

# Corpus oral de la comunidad de habla LGTBI

Materiales para la investigación  
en Lingüística *queer* hispánica

UAH MONOGRAFÍAS  
HUMANIDADES 93

## **Comité científico**

- JAIME ALVAR ÉZQUERRA  
Catedrático de Historia Antigua de la Universidad Carlos III.
- MANUEL ALEJANDRO RODRÍGUEZ DE LA PEÑA  
Profesor titular de Historia Medieval de la Universidad San Pablo-CEU.
- CARLOS ALBERTO GONZÁLEZ SÁNCHEZ  
Catedrático de Historia Moderna de la Universidad de Sevilla.
- FERNANDO MARIAS FRANCO  
Catedrático de Historia del Arte de la Universidad Autónoma de Madrid.
- RAFAEL BONILLA CEREZO  
Profesor titular de Literatura de la Universidad de Córdoba.
- CELIA CASADO FRESNILLO  
Profesora titular de Filología Hispánica de la Universidad Nacional de Educación a Distancia.
- JULIO DE LA CUEVA  
Profesor contratado doctor de Historia Contemporánea de la Universidad de Castilla-La Mancha.

# Corpus oral de la comunidad de habla LGTBI

Materiales para la investigación  
en Lingüística *queer* hispánica

Carles Navarro Carrascosa



Universidad  
de Alcalá

EDITORIAL  
UNIVERSIDAD DE ALCALÁ

La colección de Monografías de Humanidades de la Editorial de la Universidad de Alcalá ha sido distinguida con el Sello de Calidad en Edición Académica - Academic Publishing Quality (CEA-APQ).



El contenido de este libro no podrá ser reproducido,  
ni total ni parcialmente, sin el previo permiso escrito del editor.  
Todos los derechos reservados.

© De los textos: sus autores.  
© De las imágenes: sus autores.  
© De la ilustración de portada: Ronda Vázquez Martí  
© Editorial Universidad de Alcalá, 2023  
Plaza de San Diego, s/n  
28801 Alcalá de Henares  
[www.uah.es](http://www.uah.es)

I.S.B.N.: 978-84-19745-05-7  
Depósito legal: M-16046-2023

Composición: Solana e Hijos, A. G., S.A.U.  
Impresión y encuadernación: Solana e Hijos, A.G., S.A.U.  
Impreso en España

# ÍNDICE

<b>PRIMERA PARTE. INTRODUCCIÓN</b> .....	9
1. Objeto de estudio .....	11
2. Necesidad de un corpus de muestras reales de habla LGTBI .....	12
3. Cuestiones metodológicas .....	13
3.1. Variables .....	13
3.2. Informantes .....	16
3.3. Características de los corpus orales.....	17
3.4. Signos y convenciones de transcripción .....	19
3.5. Recogida de datos .....	21
3.5.1. Conversaciones coloquiales .....	21
3.5.2. Entrevistas .....	22
3.5.3. Corpus digital.....	29
<b>SEGUNDA PARTE. CORPUS ORAL DE LA COMUNIDAD DE HABLA LGTBI</b> .....	31
1. Conversaciones .....	33
C.MA.1 .....	33
C.MA.2 .....	41
C.MA.3 .....	45
C.VA.1 .....	50
C.VA.2 .....	93
2. Entrevistas .....	115
E.BA.1 .....	115
E.BA.2 .....	129
E.BA.3 .....	142
E.MA.1 .....	151
E.MA.2 .....	162
E.MA.3 .....	181
E.MA.4 .....	192
E.MA.5 .....	201

E.MA.6 .....	208
E.MA.7 .....	218
E.MA.8 .....	225
E.MA.9 .....	230
E.VA.1 .....	235
E.VA.2 .....	258
E.VA.3 .....	274
E.VA.4 .....	281
E.VA.5 .....	290
E.VA.6 .....	301
E.VA.7 .....	311
E.VA.8 .....	319
3. Corpus digital.....	327
<b>BIBLIOGRAFÍA.....</b>	<b>373</b>

# **PRIMERA PARTE**

## **INTRODUCCIÓN**





## 1. OBJETO DE ESTUDIO

En los últimos años se ha observado dentro del hispanismo una creciente tendencia por los estudios *queer*, trabajos que ya estaban consolidados en el ámbito estadounidense. La lingüística no se ha quedado atrás en esta tendencia al alza y cada vez son más las publicaciones científicas que trabajan el argot de la comunidad de habla LGTBI.

Este manual nace con el objetivo de ofrecer materiales para estudiar y analizar los códigos lingüísticos propios de esta subcultura y las intenciones pragmalingüísticas que se dan en la comunicación interna entre sus miembros y, de este modo, servir de herramienta en las investigaciones de la Lingüística *queer* (en adelante, LQ) hispánica.

Según Pereda (2004: 9), cualquier argot nace «de la necesidad, de la obligatoriedad de entenderse aquellos que pertenecen a un grupo diferenciado, y que ese lenguaje sea críptico de cara al resto de la sociedad. [...] Ello invita a que tanto el emisor como el receptor utilicen unos mismos códigos». El argot que se analiza en este manual, pues, surge en la época de más represión para el colectivo LGTBI, en tiempos en los que esconderse era una necesidad para sobrevivir y en los que las formas de comunicarse debían ser claves secretas para que no fueran detectadas por otros. Así, las personas pertenecientes a este colectivo no solo han desarrollado una lista de nuevas palabras, sino también una nueva manera de comunicarse y de entender la comunicación.

En la actualidad la represión y la censura hacia el colectivo LGTBI sigue latente, pero está más alejada de las instituciones que, progresivamente, han reconocido la presencia de gays, lesbianas, personas transgénero, etc., y han regulado parte de la legislación para equiparar los derechos de toda la ciudadanía, procurando que la identidad de género y la orientación sexual no sean motivo de discriminación. A pesar de que el camino todavía es largo, en las últimas décadas la comunicación clandestina ya no es una necesidad; sin embargo, esos códigos se mantienen y se extienden gracias a determinados medios de comunicación (como algunas revistas de temática gay) e incluso publicaciones literarias, cuyos principales protagonistas son personas del colectivo (Rodríguez, 2008). Así pues, este argot desempeña otras funciones en

la actualidad, como «reflejar los intereses y las necesidades de quienes integran la comunidad, expresar dentro de la subcultura gay una serie de roles, comportamientos y culturas sexuales, reforzar la cohesión interna y crear una realidad alternativa a la heterosexualidad hegemónica» (Bengoechea, 2015: 205).

En un estudio sobre la comunidad lingüística LGTBI, Navarro-Carrascosa (2020) realiza una caracterización del discurso propio de este grupo de hablantes, en el que destaca los siguientes rasgos:

- Formación de palabras, que puede responder a la necesidad de generar voces para nombrar referentes para los que no existe un término (como sucede con *mariendre*<sup>1</sup>); o para intensificar una burla, buscar la risa o, simplemente, intensificar la expresividad. En este último caso es habitual recurrir a determinados prefijos como *mari-* o *hetero-*, aunque también a generar palabras compuestas con el objetivo de que resulten divertidas (por ejemplo, *bollodrama*).
- Expresiones propias, que hacen referencia a la subcultura en la que se mueven los usuarios del argot. En el caso de la comunidad de habla LGTBI, la mayoría de expresiones que se registran hacen referencia a personalidades, habitualmente mujeres. Así, se encuentran colocaciones del tipo *quedarse muerta en la bañera* o *quedarse muerta y sin peluca*.
- Tendencia al uso del género gramatical femenino entre hombres homosexuales con intención afiliativa.
- Uso del género gramatical no marcado (es decir, el uso de las terminaciones en *-e* para no marcar el género del referente, en caso de que esta persona no se identifique ni con lo femenino ni con lo masculino). Se trata de una forma en gestación, pero que cada vez tiene más presencia en el discurso de esta comunidad de habla.

El corpus que se presenta en la segunda parte de este manual (el primero que recoge muestras de habla reales entre personas LGTBI y que queda a disposición de la comunidad investigadora) sirve como herramienta para la investigación de las estrategias lingüísticas que caracterizan la comunicación propia de la comunidad de habla LGTBI. Así, será beneficioso para completar estudios que hasta el momento quedaban incompletos, como cuestiones fonético-fonológicas, léxicas, fraseológicas, morfológicas, etc.

## 2. NECESIDAD DE UN CORPUS DE MUESTRAS REALES DE HABLA LGTBI

Para llevar a cabo un análisis pragmalingüístico con éxito, la mejor opción es recurrir a muestras reales de habla, tanto orales como escritas. Esta forma metodológica proviene de la Lingüística de Corpus que, como señala Parodi (2008), garantiza el estudio de la lengua con muestras reales, tanto orales como escritas:

---

<sup>1</sup> De *María* y *liendre*, para referirse a aquellas mujeres heterosexuales que tienen amistad con hombres homosexuales.

permite llevar a cabo investigaciones empíricas en contextos auténticos y que se constituye en torno a ciertos principios reguladores poderosos. Desde este enfoque, se estudia información lingüística original y completa, compilada a través de corpus, dado que la LC no se apoya la indagación de datos fragmentados, inconexos o de textos incompletos, sino que de unidades de sentido y con propósitos comunicativos específicos.

(Parodi, 2008: 96)

Un corpus es una serie de textos orales y/o escritos que se utilizan para la investigación lingüística. Son los ejemplos reales que sirven a los investigadores para observar los comportamientos (lingüísticos o no) de determinadas estructuras del sistema. Alcántara (2007) añade que los textos, en caso de que sean orales, deben estar transcritos para facilitar su estudio.

Hasta el momento, un corpus de las características que las investigaciones en LQ demanda es inexistente. Por eso, el propósito de este manual es poner a disposición de la comunidad investigadora el primer corpus *realia* que recoge el habla propia de la comunidad lingüística LGTBI para poder analizar sus peculiaridades.

La necesidad de disponer en el ámbito hispánico de un corpus de habla de la comunidad LGTBI es cada vez mayor. La LQ es una disciplina que cada vez va ocupando un lugar mayor en los estudios hispanistas<sup>2</sup>, por lo que la elaboración del primer corpus oral en español de la comunidad de habla LGTBI supone empezar a preparar el terreno para que siga abriéndose camino y consolide su posición en las investigaciones científicas.

### 3. CUESTIONES METODOLÓGICAS

#### 3.1. Variables

Para elegir de forma precisa a los participantes en la recogida de datos en una investigación lingüística centrada en un grupo específico, resulta fundamental precisar cuáles serán las variables necesarias que deben caracterizarlos. Las variables sociolingüísticas son aquellas que se relacionan con los informantes y sus características personales y sociales más relevantes. Labov (1983: 299) las define como «variables de naturaleza lingüística que están correlacionadas con alguna variable no lingüística del contexto social (hablante, receptor, público, etc.)».

Para Moreno (1990: 111), los rasgos lingüísticos que son focos de determinadas investigaciones suelen denominarse *indicadores*, y «las variables sociolingüísticas más desarrolladas suelen recibir el nombre de *marcadores*».

En sociolingüística, la observación, independientemente del método que se use, se

---

<sup>2</sup> Muestra de ello es la celebración del *I Congreso Internacional sobre Lengua y Aspectos LQTBIQ+* en 2019, de la segunda edición, en otoño de 2021, así como la publicación de un volumen con artículos científicos sobre LQ y otros aspectos relacionados (Sánchez Ibáñez, Fernández Cano, Pérez Bernabéu y Fernández de Pablo (eds.), 2019).

lleva a cabo teniendo en cuenta las distintas variables (o marcadores) que permitirán establecer una clasificación de los datos atendiendo a los informantes. Estas suelen ser edad, género<sup>3</sup> y nivel de instrucción, así como el rol que los informantes desempeñan en la sociedad (Calderón y Durán, 2009: 141). Sin embargo, aunque estas variables suelen ser constantes en la mayoría de estudios (Medina, 2010), se pueden encontrar otros trabajos en los que adquieren más relevancia otro tipo de variables, por ejemplo, en la investigación del argot de la delincuencia llevada a cabo por Sanmartín (1998), una de las variables relevantes es el tipo de delito cometido por el informante.

Por ello, para determinar los perfiles de los informantes que han participado en la elaboración de este corpus de muestras de habla reales, los marcadores que se han tenido en cuenta son la *orientación sexual* (el género hacia el que una persona siente atracción sexual) y la *identidad de género* (el género con el que una persona se identifica, independientemente de su biología), que son los ejes en torno a los cuales gira la pertenencia (o no) al colectivo LGTBI.

Por lo tanto, las personas que participan en este acopio deben estar situadas en una posición no normativa<sup>4</sup> en uno de los ejes o en ambos. Es decir, serán objeto de interés el habla de las personas cisgénero homosexuales, las personas transgénero (independientemente de su orientación sexual) y todas aquellas cuya identidad no quede enmarcada dentro de la dicotomía de género.

Las variables sociolingüísticas (o marcadores) que se han considerado para elaborar el corpus (y que constan en las fichas de presentación de cada grabación) son: la orientación sexual, la identidad de género, la edad y si el círculo social en el que se mueve el informante es LGTBI o no. A continuación, desarrollaremos la importancia para este estudio de cada una de las variables:

*La orientación sexual.* Distinguimos tres: (1) la *heterosexualidad* (personas que se sienten atraídas hacia personas del género opuesto), (2) la *homosexualidad* (personas que se sienten atraídas hacia personas del mismo género) y (3) la *bisexualidad* (personas que sienten atracción sexual hacia personas de ambos géneros). Es importante tener en cuenta que los estudios *queer* están abiertos a más de dos géneros y admiten una ruptura con la dicotomía social *hombre-mujer*. Así, podríamos contemplar una cuarta opción, la *pansexualidad*, que es la atracción hacia las personas de cualquier género, entendiendo que hay más de dos. No obstante, en algunas divisiones del colectivo se considera que hablar de *pansexualidad* es no asumir que las personas transgénero son hombres o mujeres propiamente dichos y, por lo tanto, se recomienda no hacerlo por considerarse un término transfobo. Así, para la elaboración de este corpus se ha decidido considerar ambas, *bisexualidad* y *pansexualidad*, dentro de una misma opción y solo se hablará de *bisexualidad*.

---

<sup>3</sup> En este trabajo no se habla de sexo sino de género para referirse al grupo al que pertenecen las personas en función de su identidad, ya que no nos referimos tanto a una cuestión biológica como identitaria.

<sup>4</sup> Entendiendo que la posición normativa es aquella que cumple con los requisitos de estilo de vida socialmente aceptados: identidad de género acorde con el sexo biológico y orientación heterosexual.

- *La identidad de género.* Esta variable, junto a la anterior, son las máximas protagonistas de los estudios de LQ, pues es en ellas donde se reconoce a las personas LGTBI y cada una de sus variedades. Las opciones que se encuentran en la identidad de género son, de nuevo, tres: (1) *personas cisgénero*, aquellas cuyo sexo biológico se corresponde con su identidad, pueden ser tanto hombres como mujeres; (2) *personas transgénero*, al contrario que en el caso anterior, la identidad de género de estas personas no se identifica con su sexo biológico, hay tanto mujeres transgénero (mujeres que nacieron con genitales masculinos) como hombres transgénero (hombres que nacieron con genitales femeninos); (3) *personas no binarias* (NB) o *intergénero*, aquellas que no se identifican ni como hombres ni como mujeres. No se ha considerado para la gestación del corpus un grupo de personas intersexuales<sup>5</sup> por considerar que, en cuanto a identidad de género, pueden ser identificadas en el resto de grupos.
- *Círculo social.* Después de los dos anteriores, este es el factor que más importancia tiene en LQ, pues las formas de habla específicas que se buscan en estos estudios solo se activarán en aquellas personas que tengan contacto social con miembros del colectivo LGTBI. Por tanto, se diferenciará entre aquellos informantes que sí tienen contacto social con personas LGTBI y aquellos que no lo tienen.
- *Edad.* Esta variable se menciona en la mayoría de estudios sociolingüísticos como una de las más relevantes para la variación lingüística (Zimmermann, 1996). Es de interés también para los estudios de LQ observar si hay diferencia en el uso de determinados recursos léxicos o gramaticales característicos de la comunidad de habla LGTBI (como puede ser el uso del G<sup>o</sup>GF entre hombres homosexuales) en función de la edad de los interlocutores y, por tanto, del contexto social durante el cual han crecido y la visión que se reflejaba en la sociedad de ese momento. Por tanto, en la elaboración de este compendio de muestras de habla no hemos seguido la división de franjas de edad que hacen varios estudios sociolingüísticos hispánicos (Gómez, 2003; Roselló, 2018; PRESEEA, 2021; entre otros). Teniendo en cuenta todo lo expuesto, se ha dividido a los informantes (todos ellos mayores de edad) en cuatro grupos: (1) de 18 a 25 años<sup>6</sup>; (2) de 26 a 35 años<sup>7</sup>; (3) de 36 a 45 años<sup>8</sup>; y (4) a partir de 46 años<sup>9</sup>.

---

<sup>5</sup> La intersexualidad es un fenómeno biológico por el cual una persona posee características físicas masculinas y femeninas. La proporción es variable en función del sujeto.

<sup>6</sup> Este grupo de edad ha crecido en una sociedad donde la realidad LGTBI ya está más visibilizada, lo que deriva en la aprobación de medidas legales, como la ley del matrimonio igualitario en el año 2005.

<sup>7</sup> Los encuestados pertenecientes a esta franja de edad han vivido una infancia donde las personas LGTBI ya están visibilizadas y aceptadas, pero, en general, la imagen que se da de ellas se relaciona con la mofa o la parodia. Todavía hay aspectos que se consideran un tabú (Mira, 2004).

<sup>8</sup> Este grupo de informantes ha crecido en un contexto en el que las personas LGTBI empiezan a reconocerse en la sociedad, pero el grado de aceptación hacia ellas es muy reducido (Mira, 2004).

<sup>9</sup> Estos informantes se han criado en un entorno de represión dictatorial (o con esta recién finalizada) en el que la imagen que se proyecta de los gais y las lesbianas es muy negativa y el resto del colectivo está

### 3.2. Informantes

En las grabaciones orales han participado 60 informantes<sup>10</sup> (17, en las conversaciones coloquiales; 43, en las entrevistas). El 60,67% de los participantes en los registros orales son hombres, de los cuales el 63,33% son homosexuales, el 1,66%, bisexuales y el 1,66%, heterosexuales; el 31,67% de los informantes son mujeres, de las cuales el 11,68% son homosexuales, un 6,67% se consideran bisexuales, un 1,66% han afirmado tener otro tipo de orientación sexual y son heterosexuales el 11,68% de las participantes; tan solo una persona no binaria participó en las entrevistas, lo que supone el 1,66%.

En las tablas posteriores se expone información relativa a los participantes en los registros relativa a la edad, el género y la orientación sexual. En la barra horizontal superior se apunta a la identidad de género de los participantes; en la lateral, la orientación sexual; y en la barra horizontal inferior se marca la franja de edad de los participantes:

**TABLA 1. INFORMANTES DE LAS CONVERSACIONES COLOQUIALES**

	Hombres				Mujeres				No binarias			
Homosexual	2	9										
Heterosexual					4	1						
Bisexual		1										
Otro												
	18-25	26-35	36-45	+45	18-25	26-35	36-45	+45	18-25	26-35	36-45	+45

**TABLA 2. INFORMANTES DE LAS ENTREVISTAS**

	Hombres				Mujeres				No binarias			
Homosexual	1	16	7	3	1	3	3					
Heterosexual			1			2						
Bisexual						3	1					
Otro					1				1			
	18-25	26-35	36-45	+45	18-25	26-35	36-45	+45	18-25	26-35	36-45	+45

La mayoría de los informantes son españoles nativos. Otros, son hispanoamericanos o procedentes de un país no hispanohablante. En estos dos últimos casos, los participantes, en el momento del registro, han pasado un mínimo de diez años viviendo en España.

---

totalmente invisibilizado (Mira, 2004). Por ejemplo, apenas se diferencia entre un hombre gay y una mujer transexual.

<sup>10</sup> El único participante que repite en más de una grabación es el investigador principal, que aparece en las conversaciones registradas en Valencia y en todas las entrevistas.

### 3.3. Características de corpus orales

La Lingüística con corpus muestra interés por los textos reales en uso, pero también por «la variabilidad inherente a ellos y a las situaciones y contextos de su producción» (Parodi, 2008: 96); por tanto, la obtención de esos textos reales requerirá de una metodología que ya ha sido puesta en práctica por muchos lingüistas en los últimos años. Para Biber, Conrad y Reppen (2006) esta metodología se resume en los siguientes puntos: es *empírica*, porque analiza ejemplos reales; usa *técnicas informáticas* para gestionar los textos; y el análisis de los datos es tanto *cualitativo* como *cuantitativo*:

The essential characteristics of corpus-based analysis are:

- it is empirical, analyzing the actual patterns of use in natural texts;
- it utilizes a large and principled collection of natural texts, known as “corpus”, as the basis for analysis;
- it makes extensive use of computers for analysis, using both automatic and interactive techniques;
- it depends on both quantitative and qualitative analytical techniques.

(Biber, Conrad y Reppen, 2006: 4<sup>11</sup>)

El grupo EAGLES<sup>12</sup> (1996) hace una serie de recomendaciones para elaborar corpus de calidad:

- Debe ser lo más extenso posible.
- Debe incluir ejemplos de distintas variantes lingüísticas.
- Todas las muestras deben tener un tamaño similar.
- La procedencia de todas las muestras del corpus debe estar especificada.

Parodi (2008) establece ocho recomendaciones importantes que se deben tener en cuenta a la hora de construir un corpus. Es un listado que, como apunta el autor, no está cerrado. Habla de extensión, formato, representatividad, diversificación, marcado o etiquetado, procedencia, tamaño de las muestras y clasificación temática. Teniendo esto en cuenta, el mismo autor concluye las siguientes características que debería tener (todas o la mayoría de ellas) un corpus lingüístico:

1. Recolección de textos en entornos naturales.
2. Explicitud de rasgos definitorios y compartidos por los textos constitutivos.
3. Formato final de tipo digital plano (\*.txt.) para cada texto o documento.

---

<sup>11</sup> «Las características esenciales del análisis basado en corpus son:

- es empírico, analizando los patrones reales de uso en textos naturales;
- utiliza una amplia y fundamentada colección de textos naturales, conocida como “corpus”, como base para el análisis;
- hace un gran uso de ordenadores para el análisis, utilizando tanto técnicas automáticas como interactivas;
- depende de técnicas analíticas tanto cuantitativas como cualitativas» (Traducción propia).

<sup>12</sup> Expert Advisory Group on Language Engineering Standards.

4. Tamaño, preferentemente, extenso.
5. Respeto a principios ecológicos.
6. Etiquetaje computacional semiautomático de naturaleza morfosintáctica u otra para cada texto.
7. Disponibilidad a través de medios computacionales.
8. Acceso a visualización completa de los textos que lo componen en formato plano.
9. Búsqueda de principios de proporcionalidad o representatividad (posiblemente estadística).
10. Sustento o procedencia inicial especificada.
11. Identificación de una organización en torno a temas, tipos de textos, registros, géneros, etc.
12. Registro de datos cuantitativos que permita la comparación y posible normalización de cifras.

(Parodi, 2008: 107)

Y añade que los textos que componen el corpus deben ser unidades completas; tener la modalidad (escrita u oral) identificada; tener contabilizados tanto el número de palabras como de oraciones; registrar los datos de procedencia relevantes, como la fecha y el lugar de registro, el investigador, el contexto de producción, etc.

Basándonos en las indicaciones de Parodi y el grupo EAGLES, las características de nuestro corpus pueden sintetizarse en los siguientes puntos:

1. Las grabaciones se han realizado en entornos familiares para los informantes y/o en situaciones lo más cotidianas para ellos.
2. Es extenso. Consta de 1458 minutos con 21 segundos, entre las 20 entrevistas y las 5 conversaciones coloquiales.
3. Incluye ejemplos de distintas variantes con el fin de ser lo más representativo posible. En la recopilación de entrevistas y conversaciones, hemos intentado obtener un muestrario de distintas orientaciones e identidades de género, especialmente a través de las entrevistas (que permiten una mayor programación y selección de los participantes). Así, no solo se ha entrevistado a hombres y mujeres cisgénero homosexuales y bisexuales, sino también a una persona NB. Aunque no se ha podido encontrar informantes que representen todas las identidades de género, sí que se ha conseguido una gran variedad.
4. Sobre el tamaño de las muestras, se ha procurado que las entrevistas tuvieran una duración similar (entre 45 minutos y una hora); lo que no ha sido posible conseguir con las conversaciones (que se han registrado con el método de grabación secreta), que tienen un carácter imprevisible y, además, las fuentes han sido distintas en cada caso.
5. Para identificar la procedencia de las muestras, los temas tratados, los tipos de participantes, el lugar de la grabación, etc., se ha recurrido a una ficha técnica de conversaciones coloquiales, que encabeza cada transcripción, y cuya plantilla hemos tomado de Briz y el grupo Val.Es.Co (2002).



6. Para etiquetar cada una de las muestras, se han seguido unos criterios para reconocer los rasgos principales de cada documento y que sean fácilmente reconocibles. En los siguientes apartados se explicarán estas etiquetas.

### 3.4. Signos y convenciones de transcripción

El primer paso para realizar un procesamiento de datos de los corpus orales es realizar las transcripciones de las grabaciones. Hay varios trabajos que describen las formas de transcribir los corpus orales, como el de Briz y grupo Val.Es.Co (2002), Hidalgo y Sanmartín (2005) y Albelda (2005). Estos trabajos hacen un repaso a las distintas formas de transcribir los corpus orales, señalando la transliteración ortográfica convencional y los etiquetados y todas las ventajas y desventajas de ambos. Apunta Albelda que las carencias de estos sistemas son varias, pues no recogen, por ejemplo, rasgos prosódicos, inflexiones entonativas o aspiraciones, fenómenos que sí quedan cubiertos con el sistema del grupo Val.Es.Co (2002), que «conjuga lo ortográfico con otros signos tipográficos y fonéticos» (2005: 386):

el sistema Val.Es.Co. (2002) intenta que el lector de la transcripción pueda reproducir de la manera más fielmente posible la emisión original, obtener una rápida captación cognitiva sin precisar de ninguna destreza especial, mostrar ciertas características de lo oral y, sobre todo, cuenta con un producto que facilita el análisis de la conversación. En este sentido, sigue los criterios apuntados por Payrató (1995) en relación con el sistema: supone una interpretación de datos fónicos, selecciona determinados fenómenos, se adecua a su objeto de investigación, es coherente con su base teórica y, sobre todo, es fiel a la representación de los datos.

(Albelda, 2005: 386)

Este sistema de transcripción pretende reflejar los rasgos propios de la conversación y que estos sean fácilmente interpretados por el lector, el cual podrá reproducir la conversación original descifrando los códigos que lo forman:

el sistema Val.Es.Co. prima la fidelidad a los datos de la muestra de habla, teniendo presente el tipo de discurso (conversacional) y la facilidad de lectura, frente a un exceso de representación de fenómenos diversos y un tipo de codificación que entorpece o impide apreciar con nitidez la realidad del proceso de habla. Es un sistema positivo para la transcripción de la conversación coloquial, pues con este objetivo fue creado, y atiende con precisión y a la vez con flexibilidad a las distintas necesidades de uso en función de las necesidades que dirijan el análisis del corpus.

(Albelda y Fernández, 2005: 71)

Tanto para las entrevistas como para las conversaciones, se ha optado por este sistema de transcripción, ya que nos ha facilitado reflejar los ejemplos orales para ilustrar fenómenos que se han detectado y permite el estudio de las muestras de habla

recogidas en distintos niveles, como puede ser, por ejemplo, el prosódico. De este modo, se ha llevado a cabo la creación de un corpus oral específico sobre el español en la comunidad de habla LGTBI, inexistente hasta el momento.

A continuación, presentamos el sistema de transcripción que se ha utilizado para las conversaciones coloquiales y las entrevistas, el de Briz y grupo Val.Es.Co (2002: 15-16):

:	Cambio de hablante
A:	Intervención de un hablante identificado como A
?:	Interlocutor no reconocido
§	Sucesión inmediata, sin pausa apreciable, entre emisiones de distintos hablantes
=	Mantenimientos del turno de un participante en un solapamiento
[	Lugar donde se inicia un solapamiento o superposición
]	Final del habla simultánea
-	Reinicios y autointerrupciones sin pausa
/	Pausa corta, inferior al medio segundo
//	Pausa entre medio segundo y un segundo
///	Pausa de un segundo o más
(5»)	Silencio (lapso o intervalo) de 5 segundos; se indica el número de segundos en las pausas de más de un segundo, cuando sea especialmente significativo
↑	Entonación ascendente
↓	Entonación descendente
→	Entonación mantenida o suspendida
PESADO	Pronunciación marcada o enfática
pe sa do	Pronunciación silabeada
(( ))	Fragmento indescifrable
((siempre))	Transcripción dudosa
((...))	Interrupciones de la grabación o de la transcripción
(en)tonces	Reconstrucción de una unidad léxica que se ha pronunciado incompleta, cuando pueda perturbar la comprensión
pa'l	Fenómenos de fonética sintáctica entre palabras, especialmente, marcados o que puedan dificultar la lectura
°( )°	Fragmento pronunciado con intensidad baja o próxima al susurro
h	Aspiración de la «s» implosiva
(RISAS, TOSES, GRITOS)	Cuando aparecen al margen de los enunciados. En el caso de las risas, si acompañan a lo dicho, se transcribe el enunciado y en nota al pie se indica «entre risas»
aa	Alargamientos vocálicos
aaa	Alargamientos prolongados
nn	Alargamientos consonánticos
¿i !?	Interrogaciones exclamativas
¿ ?	Interrogaciones. También para los apéndices del tipo «¿no?, ¿eh?, ¿sabes?»
Letra cursiva	Reproducción o imitación de emisiones. Estilo directo, característico de los denominados relatos conversacionales
Notas a pie de página	Anotaciones pragmáticas que ofrecen información sobre las circunstancias de la enunciación. Rasgos complementarios del canal verbal. Añaden informaciones necesarias para la correcta interpretación de determinadas palabras (la correspondencia extranjera de la palabra transcrita en el texto de acuerdo con la pronunciación real, siglas, marcas, etc.), enunciados secuencias del texto (p.e., los irónicos), de algunas onomatopeyas, etc.

\* Las incorrecciones gramaticales (fónicas, morfosintácticas y léxicas) no aparecen marcadas por lo general.

\* Los antropónimos y topónimos no se corresponden con los reales.

### 3.5. Recogida de datos

#### 3.5.1. *Conversaciones coloquiales*

Las conversaciones coloquiales han sido registradas en modo oculto, lo que supone que los informantes no son conscientes de que están siendo grabados hasta el final de la grabación. Esto garantiza la naturalidad de sus emisiones y producciones lingüísticas.

En este punto es interesante hablar de la *paradoja del observador* (Labov, 1983) que plantea la dificultad del investigador para aproximarse a un lenguaje coloquial o informal con personas fuera de su entorno o en situaciones artificiales. Silva-Corvalán (2001: 52) señala que los «estudios sociolingüísticos realizados durante los últimos quince años han tratado de resolver de diferentes e ingeniosas maneras la *paradoja del observador*. Una de estas ha sido la grabación secreta» en la que el investigador no tiene por qué estar presente. Es por esto por lo que desde un primer momento se ha tenido como la principal técnica de recogida de datos. Además, la naturalidad y espontaneidad de los participantes será total si estos desconocen que están siendo grabados. Se han conseguido cinco grabaciones de conversaciones coloquiales.

Una vez finalizada cada grabación y los participantes han sido informados del registro, se les ha pedido que firmen un permiso para poder utilizar dicha grabación. En este permiso, el investigador se compromete a anonimizar cualquier indicio por el cual el informante pueda ser identificado.

Dos de las grabaciones han sido efectuadas en la ciudad de Valencia y las otras tres, en la ciudad de Madrid. Ninguna conversación fue grabada en Barcelona, donde sí fueron registradas varias entrevistas. Los participantes son, fundamentalmente, hombres homosexuales y mujeres heterosexuales muy integradas en el colectivo. Dos de los registros se han producido durante el descanso del espacio de trabajo de los participantes; otras dos, en un restaurante; y la última, en los jardines de El Retiro de Madrid. Además, dos de ellas han estado a cargo del autor de este trabajo y las otras tres (las que han tenido lugar en Madrid) han sido conseguidas por colaboradores externos. El tiempo de cada registro es diferente, pues depende de la espontaneidad de la situación que se desarrolla (se encuentran, así, duraciones muy diversas si se trata de una cena entre amigos o de un breve descanso en el trabajo).

En total, se han grabado 282 minutos y 38 segundos.

Cada una de las conversaciones (y sus respectivas transcripciones) se han etiquetado de la siguiente manera:

- En primer lugar, la etiqueta está encabezada por una letra *C* que indica que se trata de una conversación coloquial y no de una entrevista.
- En segundo lugar, se añade el código de la ciudad: *MA*, para aquellas grabadas en Madrid y *VA*, a para las que se han registrado en Valencia.
- Cierra el código de la etiqueta el número que indica el orden en el que fueron grabadas.

Así, por ejemplo, el documento etiquetado como *C.MA.2* se corresponde con la segunda conversación coloquial grabada en Madrid. El corpus de conversaciones coloquiales se compone de los siguientes documentos:

**TABLA 3. INVENTARIO DE CONVERSACIONES COLOQUIALES**

Documento	Ciudad de grabación	Duración	N.º de participantes
C.VA.1	Valencia	156' 36''	4
C.VA.2	Valencia	106' 36''	5
C.MA.1	Madrid	11' 31''	2
C.MA.2	Madrid	03' 52''	3
C.MA.3	Madrid	04' 03''	3

### 3.5.2. Entrevistas

Las entrevistas han surgido en esta recolección de muestras de habla, en primer lugar, para cubrir las pocas grabaciones de conversaciones coloquiales que se pueden obtener dada la complejidad de crear un corpus de lengua hablada informal, especialmente con un perfil de informantes tan característico y acotado; en segundo lugar, para registrar opiniones y actitudes lingüísticas de los informantes de una manera más extensa que con la encuesta. Esto es así dado el carácter dialógico de la entrevista, que «permite recolectar datos de una manera más expositiva» (Catellano, 2012: 14).

La principal diferencia entre las conversaciones y las entrevistas es que estas últimas inician su grabación con el informante consciente de ella. Además, mientras las conversaciones tienen un desarrollo natural, las entrevistas están semidirigidas, lo que quiere decir que hay una planificación y que el investigador formula preguntas y propone temas, aunque deja que las respuestas e intervenciones cambien el rumbo de la entrevista.

Se ha considerado que las entrevistas deben dirigirse a más de una persona cada vez (Silva-Corvalán, 2001; Moreno, 2011) para poder registrar interacción entre los interlocutores, los cuales deben conocerse entre sí, y que sea más viable utilizar rasgos lingüísticos característicos del habla LGTBI que puedan resultar de interés para nuestro estudio. Aunque esto no siempre ha sido posible (hay dos entrevistas con un único informante).

La grabación del grupo parece ser la que ofrece la mejor posibilidad de obtener una muestra natural de conversación social entre conocidos.

[...] En todo caso, la inclusión de grabaciones individuales o de grupo depende de la finalidad del estudio.

(Silva-Corvalán, 2001: 57)

La grabación en grupo consiste en que amigos o conocidos interaccionen entre sí en vez de hacerlo con el entrevistador, en espera de que este sistema rebaje la tensión provocada al saberse objeto de las observaciones de un extraño y estimule las estrategias habituales de interacción entre iguales. [...] Se considera que en las sesiones de grupo se produce habla más espontánea.

(Fernández, García, Recalde y Vázquez, 2008: 8)

En total, se grabaron veinte entrevistas en tres capitales españolas por considerarse que el habla en estas podrían resultar representativas de la peninsular: Madrid (nueve entrevistas), Barcelona (tres) y Valencia (ocho), de septiembre de 2018 a octubre de 2019. El entrevistador ha sido en todas ellas el autor de este trabajo. Diecisiete entrevistas han sido realizadas por pares (es decir, con dos entrevistados simultáneamente); dos de ellas, con un solo informante; una, con cuatro participantes; y otra entrevista se ha llevado a cabo con tres interlocutoras. Cada uno de los registros tiene una media de una hora de duración. Todos los participantes, al igual que en las conversaciones, son personas LGTBI o personas que conviven en su día a día con ellas.

Labov (1983) elaboró un protocolo para las entrevistas sociolingüísticas, según el cual la entrevista debe estar dividida en módulos temáticos entre los cuales deberá haber preguntas de transición. Por su parte, el entrevistador debe tratar de rebajar la formalidad en la que el contexto de la propia entrevista puede situar a los informantes. El tono, por tanto, debe ser más conversacional. Para ello, Labov (1983) propone una serie de estrategias:

- Que el entrevistador ofrezca algunas opiniones personales sobre los temas tratados y que responda a las preguntas que ha formulado;
- Que el entrevistador intercale comentarios entre las intervenciones de los informantes;
- Que el entrevistador, cuando detecte un tema de interés para los informantes, recurra a él.

Silva-Corvalán (2001: 61) coincide en algunos puntos con el lingüista estadounidense y sugiere una serie de conductas para el entrevistador con el objetivo de que la entrevista se convierta en una conversación más fluida y que el entrevistado participe de forma más natural:

- a. El investigador es un participante en la conversación. Su papel no es solamente hacer preguntas, sino también responder preguntas, hacer comentarios, ofrecer información, intercambiar ideas, etc.
- b. El investigador debe permitir al hablante suficientes oportunidades para cambiar el giro de la conversación y proponer nuevos temas. Esto se logra a veces manteniéndose en silencio por unos segundos, o llenando una pausa con expresiones semivacías o marcadores como “Así, pues, así, así”, “Bueno, qué vida esta, tantas cosas”.

- c. El investigador debe sentirse tranquilo y cómodo durante la grabación y demostrar genuino interés en lo que dice su interlocutor. Debe además adoptar una actitud despreocupada, natural y apropiada a una conversación social.
- d. El investigador debe hablar sobre sí mismo y ser específico en sus intervenciones ya que este enfoque crea una atmósfera más íntima y personal que estimula también a hablar al individuo estudiado.

Atendiendo a las estrategias de Labov (1983) y Silva-Corvalán (2001) para la recolección de muestras orales a través de entrevistas semidirigidas, se ha optado por seguir las cuatro estrategias siguientes:

- i. El entrevistador es participante y debe ser el encargado de crear un clima familiar durante la grabación. Debe hacer intervenciones, como un interlocutor más, para que los entrevistados olviden su carácter de investigador:
  - (1) B: ya/ es que yo- es que lo pienso/ es que ningún pri- no sé// pero parece muy raro/ nooo- / de hecho para mí es muy extraño estar como en un ambiente muy heteronormativo (risas) entonces→  
**A: a mí también me pasa**

[E.MA.2]

En el fragmento que se ilustra en el ejemplo (1) y en los siguientes, el entrevistador es el participante A. Como se ve en su intervención, intenta hacer comentarios a las intervenciones de los informantes para que el ambiente adquiera un clima más familiar y las posteriores intervenciones de C y B sean más fluidas y espontáneas.

- ii. El entrevistador debe comentar, en ocasiones, las intervenciones que hacen los informantes, mostrando interés por lo que dicen y acuerdo en las opiniones que expresan, generando, así, confianza para que el lenguaje de los entrevistados sea más fluido:
  - (2) C: sí/ porque lo femenino se asocia con algo malo  
**A: estáaa- en la escala social está por debajo deee- de lo masculino/ entonces→ es una degradación/ ¿no? para ellos/ supongo**  
 C: claro/ sí

[E.MA.3]

- (3) B: de hecho/ tenemos una discusión§  
 C: §y digo *mariconada*  
 B: = dice *mariconada* y entonces yo digo/ *deja de decir mariconada para algo negativo/* y siempre decimos/ *vamos a decir mariconada cuando sea algo positivo/ digamos/ qué rico está el helado/ menuda mariconada//* para darle la vuelta al lenguaje  
 C: pero luego no lo hacemos

B: no lo hacemos// pero *mariconada* lo dices tú siempre// bueno/ siempre//  
yo nunca lo digo/ tú lo dices

**A: ¿como Mecano?**

C: (risas) como Mecano

[E.MA.1]

En los ejemplos (2) y (3) el entrevistador (nuevamente, el participante A) comenta las intervenciones de los informantes para intentar mantener la conversación y que surjan (o sigan surgiendo) las formas que busca o sigan hablando con normalidad de los temas que le interesan.

- iii. El entrevistador tiene que procurar proporcionar información personal (experiencias, vivencias...) relativa a los temas que se tratan. Especialmente si los informantes son conocidos, debe procurar enfatizar estos aspectos para que el hecho de que se encuentran en un entorno de grabación sea olvidado lo más rápido posible por los informantes y el ambiente resulte mucho más familiar y cotidiano, de tal modo que las producciones de habla serán más naturales:

(4) A: es verdad/ que yo le regañé luegooo/ aal

C: sí

A: al chaval

C: yo eso me lo perdí/ gracias/ señor

**A: yo creo que te hubiera gustao/ estuve muy bien**

C: no/ a mí nooo

B: ¿a quién le reñiste?

**A: al chaval que nos vendió las entradas↑ porque no nos había dicho lo dee**

[E.VA.1]

- iv. El entrevistador debe insistir en los temas que susciten interés en los informantes, bien sea de una manera positiva o negativa, especialmente si el tema genera pasión en las intervenciones de los entrevistados:

(5) B: especialmente con la palabra *pasiva*// es TERRIBLEE// es decir/ *es que es súper pasiva*/ ¡aviam!<sup>13</sup> (2») [¡aviam!]

C: [¿(( ))] *pasiva* femenino?

B: es decir/// ¿qué estamos haciendo?

**A: hay una canción// ¿sabéis? de una travesti/ ¡bueno! de [una drag queen]**

B: [lo que pasa es]

que es una versión de la versión original en inglés/ *the boy is a bottom* y clarooo// es- es lo guay/ es como no tienen género los adjetivos en inglés/ pues es mucho más cómodo↑ yo también lo entiendo

[E.BA.1]

<sup>13</sup> Interjección catalana, equivalente a la expresión castellana *¡vamos a ver!*

Por último, en (5) A detecta un tema que parece interesar mucho a los informantes e intenta mantenerlo en la conversación para que siga fluida y ellos continúen expresándose con absoluta libertad.

Sin embargo, Wolfson (1976) advierte que no se puede considerar la entrevista como habla espontánea, ya que hay cierta planificación previa (por parte del investigador) y los informantes son conscientes de que están siendo grabados.

Los estudios sociológicos señalan tres tipos de entrevistas (Acevedo y López, 1988; Castellano, 2012; Selltiz, Jahoda, Deutsh y Cook, 1970; López, 1994): la entrevista estructurada, la semidirigida o semilibre y la entrevista libre. Del tipo de entrevista dependerán la estructuración de la situación de la entrevista, el grado de planificación y las formas y contenidos de las preguntas.

- *La entrevista estructurada*: Está elaborada minuciosamente. Las preguntas han sido preparadas (López, 1994) y el desarrollo de la entrevista no puede salirse del plan estipulado. El investigador debe tener preparadas preguntas a posibles respuestas del entrevistado.
- *La entrevista semidirigida (o semilibre)*: Las preguntas han sido programadas, existe un plan de desarrollo por parte del entrevistador; sin embargo, los entrevistados gozan de absoluta libertad en sus respuestas. El papel fundamental del entrevistador es que la dirección de la conversación que se pueda derivar de las preguntas no se aleje del tema propuesto. El propósito de las preguntas es «buscar registros de actuación lingüística» (Castellano, 2012: 14) y opiniones y actitudes sobre dichos registros. Selltiz, Jahoda, Deutsh y Cook (1970: 303-304) destacan que en este tipo de entrevistas

la función del entrevistador es simplemente exhortar al entrevistado a charlar acerca de un tema propuesto con un mínimo de guía o preguntas. Se anima al interrogado a charlar extensa y libremente estando alerta a los sentimientos expresados en sus afirmaciones y por medio del reconocimiento caluroso, si bien no comprometido, de los sentimientos del sujeto.

Castellano Asencio las cataloga como el tipo de entrevistas más utilizadas en los estudios sociolingüísticos.

- *La entrevista libre*: No hay ningún tipo de guion ni de preparación, a excepción del tema del que se quiere hablar. El entrevistador improvisa en función de las respuestas que va obteniendo. Según Fernández, «la idea fundamental consiste en que es el discurso de los hablantes el que construye el texto, de manera que [...] la acción comunicativa de los entrevistadores tiene como objetivo ir creando una situación interaccional [...] más cercana, informal y flexible» (2018: 149-150). La diferencia fundamental respecto a los tipos anteriores es que, en la entrevista libre, aunque pueda partir de una idea previa, el entrevistador se deja llevar y sigue la conversación de un modo más orgánico.



Para cumplir con nuestro propósito, se ha preparado una entrevista semidirigida, pues se pretende que los entrevistados hablen con libertad, de forma espontánea y relajada, para obtener muestras de habla naturales. Con todo, se busca cierto control, ya que los temas tratados son significativos para las posibles investigaciones que se puedan apoyar en estos materiales y se busca que hablen de ellos durante toda la grabación.

La entrevista preparada para este estudio se ha estructurado en dos partes, cada una de ellas con dos apartados:

1. Preguntas rompehielo.
  - 1.1. Relación entre los informantes
  - 1.2. Su situación LGTBI
2. Preguntas sobre el lenguaje de la comunidad de habla LGTBI
  - 2.1. Debates sobre términos o formas
  - 2.2. Preguntas sobre el uso de términos y formas apelativas

El tipo de discurso que se busca es de tipo argumentativo, pues se pretende que los informantes expongan su opinión y reflexionen sobre el uso que ellos mismos hacen sobre determinadas voces.

Sobre las preguntas para romper el hielo, se ha optado por elegir aquellas que implicaban a todos los informantes de la entrevista, del tipo *¿Cómo os conocisteis?* o *¿Cómo os lleváis?*, de modo que se vean obligados a interactuar entre ellos recordando o corrigiéndose. En una segunda parte, las preguntas pasan a ser más específicas sobre la situación LGTBI de los informantes (si no son específicamente miembros del colectivo, se les preguntará sobre la situación LGTBI de su interlocutor o interlocutora).

El segundo bloque de preguntas es el que introduce el tema de la investigación. La primera parte de este bloque propone debates sobre temas lingüísticos que afectan a la comunidad LGTBI. El objetivo es, por un lado, que interactúen todos los participantes, buscando así alguna muestra de habla que resulte relevante para nuestro corpus y, por otro lado, buscar opiniones, actitudes y reflexiones sobre los temas propuestos. Así, se proponen los siguientes asuntos para poder debatir:

- El G°GF usado entre hombres homosexuales.
- El G°GNM o no binario.

La segunda parte del bloque introduce preguntas propiamente dichas sobre el uso de palabras o formas más concretas, consideradas términos o formas propias del argot de la comunidad lingüística LGTBI.

El final de este segundo bloque está compuesto de una serie de enunciados en los que se usa la palabra *maricón*: desde *maricón de mierda* hasta *la quiero mucho, maricón*. Se les propone a los informantes dichos enunciados y ellos deben interpretar quién es el emisor y quién el receptor de cada uno, la relación entre ambos, la situación comunicativa, etc.

La duración de todas las entrevistas, como ya se ha comentado, es de entre cuarenta y cinco minutos a una hora, como recomiendan diversos autores (Hudson, 1981; López, 1994; Fernández, García, Recalde y Vázquez, 2008; Castellano, 2012), a excepción de

algunas de ellas, que tienen una duración de entre 90 y 110 minutos. En el conjunto de todas las entrevistas, se han registrado un total de 1175 minutos y 43 segundos.

En todas las sesiones de grabación descubierta, el entrevistador ha explicitado la finalidad de la entrevista, como recomiendan hacer López (1994) y Castellano (2012), y los entrevistados han firmado una autorización para ser grabados; todas ellas se han realizado en un lugar conocido para los informantes, con el fin de que se sintieran más cómodos y estuvieran más abiertos a hablar. En caso de que la grabación no pudiera tener lugar en la residencia de uno de los entrevistados, se llevó a cabo en un lugar público elegido por ellos, habitualmente un bar o cafetería.

Al igual que las conversaciones, las entrevistas se han etiquetado y, para ello, se ha seguido un criterio similar:

- En primer lugar, una *E* que la cataloga como entrevista.
- En segundo lugar, se añade el código de la ciudad: *MA*, para Madrid; *BA* hace referencia a las entrevistas grabadas en Barcelona; y *VA*, a las registradas en Valencia.
- Cierra el código de la etiqueta el número que indica el orden en el que fueron grabadas.

Así, por ejemplo, el documento etiquetado como *E.VA.4* se corresponde con la entrevista grabada en Valencia en cuarto lugar.

La siguiente tabla contiene la clasificación de las entrevistas registradas para nuestro corpus:

**TABLA 4. INVENTARIO DE ENTREVISTAS**

Documento	Ciudad de grabación	Duración	N.º de entrevistados
E.BA.1	Barcelona	51' 24''	2
E.BA.2	Barcelona	111' 00''	5
E.BA.3	Barcelona	55' 36''	3
E.MA.1	Madrid	63' 40''	2
E.MA.2	Madrid	70' 02''	2
E.MA.3	Madrid	67' 47''	2
E.MA.4	Madrid	42' 16''	2
E.MA.5	Madrid	48' 11''	2
E.MA.6	Madrid	45' 49''	2
E.MA.7	Madrid	31' 45''	1
E.MA.8	Madrid	53' 03	2
E.MA.9	Madrid	29' 04''	2
E.VA.1	Valencia	74, 14'	2
E.VA.2	Valencia	60' 03''	2
E.VA.3	Valencia	49' 33''	2
E.VA.4	Valencia	85' 54''	2
E.VA.5	Valencia	50' 36''	2
E.VA.6	Valencia	51' 13''	2
E.VA.7	Valencia	64' 59''	2
E.VA.8	Valencia	69' 34''	1

Todas las transcripciones, tanto de las entrevistas como de las conversaciones coloquiales, van encabezadas de una ficha de datos de cada registro, la cual se ha basado en las utilizadas por el grupo Val.Es.Co (2002).

### 3.5.3. *Corpus digital*

Pese a que el corpus de grabaciones en su totalidad (tanto entrevistas como conversaciones coloquiales) supera los 1400 minutos de grabación, hay muchos recursos lingüísticos que todavía no han podido ser registrados y reflejados en nuestras transcripciones (como, por ejemplo, el uso del G°GNM). Apelamos a la dificultad de conseguir una gran variedad de informantes dispuestos a ser grabados para una investigación de estas características.

Por lo tanto, el vacío que deja la ausencia de estos recursos en las grabaciones orales ha sido llenado con publicaciones en redes sociales, fundamentalmente *Twitter*, *WhatsApp*, *Facebook* y aplicaciones de contactos.

Las 203 publicaciones incluidas en el corpus han sido seleccionadas teniendo en cuenta los siguientes dos criterios:

- La fecha de publicación. Se han seleccionado publicaciones entre los años 2012 y 2021.
- Palabras utilizadas. Todas las publicaciones seleccionadas utilizan formas referidas a tipificaciones sexuales: *maricón*, *bollera*, *bujarrón*, *tortillera*, *hetero*, *mariliendre*, *machirulo* y las flexiones del G°GF dirigidas a hombres y el G°GNM.

Dada la dificultad para conocer la orientación sexual de los perfiles en redes sociales, la clasificación de las variables se ha llevado a cabo en base a la información presentada en sus perfiles, que no siempre es exacta. De este modo, los emisores se han clasificado en tres grupos de género (hombre, mujer, persona NB) y un cuarto grupo para aquellas publicaciones en las que se desconoce el género de su emisor. En cuanto a las variables que caracterizan a los miembros de la comunidad LGTBI, han sido unificados en un solo grupo por la dificultad, en muchas ocasiones, de determinar si el publicante es homosexual, bisexual o de cualquier otra identidad. Por todo ello, se han establecido tres variantes: LGTBI, heterosexual y una tercera para aquellos casos en los que esta información no se ha podido clarificar.

De este modo, los publicantes de esta parte del corpus son un 33,03% de hombres LGTBI, un 30,66% de mujeres LGTBI y un 1,87% de personas NB. No se ha podido clarificar el género de los participantes en un 1,42% de los ejemplos. En la siguiente tabla se resume la tipología de emisores atendiendo a ambos criterios: género y pertenencia al colectivo LGTBI:

TABLA 5. EMISORES DE LAS PUBLICACIONES DE REDES SOCIALES

	Hombre	Mujer	NB	?
LGTBI	70 (33,03%)	65 (30,66%)	4 (1,87%)	3 (1,42%)
No LGTBI	40 (18,87%)	29 (13,68%)		
?		1 (0,47%)		

Los datos que se ofrecen en cada una de las publicaciones que se incluyen son, en primer lugar, la muestra, en letra cursiva<sup>14</sup>. En segundo lugar, el perfil del emisor, es decir, de la persona que ha escrito y publicado el enunciado (se señala, fundamentalmente, su género y orientación sexual). Después, se ha señalado el perfil del referente también en lo que se refiere a la identidad de género y a su orientación sexual<sup>15</sup>. Los dos últimos datos presentados son las fuentes (es decir, la red social de la que se ha extraído) y la fecha de publicación:

FIGURA 3. EJEMPLO DEL CORPUS DIGITAL

<p><i>Bolli, que no sabías que era mi cumpleeee</i>  <b>Emisor:</b> Mujer cisgénero homosexual  <b>Receptor/Referente:</b> Mujer cisgénero homosexual  <b>Fuente:</b> Twitter  <b>Fecha de publicación:</b> 20 de junio de 2017</p>
---

Las publicaciones en redes sociales han sido etiquetadas a partir de dos factores: la red social de donde se ha extraído el ejemplo, el año de publicación y el número de orden de la muestra. Estos rasgos aparecen a modo de etiqueta cuando se inserta un ejemplo en el análisis y un número para distinguirlos. Las RR.SS. se identifican a partir de los siguientes códigos:

- *WA*: Whatsapp
- *FB*: Facebook
- *TW*: Twitter
- *APP*: Aplicaciones de contactos

Así, por ejemplo, la primera muestra extraída de la red social Facebook con fecha de publicación del 13 de enero del año 2016 tendrá la siguiente etiqueta:

[FB.16.1]

<sup>14</sup> Como es propio de las RR.SS., muchas de las publicaciones que se han utilizado para el acopio de muestras tenían errores ortográficos, de puntuación y, en muchos casos, se utilizaban abreviaturas de algunas palabras. Para facilitar la lectura, en el corpus se han corregido estos errores, manteniendo el contenido del texto. Por otro lado, en muchas publicaciones aparecen personas etiquetadas. En este caso, para mantener el anonimato de los informantes, cuando aparece una etiqueta personal con el nombre del perfil del usuario de la determinada red social, se ha eliminado y se ha sustituido por el símbolo XXXXXX.

<sup>15</sup> Estos datos se han podido conseguir acudiendo al perfil de los usuarios que han realizado la publicación y de los receptores, en caso de que estos estuvieran etiquetados.