues permiten analizar cómo se manifiestan las variaciones lingüísticas en distintas capas de la sociedad y cómo se utilizan diferentes registros del habla en contextos variados. Además, estos corpus son útiles para investigar fenómenos como la evolución del lenguaje, la influencia de factores sociales en el habla y la dinámica del cambio lingüístico.

Otros ejemplos son el corpus PRESEEA<sup>2</sup> (Proyecto para el Estudio Sociolingüístico del Español de España y de América), que recopila datos orales de diferentes ciudades hispanohablantes; y el CORLEC<sup>3</sup>

<sup>1</sup> [https://www.valesco.es/#/pages/cod\\_hj3y7hwvuuajtllkq0ik/cod\\_m5em4hajg5ljtpxzuc1](https://www.valesco.es/#/pages/cod_hj3y7hwvuuajtllkq0ik/cod_m5em4hajg5ljtpxzuc1).

<sup>2</sup> <https://preseea.uah.es/corpus-preseea>.

<sup>3</sup> <http://www.lllf.uam.es/ESP/Corlec.html>.

(Corpus Oral de Referencia de la Lengua Española Contemporánea), que se centra en la recopilación de muestras de habla de diversas regiones y contextos.

## 2.2. *Corpus orales parciales o sectoriales*

Son aquellos que se limitan a un sector social específico, como determinados grupos sociales, profesionales o regionales. Estos corpus están diseñados para estudiar y describir la variedad del lenguaje en contextos más acotados y concretos, proporcionando información detallada sobre el uso del lenguaje en comunidades particulares.

Al tener los corpus parciales objetivos mucho más definidos que los corpus generales, los investigadores pueden realizar análisis más detallados y específicos, facilitando el estudio de fenómenos lingüísticos particulares dentro de esos grupos.

Algunos ejemplos de corpus parciales son el COVJA (Corpus Oral de la Variedad Juvenil Universitaria del Español de Alicante) (Azorín Fernández 2005); otro ejemplo más reciente es el corpus del proyecto ESPRINT<sup>4</sup>, que se enfoca en las interacciones de parejas; el COSER<sup>5</sup> (Corpus Oral y Sonoro del Español Rural), que proporciona muestras de las variantes dialectales rurales del español de España, en concreto, de personas mayores; el CHCS (Corpus del Habla Culta de Salamanca), de Fernández Juncal (2005); el COJEM<sup>6</sup> (Corpus Oral Juvenil del Español de Mallorca); o el COLEM (Corpus Oral de la Lengua Española en Montreal), de Pato (2020).

## 3. Fases para la elaboración de un corpus oral parcial de una comunidad lingüística específica

Moreno Fernández (2005) enumera los fundamentos metodológicos que él y su equipo llevaron a cabo para la elaboración de los distintos corpus orales del proyecto PRESEEA. Nosotros los hemos sintetizado en cuatro procedimientos metodológicos que deben seguirse para la construcción de un corpus oral: la elección de informantes, el tamaño de la muestra, el método de recolección y la transcripción. Todos estos criterios responden a decisiones que los investigadores tendrán que afrontar para desarrollar su tarea. Sin embargo, todos los puntos son mucho más delicados cuando el objetivo es la elaboración de un realia oral de una comunidad de habla específica.

<sup>4</sup> Estrategias pragmático-retóricas en la interacción conversacional conflictiva entre íntimos y conocidos: intensificación, atenuación y gestión interaccional. Ministerio de Ciencia e Innovación (PID2020-114805GB-I00). Directoras: Marta Albelda / Maria Estellés (Universitat de València).

<sup>5</sup> <http://www.corpusrural.es>.

<sup>6</sup> [http://www.linred.es/numero13\\_corpus-1.html](http://www.linred.es/numero13_corpus-1.html).

### 3.1. *La elección de informantes*

Sobre este aspecto, Briz (2012) ya señalaba la falta de representatividad de todos los corpus orales, una carencia que se acentúa aún más en aquellos de carácter parcial, pues en ellos se pone el foco en un sector de la sociedad más acotado. Para definir el tipo de informante necesario para un corpus oral sectorial es importante plantear algunas cuestiones específicas. Generalmente, se trata de criterios demandados por los objetivos de las investigaciones para las que se acopian los textos orales: «núcleos urbanos, monolingües o bilingües, con una población hispanohablante bien asentada, con conciencia de comunidad y suficientemente diversa desde un punto de vista sociológico» (Moreno Fernández 2005: 127). Las decisiones que deben tomarse, a este respecto, tienen que ver con criterios demográficos, como la edad, los años de residencia, el género, el nivel de instrucción, etc. (Carcelén y Uclés 2019).

Estos son los procedimientos más habituales, pero, dependiendo del tipo de investigación, podrían surgir otros más específicos o particulares que acotarán mucho más el perfil de los informantes. Por ejemplo, Sanmartín (1998, 1999a y 1999b), para su trabajo sobre el español de la delincuencia elaboró un corpus de habla cuyos informantes eran reclusos de la prisión de Valencia. Manejó variables generales y otras más específicas, como el tipo de delito que había cometido el informante o los años que llevaba recluido en el momento de la entrevista.

### 3.2. *El tamaño de la muestra*

El tamaño de la muestra es una de las cuestiones más polémicas entre los distintos investigadores en sociolingüística (Briz y Val. Es.Co. 2002a). Lo habitual es aplicar principios estadísticos para obtener muestras representativas de la población lingüística (Torrueala y Llisterri 1999). Esta cuestión va aparejada al número de participantes.

El plan de muestreo diseñado para una muestra aleatoria simple supone un nivel de confianza del 90 % en los resultados obtenidos, con un margen de error en esa confianza del 6 %. Este nivel de precisión exige una muestra de 189 informantes, valor que nos permitirá un alto grado de seguridad en las inferencias obtenidas (Briz y Grupo Val.Es.Co. 2002a: 14).

Para el corpus PRESEEA se estableció que el tamaño de la muestra debe estar relacionado con la población total del municipio de la recopilación:

El estudio de las comunidades con poblaciones de entre medio millón y un millón de habitantes se aborda mediante muestras formadas por 54 informantes. Cuando la ciudad es de un tamaño mayor y de una mayor complejidad sociológica, se recomienda el empleo de muestras de 72 a 108 hablantes (Moreno Fernández 2005: 128).

No obstante, este criterio puede variar atendiendo a varios factores, entre ellos el tipo de recolección (que se verá en el siguiente apartado). Los corpus vivos (aquellos cuya elaboración nunca finaliza, pues siempre van ampliándose, como es el caso del corpus de conversaciones coloquiales del grupo Val.Es.Co), al ser ilimitados y estar en constante elaboración, no siempre establecen un número de informantes.

En los corpus que buscan plasmar los modos de habla de comunidades específicas, se pueden presentar diversos problemas a la hora de establecer un criterio cuantitativo en el número de participantes. En primer lugar, si las comunidades son minoritarias, la dificultad puede girar en torno a encontrar informantes o a que estos deseen participar. Por ejemplo, el corpus de Montecino Soto (2008), limitado a personas sin hogar de Santiago de Chile, necesitó de más de dos años para conseguir once entrevistas.

### 3.3. *El método de recolección*

La codificación y presentación de datos orales implica decisiones sobre el formato de registro y las convenciones de transcripción (Pons y Ruiz Gurillo 2005; Recalde y Vázquez Rozas 2009). Los corpus orales suelen recurrir a dos técnicas de acopio de muestras: la entrevista sociolingüística y la grabación secreta.

#### 3.3.1. *La entrevista sociolingüística*

La entrevista sociolingüística es una técnica utilizada por lingüistas investigadores para obtener muestras de habla reales en determinados contextos sociales. Es una metodología original de William Labov, diseñada en los años 60 y con el objetivo de estudiar la variación dialectal y el cambio lingüístico a través del registro y el análisis de las muestras acopiadas.

Existen varias técnicas metodológicas empleadas para desarrollar una entrevista sociolingüística; algunas de ellas están diseñadas para minimizar la paradoja del observador<sup>7</sup> (Labov 1972). Estas técnicas se clasifican de la siguiente manera:

---

<sup>7</sup> La paradoja del observador se refiere al fenómeno donde los sujetos modifican su comportamiento cuando saben que están siendo observados.

– **La entrevista estructurada:** está basada en un conjunto de preguntas preparadas con anterioridad (en ocasiones, acordadas), lo que facilita la comparación de respuestas entre los distintos entrevistados. No obstante, varios autores destacan que las respuestas de estas entrevistas suelen resultar menos espontáneas y, por tanto, su validez es menor (Calderón Noguera y Alvarado Castellanos 2011).

– **La entrevista semiestructurada o semidirigida:** el entrevistador sigue un guion básico, pero puede adaptar las preguntas según las respuestas del informante, fomentando así una conversación más natural y detallada. «Es una estrategia de recolección de materiales de punto medio, en la que las preguntas planificadas con anterioridad tienen como propósito buscar registros de actuación lingüística [...] o datos lingüísticos de manera directa» (Calderón Noguera y Alvarado Castellanos 2011). Según Moreno Fernández (2005: 128), las entrevistas sociolingüísticas «suelen recogerse mediante conversaciones semidirigidas y grabadas con magnetófono [o cualquier dispositivo digital habilitado] a la vista en situación de entrevista».

– **La entrevista no estructurada:** también llamada entrevista libre (Fernández Sanmartín 2022). Su principal característica es la total flexibilidad y espontaneidad, lo que permite una conversación más natural. Es útil para registrar el habla vernácula, aunque dificulta un análisis sistemático.

Tanto en la entrevista semidirigida como en la no estructurada, se pueden utilizar diferentes estrategias con el objetivo de mejorar la naturalidad de las respuestas, como la inclusión de múltiples entrevistadores (Calderón Noguera y Alvarado Castellanos 2011) o la realización de entrevistas grupales con más de un informante a la vez donde los participantes se sientan más cómodos y menos observados (Moreno Fernández 2011; Navarro-Carrascosa 2023a).

### 3.3.2. *La grabación secreta*

La grabación secreta es una técnica de recopilación de muestras de habla reales que consiste en registrar conversaciones sin que los participantes sepan que están siendo grabados. El objetivo es que las manifestaciones lingüísticas sean lo más espontáneas posible, evitando probables moderaciones expresivas de los hablantes al ser conscientes del registro.

La grabación secreta puede ser con observación participante (el investigador que consigue la muestra participa en ella) o sin esta. Ambas modalidades son cada vez más habituales en la investigación sociolingüística por considerarse «la forma más eficaz de obtención de datos

del español coloquial y permite soslayar inconvenientes teóricos como la llamada *paradoja del entrevistador*» (Briz y Val.Es.Co. 2002a: 17).

Con la legislación anterior se permitía que el investigador informara debidamente, una vez finalizado el registro, a los participantes de que la grabación había tenido lugar y les solicitara, *a posteriori*, un consentimiento por escrito firmado para que el documento resultante fuera utilizado. En la actualidad, el consentimiento debe otorgarse siempre antes del registro, lo cual dificulta más la espontaneidad de las muestras, especialmente con miembros de comunidades específicas, cuyos informantes tienen características más concretas y menos habituales. Además, cualquier dato que permita la identificación de cualquiera de los informantes deberá ser falseado.

### 3.4. La transcripción

La transcripción de textos orales con fines lingüísticos implica la representación escrita de lo dicho, pero también la preservación de elementos paralingüísticos y contextuales esenciales que permitirán un análisis exhaustivo. Cabe añadir que este proceso no es una mera traducción entre modos de expresión; es, además, un acto forzosamente interpretativo, en el que quien transcribe incorpora evaluaciones y toma decisiones sobre cómo representar cada detalle del habla. Algunos autores consideran que es un método fundamental para el análisis lingüístico actual, ya que facilita el manejo de datos mediante *software* especializado para la visualización, etiquetado morfosintáctico, entre otros objetivos (Ridao Rodrigo 2021).

Podemos enumerar tres tipos de transcripción: la ortográfica, la fonética y fonológica y la anotada:

– La transcripción ortográfica plasma la grabación oral usando los recursos ortográficos estándar. Se trata de trasladar al papel el texto oral (Bejarano, Llanos, Rubio y Bonilla 2018).

– La transcripción fonética y fonológica procura capturar con mayor precisión los sonidos del habla, utilizando una serie de símbolos para representar detalles prosódicos, entre otros.

– La transcripción anotada incluye anotaciones adicionales sobre aspectos paralingüísticos y contextuales, como elementos no verbales que acompañan al discurso (Ridao Rodrigo 2021).

Para facilitar la transcripción y garantizar la uniformidad se han desarrollado diversos protocolos. Por ejemplo, el Instituto Caro y Cuervo ha creado el protocolo de transcripción ortográfica CLICC (Bejarano, Llanos, Rubio y Bonilla 2018), que establece pautas claras

para transcribir corpus orales. Sin embargo, los métodos más conocidos y utilizados en sociolingüística son el sistema PRESEEA y Val. Es.Co., que aplican sistemas de transcripción de tipo anotada.

El sistema de Briz y el grupo Val.Es.Co (2002b), uno de los más populares en la actualidad, se basa en tres principios fundamentales: transcripción ortográfica adaptada, es decir, respecta la normativa ortográfica e incluye símbolos convencionales, como signos de exclamación o interrogación; incorporación de elementos paralingüísticos, como risas, vacilaciones, reformulaciones, cambios de tono, etc. Estos se marcan como etiquetas específicas dentro del texto para que el lector pueda interpretarlas de manera clara. Por último, el uso de etiquetas para señalar aspectos relevantes de la interacción, como las interrupciones, los solapamientos y las palabras ininteligibles, entre otros.

#### 4. Necesidad de un corpus oral de la comunidad de habla LGTBI

Desde hace algunas décadas, la academia estadounidense ha abierto un espacio para los estudios *queer*, es decir, aquellos estudios pertenecientes a las distintas áreas de conocimiento que desarrollan sus investigaciones desde una perspectiva de las identidades de género y sexoafectivas disidentes. En los últimos años, esta tendencia ha empezado a surgir en el hispanismo y podemos encontrar varios trabajos académicos de diferentes disciplinas (arquitectura, literatura, educación, derechos, etc.) desde un prisma LGTBI. La lingüística *queer* no ha sido una excepción y está ganando un espacio significativo en los estudios hispanistas.

En este sentido, para poder ofrecer materiales de estudio para una lingüística *queer* se hace necesario el Corpus oral de la comunidad de habla LGTBI (Navarro-Carrascosa 2023a), cuyo objetivo es

ofrecer materiales para estudiar y analizar los códigos lingüísticos propios de esta subcultura y las intenciones pragmlingüísticas que se dan en la comunicación interna entre sus miembros y, de este modo, servir de herramienta en las investigaciones de la Lingüística *queer* hispánica (Navarro-Carrascosa 2023a: 11).

Cada colectivo alberga una serie de características actitudinales que las distinguen del resto. Entre ellas se pueden destacar comportamientos lingüísticos determinados. En el caso de la comunidad de habla LGTBI, estos se hacen especialmente particulares, ya que provienen de necesidades históricas de mantenerse en la clandestinidad e interactuar con otros miembros del colectivo. Este contexto represivo para sus miembros motivó una serie de mecanismos lingüísticos secretos

que hoy en día ya no son necesarios. Sin embargo, algunos de ellos se han mantenido y han evolucionado, ya descodificados para el resto de la sociedad gracias a publicaciones literarias y a los medios de comunicación (Rodríguez 2008).

Este argot propio del colectivo LGTBI no solo refleja los intereses y necesidades de esta comunidad, sino que también expresa roles, comportamientos sexuales específicos dentro de esta, reforzando la cohesión interna y creando una alternativa a la heterosexualidad hegemónica, tal y como apunta Bengoechea (2015).

Para poder clasificar, organizar y analizar todas estas estrategias lingüísticas se hizo necesaria la compilación de todas las muestras de habla reales que conforman el corpus y se mantiene hoy día la necesidad de seguir ampliándolo con más conversaciones y entrevistas que mantengan el carácter sincrónico del *realia* y amplíen la variedad de informantes en todas las variantes que un corpus de estas características debe manejar (la identidad de género y la orientación sexual, fundamentalmente).

## 5. El corpus oral de la comunidad de habla LGTBI

Como se señalaba en el apartado anterior, uno de los corpus orales parciales o sectoriales del español es el que recogen muestras de habla reales de la comunidad LGTBI<sup>8</sup>. El corpus consta de un total de cincuenta y nueve informantes distribuidos en dos tipos de muestras: conversaciones coloquiales y entrevistas semiestructuradas.

Las conversaciones coloquiales incluyen diecisiete informantes, organizados en cinco grabaciones: dos en Valencia (C.VA.1 con cuatro participantes y C.VA.2 con cinco) y tres en Madrid (C.MA.1 con dos, C.MA.2 con tres y C.MA.3 con tres). Estas conversaciones fueron obtenidas mediante grabación secreta para preservar la naturalidad de las interacciones, con una duración total de 282 minutos y 38 segundos.

Las entrevistas cuentan con cuarenta y dos informantes, realizadas en Barcelona, Madrid y Valencia. Se han llevado a cabo tres entrevistas en Barcelona con diez participantes; nueve en Madrid con diecinueve participantes; y ocho en Valencia con tres participantes. La duración acumulada de las entrevistas es de 1175 minutos y 43 segundos.

---

<sup>8</sup> Se entiende por comunidad LGTBI a aquel grupo social compuesto «por personas que representan realidades sexoafectivas y/o de género que no están integradas dentro de lo que socialmente es considerado como normal» (Navarro-Carrascosa 2023b: 45). Por tanto, se considerará que los miembros de esta comunidad serán hombres y mujeres cisgénero homosexuales, bisexuales, pansexuales, mujeres trans, hombres trans y personas no binarias.

Este corpus, iniciado por Navarro-Carrascosa (2023a), sigue una serie de pautas y criterios de elaboración atendiendo a las particularidades que el contexto de acopio requiere, los cuales se desarrollan a continuación para la posible ampliación.

### 5.1. Informantes

Para seleccionar con precisión los participantes en la recolección de datos de una investigación lingüística enfocada en un grupo social específico se debe identificar las variables sociolingüísticas pertinentes que definan a los miembros de la comunidad de estudio y que los diferencien del resto. En el caso del colectivo LGTBI estas variables se centran en la orientación sexual, la identidad de género y el círculo social del informante, determinando si las personas que lo conforman son miembros del colectivo o no. A continuación, se presentan las variables más importantes que resultan determinantes para la selección de informantes del corpus oral LGTBI<sup>9</sup>:

— La orientación sexual: las variables que se manejan en esta categoría son tres: la heterosexualidad (aquellos que se sienten atraídos hacia el género opuesto); la homosexualidad (la atracción sexual hacia el mismo género); y la bisexualidad<sup>10</sup> (correspondiente a personas que se sienten atraídas a otras personas, independientemente del género de estas).

— La identidad de género: no se corresponde la identidad de género con la biología de los genitales de cada individuo. La primera, se refiere a los roles de género que cada uno siente que le son propios. Nuevamente, destacamos tres variables: personas cisgénero (la identidad de género está alineada con lo socialmente atribuido a cada uno de los géneros; es decir, una persona con genitales masculinos se identifica con los roles de género socialmente asociados al hombre); personas transgénero, esto es, la identidad del informante no está determinada por sus genitales de nacimiento<sup>11</sup>. Por último, en cuanto a la identidad de género destacamos una última categoría para identificar a los informantes del corpus: se trata de las personas no binarias (NB) o intergénero, las cuales no se identifican identitariamente ni como hombres ni como mujeres. Llegados a este punto, nos planteamos la posibilidad de un cuarto grupo en cuanto al género: la intersexualidad<sup>12</sup>,

<sup>9</sup> Los términos referidos a identidades de género y orientaciones sexuales se definen en profundidad en obras lexicográficas parciales, como las de Mira (1999), Pereda (2004) y Rodríguez (2008).

<sup>10</sup> Los estudios *queer* consideran la ruptura entre la dicotomía de género masculino-femenino, lo que cuestionaría el término bisexualidad (si se asume que existen más de dos géneros) y haría necesario hablar de pansexualidad, la atracción hacia las personas independientemente de su género, entendiendo que hay más de dos.

<sup>11</sup> Dentro de este grupo de pueden distinguir mujeres transgénero (mujeres que nacieron con genitales masculinos) y hombres transgénero (hombres que nacieron con genitales masculinos).

<sup>12</sup> La intersexualidad es una condición biológica en la que una persona nace con características

es decir, personas que han nacido con genitales ambiguos o variaciones cromosómicas.

— El círculo social del informante: como ya se ha apuntado, es muy relevante para la elección de participantes de este corpus, pues es relevante que un hombre cisgénero heterosexual forme parte del inventario si su círculo social está conformado, en parte, por personas LGTBI, ya que probablemente esto afecte a su expresión en contextos de familiaridad; del mismo modo que un miembro del colectivo LGTBI que no se relacione con otros no es considerado un hablante representativo del argot de esta comunidad. Por tanto, las posibilidades son dos: informantes con contacto social con personas LGTBI y los informantes que no lo tienen.

— La edad: es una variable muy relevante en la mayoría de los estudios sociolingüísticos. Para las investigaciones en lingüística *queer* también lo es, pues resulta de interés observar si hay cambios en las formas de expresarse de los miembros de esta comunidad lingüística atendiendo a este factor. La división en las franjas de edad que se propone para catalogar a los informantes difiere de la clasificación clásica en sociolingüística (aplicada a estudios como los de PRESEEA 2021; entre otros), pues se parte de la base de que el contexto social de los miembros del colectivo LGTBI tiene particularidades socio-históricas relevantes y, por tanto, la clasificación de los grupos de edad que se ha establecido es la siguiente:

a) De 18 a 25 años: este grupo ha crecido en una sociedad en la que el colectivo LGTBI está totalmente visibilizado, han alcanzado varios derechos sociales y, por tanto, su contexto no ha experimentado tantas censuras como los grupos de mayor edad.

b) De 26 a 35 años: los participantes de esta franja de edad también han crecido en un contexto donde las personas del colectivo están más o menos visibilizadas, pero la imagen que se proyecta de ellas se relaciona con la burla. Además, algunos aspectos relacionados con este colectivo todavía se consideran un tabú (Mira 2004).

c) De 36 a 45 años: el contexto en el que han crecido las personas de esta franja de edad tiene un grado de aceptación hacia el colectivo que, aunque empieza a abrirse, todavía es reducido (Mira 2004).

d) Más de 45 años: se han criado durante la dictadura o durante la transición, es decir, en la época en la que la homosexualidad estaba penalizada. Por tanto, la imagen que se proyecta de los gays y las

---

sexuales físicas que no encajan en las típicas definiciones binarias de masculino o femenino. Estas características pueden incluir variaciones en los genitales, las gónadas (ovarios o testículos), los cromosomas sexuales (XX, XY o combinaciones atípicas), o los niveles hormonales.

lesbianas es muy negativa. Por otro lado, el resto del colectivo está totalmente invisibilizado (Mira 2004).

## 5.2. *Recopilación de datos*

La recopilación de datos en los estudios sociolingüísticos y dialectológicos sigue un conjunto de características generales que aseguran la calidad, ética y confidencialidad de la información obtenida. A continuación, se describen las características fundamentales para la recopilación de datos en este contexto:

— La **transcripción** de los datos: se realiza utilizando el sistema Val.Es.Co, desarrollado por Briz y el Grupo Val.Es.Co. (2002a). Este sistema de transcripción es detallado y permite capturar con precisión las características del habla, incluyendo pausas, entonación y otros fenómenos prosódicos.

— Para proteger la identidad de los participantes y cumplir con la legalidad vigente se lleva a cabo la **anonimización** de los datos. Esto implica el cambio de nombres y cualquier otra información que pudiera permitir el reconocimiento de los informantes.

— Los investigadores redactan un **compromiso de confidencialidad**, en el cual se comprometen a mantener la anonimización y a tratar las grabaciones de manera correcta y ética. Este documento debe ser archivado y conservado como parte de la documentación del estudio, asegurando así el cumplimiento de las normas éticas.

— Tanto para las entrevistas como para las grabaciones secretas de conversaciones coloquiales, se utiliza una **ficha identificatoria** basada en la metodología de Briz y el Grupo Val.Es.Co. (2002a). Esta ficha incluye información relevante del registro, como la fecha y lugar de la grabación, el investigador principal, el número de participantes, sus edades, profesiones y la duración del registro, entre otros datos. Esta información es esencial para contextualizar los datos y facilitar su análisis posterior. Dadas las particularidades de los participantes en las grabaciones del corpus oral de la comunidad de habla LGTBI, la ficha ha sido completada para que incluya rasgos como la identidad de género y la orientación sexual.

### 5.2.1. *Entrevistas*

En el caso de las entrevistas, para garantizar su clasificación e identificación es importante el sistema de etiquetado, que sigue las siguientes pautas:

— Cada entrevista se marca con una *E* al inicio del etiquetado para diferenciarlas de las grabaciones secretas.

— Se indica la ciudad en la que se ha llevado a cabo el registro mediante un código específico (*VA*, para Valencia; *BA* indica que el registro se ha hecho en Barcelona; y *MA* señala que la entrevista ha tenido lugar en Madrid). Para ciudades nuevas en las que todavía no se han recopilado muestras, se asigna una marca específica que procure no coincidir con otras ya asignadas. Por ejemplo: *SE*, para Sevilla; *VALL*, para Valladolid; o *MAL*, para Málaga.

— El etiquetado finaliza con el número asignado al documento, generalmente siguiendo el orden en el que fueron recopiladas: *E.VA.3*, por tanto, es la tercera entrevista obtenida en Valencia; mientras que *E.BA.2* se corresponde con la segunda grabada en Barcelona.

El sistema de etiquetado permite una rápida localización de los documentos (tanto de los audios como de las transcripciones) y un sencillo manejo de los datos, facilitando el análisis comparativo entre diferentes registros.

Para maximizar la naturalidad del discurso y mitigar la paradoja del observador (Labov 1972), es recomendable propiciar un ambiente cómodo para los entrevistados. Para ello la metodología de registro de entrevistas del Corpus Oral de la Comunidad de Habla LGTBI establece las siguientes directrices:

— El **lugar de la grabación** debe ser un espacio familiar para los entrevistados. Idealmente, el hogar de uno de los participantes o un comercio habitual para ellos, siempre de su elección. Esto permite que los entrevistados se sientan en un entorno seguro y conocido, lo que facilita una mayor naturalidad en sus intervenciones.

— El **número de participantes** por entrevista se recomienda que no sea inferior a dos. Además, la **relación** entre ellos debe ser afectiva (familiares, amigos, pareja). Esta vinculación emocional promueve un ambiente relajado y permite que los participantes se sientan más cómodos al expresarse.

— El entrevistador lleva un esquema con temas específicos para la entrevista, enfocándose en asuntos relativos a la comunidad LGTBI. Sin embargo, este esquema no condiciona el desarrollo de la sesión y se permite que los participantes introduzcan otros temas si así lo desean. Este enfoque flexible facilita una conversación más espontánea y fluida, ya que los entrevistados pueden hablar sobre asuntos que les resulten más naturales e interesantes durante el registro.

En definitiva, para lograr un entorno propicio y tener más posibilidades de obtener muestras de habla espontáneas y, sobre todo, coloquiales, el entrevistador adopta una actitud relajada y amistosa, interviniendo como un participante más, manifestando sus opiniones en determinados asuntos y evitando imponer una estructura rígida en la conversación.

### 5.2.2. *Grabaciones secretas*

Las grabaciones secretas son registros realizados de manera oculta para captar el habla más natural y espontánea posible. Por lo general, suelen realizarse los registros sin el conocimiento expreso del informante.

El papel del investigador en este tipo de acopio de muestras de habla es la de infiltrarse en un grupo social (con al menos dos interlocutores, además del recopilador) y grabar sus interacciones lingüísticas<sup>13</sup>. A los informantes se les comunica que han sido grabados y se les solicita un permiso para utilizar el registro, siempre con fines científicos. Dicha autorización es firmada por los participantes, que aceptan a cambio de que cualquier dato personal que pudiera identificarlos en la grabación sea suprimido o anonimizado.

Al igual que en el caso de las entrevistas, los documentos generados por las grabaciones secretas (tanto los archivos de audio como las transcripciones) se etiquetan. El sistema utilizado es el mismo que el mencionado en las entrevistas, pero se marca con una C para diferenciarlo, indicando que se trata de una conversación coloquial espontánea.

### 5.3. *Investigaciones en las que se puede integrar el corpus*

El corpus oral de la comunidad de habla LGTBI es una herramienta valiosa, no solo porque supone el primer acopio de registros reales de este colectivo, sino porque, además, abre un camino en las investigaciones sobre el argot de esta comunidad.

A continuación, se describen algunas de las posibles investigaciones en las que este corpus (tanto en su estado actual como cuando sea ampliado) puede resultar particularmente útil.

---

<sup>13</sup> La problemática ética y legal que este tipo de técnicas puede presentar en la actualidad se plantea y se desarrolla en el apartado 5.4.3.

### 5.3.1. Estudios fonético-fonológicos

Este corpus puede facilitar el análisis de las características fonéticas y fonológicas del habla de la comunidad LGTBI. Investigaciones sobre patrones entonativos distintivos, uso marcado de ciertos fonemas y modulaciones prosódicas que reflejan la construcción de identidad en contextos de interacción social, especialmente en ámbitos como el uso de la entonación afectiva o variaciones en la pronunciación de vocales, consonantes o el seseo para proyectar afiliación a subgrupos, serían de gran relevancia.

### 5.3.2. Estudios léxicos y fraseológicos

El corpus permite estudiar el léxico específico y las expresiones idiomáticas que son propias de la comunidad LGTBI, enfocándose en el uso y evolución de términos específicos como *maricón* (Navarro-Carrascosa, 2019), *pluma* o *drag*, que han sido resignificados y reapropiados. La investigación podría abordar la creatividad léxica, incluyendo el uso de prefijos y sufijos con connotaciones humorísticas o irónicas (la formación de palabras, el uso de prefijos y sufijos, y las expresiones características ofrecen un campo amplio para investigar cómo se crean y utilizan estos términos en diferentes contextos), así como expresiones que forman parte del argot LGTBI en distintos contextos sociales y culturales, tales como el *ball culture* o *voguing*.

### 5.3.3. Estudios morfológicos

La morfología en el habla LGTBI representa una herramienta para reflejar identidades de género diversas. Este corpus permite investigar la adaptación de estructuras morfológicas para visibilizar el género no binario y analizar cómo se emplean innovaciones como el uso de terminaciones neutras, referidas al género gramatical no marcado (por ejemplo, *-e*), o formas flexionadas alternativas en un contexto donde se resignifican los sistemas tradicionales de género.

### 5.3.4. Estudios pragmáticos y sociolingüísticos

El corpus puede resultar de utilidad para estudios pragmáticos que analicen las intenciones comunicativas y el uso del lenguaje en interacciones sociales dentro de la comunidad LGTBI, como el uso de la ironía, el doble sentido o la subversión de normas sociales mediante el lenguaje. Además, permite analizar cómo las identidades de género y las orientaciones sexuales se expresan a través de actos de habla específicos, tales como saludos, cumplidos o estrategias de afrontamiento en contextos de discriminación o solidaridad.

### 5.3.5. *Estudios sobre identidades de género y uso del lenguaje*

Otras investigaciones que pueden sacar provecho de este corpus son aquellas que busquen entender cómo las personas no binarias y transgénero utilizan el lenguaje para expresar su identidad. Este compendio de muestras de habla proporciona datos sobre el uso de pronombres y formas no marcadas de género, así como sobre la adopción de términos específicos para describir experiencias únicas.

### 5.3.6. *Estudios comparativos*

El corpus también puede utilizarse en estudios comparativos entre el lenguaje de la comunidad LGTBI y el lenguaje de otros grupos sociales. Esto puede incluir comparaciones entre diferentes ciudades, edades y niveles socioeconómicos, proporcionando una visión más amplia de la variación lingüística.

También permite estudios comparativos internos dentro del colectivo, tales como el análisis de diferencias entre el habla de personas gais, lesbianas, bisexuales, trans y no binarias, o entre grupos de diferentes contextos culturales. Comparar el lenguaje empleado en ciudades específicas o en función de factores como la clase social o la edad puede proporcionar una perspectiva más matizada de la variación y el cambio lingüístico en esta comunidad.

## 5.4. *Problemas del corpus*

A pesar de los esfuerzos por construir un corpus oral amplio y representativo de la comunidad de habla LGTBI, existen varios problemas importantes que deben ser enfrentados para mejorar la calidad y la integridad de los datos recopilados. Los obstáculos más destacados incluyen la obtención de muestras de subgrupos menos visibilizados y la necesidad de incrementar los registros en modalidad de grabaciones secretas.

### 5.4.1. *Obtención de muestras en subgrupos menos visibilizados*

Uno de los retos más significativos en la recopilación de datos es la inclusión de subgrupos menos visibilizados, como las personas no binarias o las intersexuales. Estos grupos suelen estar subrepresentados en los estudios lingüísticos, lo que limita nuestra comprensión de la diversidad y riqueza del uso del lenguaje en diferentes identidades de género. Su representación en las primeras muestras del corpus publicadas por Navarro-Carrascosa (2023a) es mínima.

Para abordar este problema es esencial desarrollar estrategias inclusivas y sensibles que permitan contactar y reclutar participantes de estos subgrupos, asegurando así su representación adecuada en el corpus. A continuación, se enumeran algunas posibilidades:

— Buscar la colaboración de organizaciones LGTBI para facilitar el acceso a participantes de estos subgrupos.

— Desarrollar campañas de sensibilización que informen y animen a la participación de personas no binarias e intersexuales, así como otras minorías menos representadas.

— Adaptar las metodologías de recolección de datos para ser más inclusivas y respetuosas con todas las identidades de género, asegurando que todos los participantes se sientan cómodos y representados.

#### 5.4.2. *Tamaño de la muestra*

Como se ha indicado anteriormente, este es uno de los principales obstáculos en la elaboración de los corpus orales parciales. En el caso del corpus oral de la comunidad de habla LGTBI para calcular cuál es una buena muestra de partida es pertinente conocer el número de personas en España que pertenecen al colectivo (asumiendo que el corpus se limita a este país, aunque posteriormente pueda crecer con otros países hispanohablantes). Sin embargo, «no todas las personas con una orientación sexual o una identidad de género no normativas lo declaran en público y, por lo tanto, los censos pueden ser engañosos» (Navarro-Carrascosa 2023b: 92). No obstante, las asociaciones LGTBI ofrecen algunos datos a partir de los cuales se pueden realizar cálculos aproximativos<sup>14</sup>.

Por otro lado, además, es importante que las muestras sean representativas de todas las realidades LGTBI en proporción. Así, se establece un número aproximado de participantes que sean hombres cisgénero homosexuales, mujeres cisgénero homosexuales, mujeres transgénero, hombres transgénero, personas bisexuales (o pansexuales), personas intergénero, etc. Conseguir muestras representativas de todos los subgrupos resulta uno de los problemas que se presentan a la hora de elaborar el corpus. En la publicación de Navarro-Carrascosa (2023a), de los 60 informantes que participaron tanto en las entrevistas semidirigidas como en las grabaciones secretas, solo cinco de ellas eran bisexuales; y solo una, intergénero.

<sup>14</sup> En 2023 se proporcionan datos de la asociación FELGTBI+ que señalan que entre el 7 y el 8 % de la población española pertenece al colectivo LGTBI: [https://felgtbi.org/wp-content/uploads/2023/11/I-Informe-Estado-socioeconomico\\_felgtbi.pdf](https://felgtbi.org/wp-content/uploads/2023/11/I-Informe-Estado-socioeconomico_felgtbi.pdf).

### 5.4.3. Incrementar los registros en modalidad de grabaciones secretas

Otro problema que se debe abordar es aumentar el número de registros obtenidos mediante grabaciones secretas. Este método resulta valioso porque puede proporcionar una visión más auténtica y espontánea del habla, minimizando la paradoja del observador, donde la presencia del investigador puede influir en el comportamiento lingüístico de los participantes. Sin embargo, este tipo de registro presenta varios desafíos legales y éticos. La legislación vigente no permite grabar sin el conocimiento y consentimiento previo de los participantes, por tanto, es importante preparar un documento en el que se informe a los informantes de que van a ser grabados en algún momento futuro sin previo aviso y estos lo firman a modo de autorización. Tal y como se especifica en el proyecto Ameresco<sup>15</sup>, el consentimiento se firma en múltiples ocasiones para asegurar la claridad y el acuerdo continuo:

Esta primera firma manifiesta el consentimiento del hablante a ser grabado y recoge la fecha en que lo autorizó. En el segundo apartado, una vez realizada la grabación y habiendo sido informado el hablante de que acaba de ser grabado, se le da la posibilidad de escuchar la conversación y, si consiente en cederla para su análisis lingüístico, deberá firmar y fechar el segundo apartado de la autorización. Con esta firma, por tanto, el hablante manifiesta que ha escuchado la conversación y que está de acuerdo con que se haga pública para fines de investigación, previa anonimización de nombres y lugares. En último lugar, los hablantes deben firmar la sección para el tratamiento de datos personales, de acuerdo con la normativa vigente, y aceptar los términos. En el caso de no obtener dicha autorización o de que esta esté incompleta, el archivo no podrá utilizarse y deberá ser destruido (Carcelén y Uclés 2019: 22).

Otro obstáculo en la implementación de la grabación secreta, relacionado con el cumplimiento de la ley de protección de datos, es el desconocimiento de algunos participantes por parte del investigador o la persona encargada de realizar la grabación. En este caso, encontrar el momento adecuado para la grabación sin que los participantes lo esperen puede ser más complicado. Se requiere, por tanto, de una planificación cuidadosa y un enfoque sensible para no comprometer la ética del estudio.

Tratar estos retos es complejo, no obstante, existen algunas acciones que se pueden proponer al respecto:

— Delegar la responsabilidad de las grabaciones secretas en otros participantes que, previamente, hayan firmado el consentimiento, junto con el resto de los informantes. A modo de infiltrado,

<sup>15</sup> Proyecto que estudia y analiza el español coloquial de los países americanos, dirigido por Marta Albelda y Maria Estellés (Universitat de València).

grabará las conversaciones sin que el resto de los participantes sepa que va a hacerlo y su presencia no condicionará la actuación lingüística y conversacional.

— Proponer varios encuentros del investigador con el grupo que va a ser grabado sin que los participantes sepan en cuál se desarrollará el registro.

— Uso de tecnologías discretas, siempre que haya una financiación por parte de un proyecto y/o institución. En caso de que existan fondos, se pueden implementar micrófonos ambientales, colocados en lugares comunes donde los participantes suelen estar, o dispositivos portátiles pequeños, como grabadoras que puedan ser llevadas en bolsillos o mochilas.

## 6. El futuro del corpus oral de la comunidad de habla LGTBI

El Corpus Oral de la Comunidad de Habla LGTBI ha sentado una base sólida para el estudio de la lingüística *queer* hispánica, pero su desarrollo y expansión son esenciales para una comprensión más completa y diversa del lenguaje dentro de esta comunidad. A continuación, y a modo de conclusión, se detallan los principales objetivos y planes futuros para el corpus:

— Inclusión de grupos más invisibilizados: como ya se ha especificado en apartados previos, resulta fundamental obtener más muestras de habla reales de personas pertenecientes a grupos más invisibilizados dentro del colectivo LGTBI. Esto incluye a personas transexuales binarias, intergénero, bisexuales, y categorías que hasta ahora no han sido registradas en el corpus, como asexuales, demisexuales, entre otros. La inclusión de estos subgrupos permitirá una visión más completa y representativa de la diversidad lingüística y cultural del colectivo LGTBI.

— Registro de habla de generaciones mayores: registrar más grabaciones con personas de la franja de edad más mayor es un objetivo clave para poder comparar la evolución del argot LGTBI a través de distintas generaciones. Esto abrirá la puerta a estudios diacrónicos en lingüística *queer*, permitiendo analizar cómo ha cambiado y se ha adaptado el lenguaje de la comunidad a lo largo del tiempo, reflejando cambios sociales, culturales y estructurales.

— Expansión geográfica del corpus: la primera versión del corpus se limitó a las tres principales capitales de España: Madrid, Barcelona y Valencia. Resulta fundamental ampliar los registros a más ciudades

y municipios para capturar una mayor diversidad de dialectos y usos lingüísticos. Además, buscar registros en núcleos más rurales será interesante para identificar y estudiar diferencias lingüísticas significativas que puedan existir entre las áreas urbanas y rurales.

— Inclusión de variantes hispanoamericanas: siguiendo la misma línea que el punto anterior, el corpus debe, en su futura expansión, buscar registros de las distintas variantes hispanohablantes en otros países de habla española. La inclusión de datos de comunidades LGTBI en América Latina y otras regiones hispanohablantes permitirá comparaciones interculturales y enriquecerá el corpus con una variedad más amplia de influencias lingüísticas y culturales.

## BIBLIOGRAFÍA

- Azorín Fernández, Dolores (2005), «Corpus oral para el estudio del lenguaje juvenil y del español hablado en Alicante. El corpus ALCORE y COVJA», *Oralia*, 8: 265-287.
- Bejarano, Daniel, Andrea Llanos, Ruth Rubio, y Jonhatan Bonilla (2018), *Protocolo de transcripción ortográfica CLICC*, Bogotá, Instituto Caro y Cuervo.
- Bengoechea Bartolomé, Mercedes (2015), *Lengua y género*, Madrid, Síntesis.
- Briz Gómez, Antonio, y Grupo Val.Es.Co. (2002a), *Corpus de conversaciones coloquiales*, Madrid, Arco/Libros.
- Briz Gómez, Antonio (2012), «Los déficits de los corpus orales del español», en Tomás Eduardo Jiménez Juliá, Belén López Meirama, Victoria Vázquez Rozas y Alexandre Veiga Rodríguez (coords.), *Cum corde et in nova grammatica: estudios ofrecidos a Guillermo Rojo*, Santiago de Compostela, Universidad de Santiago de Compostela: 115-137.
- Calderón Noguera, Donald Freddy, y Julia Alvarado Castellanos (2011), «El papel de la entrevista en la investigación sociolingüística», *Cuadernos de Lingüística Hispánica*, 17: 11-24.
- Carcelén Guerrero, Andrea, y Gloria Uclés Ramada (2019), «Diseño y construcción de un corpus oral multidialectal. El corpus AMERESCO», *Normas*, 9: 17-36.

- Fernández Juncal, Carmen (2005), *Corpus de habla culta de Salamanca (CHCS)*, Burgos, Fundación Instituto Castellano y Leonés de la Lengua.
- Fernández Sanmartín, Alba (2022), *Teoría y métodos para la elaboración de corpus orales: la entrevista sociolingüística*, Santiago de Compostela, Universidad de Santiago de Compostela.
- Labov, William (1972), *Sociolinguistic patterns*, Philadelphia, University of Pennsylvania.
- Mira, Alberto (1999), *Para entendernos: diccionario de cultura homosexual, gay y lesbica*, Barcelona, Ediciones de la Tempestad.
- Mira, Alberto (2004), *De Sodoma a Chueca: una historia cultural de la homosexualidad en España en el siglo xx*, Madrid, Egales.
- Moreno Fernández, Francisco (2005), «Corpus para el estudio del español en su variación geográfica y social. El corpus PRESEEA», *Oralia*, 8: 123-139.
- Moreno Fernández, Francisco (2011), «La entrevista sociolingüística: esquema de perspectivas», *Linred. Lingüística en la red*, 9.
- Navarro-Carrascosa, Carles (2019), «Resignificación y reapropiación en el español coloquial: el caso de *maricón*», en Adrián Cabedo Nebot y Antonio Hidalgo Navarro (eds.), *Pragmática del español hablado: hacia nuevos horizontes*, Valencia, Universitat de València: 169-183.
- Navarro-Carrascosa, Carles (2023a), *Corpus oral de la comunidad de habla LGTBI: materiales para la investigación en lingüística queer hispánica*, Alcalá de Henares, Universidad de Alcalá.
- Navarro-Carrascosa, Carles (2023b), *Lingüística queer hispánica: las formas nominales de tratamiento de la comunidad de habla LGTBI*, Berna, Peter Lang.
- Pato, Enrique (2020), «El español en contacto con el francés en Quebec y su estudio gracias al 'Corpus Oral de la Lengua Española en Montreal' (COLEM)», *Boletín Hispánico Helvético*, 35-36: 263-287.
- Pereda, Ferran (2004), *El cancaneo: diccionario petardo de argot gay, lesbi y trans*, Madrid, Laertes.
- Pons Bordería, Salvador, y Leonor Ruiz Gurillo (2005), «Corpus para el estudio de la conversación coloquial: el corpus Val.Es.C.o. (Valencia. Español Coloquial)», *Oralia*, 8: 243-264.

- PRESEEA (2021), *Metodología del «Proyecto para el estudio sociolingüístico del español de España y de América»* (PRESEEA), Disponible en: <http://preseea.linguas.net> [Fecha de consulta: 03/03/2024]
- Recalde, Montserrat, y Victoria Vázquez Rozas (2009), «Problemas metodológicos en la formación de corpus orales», en Pascual Cantos Gómez y Aquilino Sánchez Pérez (eds.), *A Survey on corpus-based research /Panorama de investigaciones basadas en corpus*, Murcia, Asociación Española de Lingüística de Corpus: 37-49.
- Ridao Rodrigo, Susana (2021), «Aproximación a la transcripción de corpus orales: los símbolos de transcripción en corpus judiciales», *Revista de Llengua i Dret*, 77: 93-110.
- Rodríguez González, Félix (2008), *Diccionario gay-lésbico*, Madrid, Gredos.
- Sanmartín Sáez, Julia (1998), *Lenguaje y cultura marginal: el argot de la delincuencia*, Valencia, Universitat de València.
- Sanmartín Sáez, Julia (1999a), *Palabras desde el talego: el argot en la prisión de Valencia*, Valencia, Edicions Alfons El Magnànim.
- Sanmartín Sáez, Julia (1999b), *Diccionario de argot*, Madrid, Espasa.
- Torruella, Joan, y Joaquim Llisterra (1999), «Diseño de corpus textuales y orales», en José Manuel Blecua, Gloria Clavería, Carlos Sánchez y Joan Torruella (eds.), *Filología e informàtica: nuevas tecnologías en los estudios filológicos*, Barcelona, Editorial Milenio/Universitat Autònoma de Barcelona: 45-77.