

## COLFI: un corpus de películas y series de televisión españolas

ANA LLOPIS CARDONA

Universitat de València

[ana.b.llopis@uv.es](mailto:ana.b.llopis@uv.es)

ORCID- iD: <https://orcid.org/0000-0002-7531-8553>

### RESUMEN

En la actualidad solo existen corporus de películas y series de televisión en lengua inglesa e italiana. El español no cuenta con corporus similares ni los grandes corporus actuales han incluido este tipo de material. Ante este vacío, surge el *Corpus oral de coloquialidad ficticia* (COLFI), compuesto por transcripciones de películas y series de televisión españolas desde los años cincuenta del siglo pasado hasta la actualidad. Este artículo presenta los aspectos relevantes del diseño y composición de COLFI. Comienza abordando cuestiones preliminares, como los requisitos legales, el título o el contenido de los archivos. A continuación, se explican en detalle los criterios que se han considerado para la selección de las obras audiovisuales. Seguidamente, se describe la ficha de metadatos de las obras y de los personajes, así como el sistema de transcripción que se ha implementado. El artículo se cierra con unas reflexiones sobre su alcance y limitaciones.

**PALABRAS CLAVE:** corpus, películas, series de televisión, inmediatez comunicativa, coloquial, ficción

## COLFI: a corpus of Spanish movies and tv series

### ABSTRACT

To date, film and television corpora have been developed exclusively for English and Italian. Spanish, by contrast, lacks comparable corpora, and major existing corpora have not incorporated this type of material. To fill this gap, *Corpus oral de coloquialidad ficticia* (COLFI) has been developed, consisting of transcriptions of Spanish films and television series produced from the 1950s to the present day. This article outlines the key aspects of COLFI's design and composition. It begins by addressing preliminary issues such as legal requirements, the corpus title, and the content of the files. It then details the criteria applied in selecting the audiovisual works. Subsequently, it describes the metadata sheets for both the works and the characters, as well as the transcription system. The article concludes with reflections on the scope and limitations of the corpus.

**KEY WORDS:** Corpus, Films, TV series, Communicative immediacy, Colloquial, Fiction

## 1. INTRODUCCIÓN<sup>1</sup>

El *Corpus Oral de Coloquialidad Ficticia* (COLFI) es una base de datos<sup>2</sup> que recoge transcripciones de películas y series de televisión que simulan diálogos caracterizados por su inmediatez comunicativa en gran parte de sus secuencias, y que fueron estrenadas desde los años 50<sup>3</sup> hasta la actualidad. El corpus surgió con el fin de llenar el vacío de obras audiovisuales de ficción en los corpus de referencia y de contribuir en cierta medida a la limitada representación del registro coloquial –especialmente en el periodo previo a la irrupción de los corpus orales–. El acceso a estos textos dialogales ofrece nuevos materiales para explorar procesos de difusión y cambio lingüístico y sociopragmático desde la segunda mitad del siglo XX. A su vez, permite comparar la mímesis conversacional producida en la ficción filmica y televisiva con las conversaciones auténticas, y examinar cuestiones pragmalingüísticas asociadas a la construcción de los guiones, de los personajes y de las escenas de la ficción audiovisual.

La siguiente sección trata de aspectos preliminares y pertinentes en la composición del corpus: 1) se ubica COLFI en el panorama de los corpus del español hablado y de los corpus históricos de oralidad, 2) se destaca su originalidad y relevancia, 3) se especifican los requisitos legales que debe cumplir un corpus de estas características, 4) se explica el título asignado en relación con las etiquetas que ha recibido el fenómeno de la mímesis de la conversación, finalmente 5) se expone qué material recogen los archivos de COLFI. La tercera y cuarta sección se ocupan de la composición y configuración del corpus. En la tercera se detallan los criterios de selección de las obras audiovisuales; seguidamente, se ilustra la estética realista mediante un esbozo que recorre géneros, autores y obras de manera sintética; en último lugar, se considera el realismo lingüístico a la luz de la variación intra e interlingüística. En la cuarta sección se explican los metadatos de las obras audiovisuales y de los personajes recogidos en las fichas de análisis e incorporados en la base de datos. En la quinta sección se indican las pautas de transcripción y etiquetado empleadas, y se presenta el repositorio con su motor de búsqueda. El artículo finaliza con unas breves reflexiones sobre las oportunidades y limitaciones del discurso telecinemático.

## 2. CUESTIONES PRELIMINARES

### 2.1. Breve recorrido por los corpus de español hablado y corpus históricos de oralidad

El inicio de los corpus de español hablado se remonta al *Proyecto Estudio coordinado de la norma lingüística culta de las principales ciudades de Iberoamérica y de la Península Ibérica*,

<sup>1</sup> Aprovecho estas líneas para expresar mi agradecimiento a los colegas y colaboradores que se han cruzado en el camino de la elaboración del corpus. Merecen una mención especial Mario Barcala, por sus explicaciones sobre cuestiones técnicas y sugerencias; Adrián Cabedo y María Estellés, por sus consejos en la visualización de algunos datos; Raquel Zapater y a Daniel Gascó, por sus muchas orientaciones en la selección de las películas.

<sup>2</sup> En la versión actual cuenta con 95 archivos, 68 de películas y 27 de episodios de series de televisión, que suman un total de 752 165 palabras ortográficas.

<sup>3</sup> En la siguiente fase de elaboración del corpus, se pretende equilibrar la representación de las décadas, así como incluir películas de la década de 1940.

coordinado por Juan M. López Blanch en el marco de Programa Interamericano de Lingüística y Enseñanza de Idiomas (PILEI). Este proyecto recopiló numerosos corpus de entrevistas semidirigidas entre las décadas setenta y noventa que culminaron con la publicación de una selección representativa del macrocorpus a finales de los noventa (Samper, Hernández y Troya 1998 en Briz y Albelda 2009). En esta década cogió el testigo del macrocorpus transatlántico el *Proyecto para el estudio sociolingüístico del Español de España y de América* (PRESEEA), coordinado por Francisco Moreno, con la novedad de la representación de los tres niveles socioculturales. A su vez, surgieron los primeros corpus de conversaciones espontáneas orientados a recolectar materiales del registro coloquial, entre los que destaca por su repercusión el corpus Val.Es.Co. (Briz y Grupo Val.Es.Co. 2002), ahora digitalizado y ampliado con la incorporación de conversaciones de las siguientes décadas (Val.Es.Co. 3.0.). También sobresale el CORLEC por incluir tipos de textos muy variados, desde conversaciones a clases, conferencias y programas radiofónicos, que comprenden el continuo de la variedad diafásica en su modalidad oral. Tanto este corpus de referencia como otros corpus orales, por ejemplo, el COVJA, terminaron volcados en el CREA para su consulta con motor de búsqueda. Una década más tarde, siguiendo la estela de un amplio espectro de registros, se publica el C-ORAL-ROM, que comprende un corpus de español, francés e italiano. En esta primera década del 2000 se elaboran corpus con el propósito de recoger muestras de grupos sociales o zonas geográficas apenas representadas en los corpus anteriores, como el lenguaje de los adolescentes (COLA) o el español rural (COSER). Ese afán por cubrir más áreas, sobre todo urbanas, o por disponer de muestras más actualizadas, ha conducido a la proliferación de nuevos corpus, como el ESLORA, el CORMA o el corpus de conversaciones Ameresco, entre otros (cf. Briz y Carcelén 2019; Rojo 2024). Asimismo, se han elaborado nuevos corpus con fines didácticos (CAES) o de investigación, entre los que cabría mencionar aquellos de enfoque pragmático. También habría que resaltar la incorporación al CORPES XXI de textos procedentes de medios orales y/o digitales, con una impronta coloquial rastreable en distintas marcas.

La increíble labor desarrollada en la elaboración de corpus del español hablado actual ha corrido pareja a la profusión de corpus históricos tanto de referencia general (CORDE, CdH, CdE), como más específicos (CHARTA, CODEA + 2022, PostScriptum). Dado el mayor predominio de textos concebidos desde la distancia comunicativa en estos corpus, ha venido creciendo el interés por disponer de textos más próximos a la lengua hablada de tiempos remotos, bien sean auténticos (cartas privadas, crónicas, declaraciones, etc.), bien de ficción (obras de teatro, novelas, relatos, coloquios) (cf. Bertolotti y Company 2024). Muchas de estas recopilaciones se han desarrollado en el marco de estudios particulares y de investigaciones de tesis doctoral. Hasta la fecha, aunque sigan en curso, se cuenta ya con corpus electrónicos de acceso libre, como ODE (*Oralia diacrónica del español*), que contiene inventarios de bienes, declaraciones de testigos en juicios penales y certificaciones de barberos y cirujanos producidos entre 1492 y 1833 en España (Calderón y Vaamonde 2020). También se puede consultar Post Scriptum, que recoge cartas privadas de la Edad Moderna, y algunas pocas de finales del siglo XV, producidas en España en su mayoría y algunas pocas en América y Filipinas. Las cartas y otros documentos como notas y recados forman parte de CORDIAM, que contiene también documentos administrativos, jurídicos y cronísticos. Asimismo, habría que mencionar proyectos en curso de gran interés mediante los que se pretende seguir cubriendo parcelas de la diacronía de la oralidad, como CorDisDial (*Corpus del Discurso Dialógico en la historia de las lenguas romances*), que incluye traducciones de textos dialógicos desde el siglo XV al XIX (Del Rey y Carmona 2024).

Finalmente, otros corpus de carácter diacrónico disponibles en la red son HumText, un corpus del humor escrito que reúne producción humorística desde el siglo XV hasta la actualidad, y Humcor, un corpus oral multimodal que recoge chistes, monólogos y sketches procedentes de programas de radio y televisión con muestras desde 1900 hasta 2024. Sin duda, la selección de estos textos audiovisuales (entretenimiento/humor) y el periodo que comprende cubre un vacío de gran atractivo e interés.

## 2.2. Relevancia y originalidad de COLFI

El desarrollo de la radio y la televisión, junto al surgimiento del cine y de artes menores como la historieta gráfica a lo largo del siglo XX, proporcionan nuevos y valiosos materiales para el estudio del español coloquial (Pons Bordería 2023; Llopis Cardona 2024). De hecho, en estudios recientes se ha reconocido que la representación de la conversación coloquial es bastante fiable en películas y, más aún, en series de televisión (Bednarek 2010; Álvarez-Peyre 2011; Jucker 2021; Jucker y Landert 2023). La proximidad a la conversación real es más prominente que en los diálogos narrativos o teatrales, bien sea por el formato audiovisual, bien por el proceso de producción-realización. Sin embargo, los corpus de referencia apenas han incorporado transcripciones procedentes de estos medios, salvo excepciones puntuales<sup>4</sup>, ni el español dispone de un corpus específico de transcripciones o subtítulos de películas y series de televisión.

En italiano<sup>5</sup> se creó en 2005 el *Pavia Corpus of Film Dialogue* (PCFD), compuesto por transcripciones de películas italianas y de películas inglesas en su versión original y en la versión doblada al italiano. En lengua inglesa existe por el momento un corpus de subtítulos de películas y series de televisión (*TV Corpus/Movies Corpus*, Davies 2021) y un corpus de transcripciones de series de televisión (*Sydney Corpus of Television Dialogue*, Bednarek 2018). A pesar de la gran cantidad de palabras que almacena el primero, reporta menor fiabilidad por las abundantes alteraciones de los subtítulos con respecto a las palabras pronunciadas.

Ante este panorama, surgió el *Corpus oral de coloquialidad ficticia* (COLFI), compuesto por transcripciones de películas y series de televisión españolas desde los años cincuenta del siglo pasado hasta la actualidad. Este corpus se inició en 2020 con el proyecto “Difusión del cambio lingüístico en el español coloquial durante los últimos cincuenta años” (GV/2020/157) y se ha desarrollado en una primera fase gracias a la Beca Leonardo a Investigadores y Creadores Culturales 2023 de la Fundación BBVA (LEO23-2-10560).

## 2.3. Requisitos legales en el uso de las obras audiovisuales

Una cuestión preliminar concierne a las prescripciones legales a las que se debe ceñir el corpus. COLFI recopila transcripciones de obras audiovisuales, principalmente de los diálogos, sin incluir la imagen de las escenas ni su audio. Estas transcripciones corresponden al guion en bastante medida, aunque el texto final varíe según la libertad interpretativa concedida por los directores durante el rodaje. En este sentido, debe

<sup>4</sup> A excepción de las transcripciones de algunos programas televisivos de los ochenta y noventa en el CREA, así como algún guion incorporado en CORPES XXI.

<sup>5</sup> En francés se viene desarrollando una amplia investigación del discurso telecinemático, no obstante, se recurre a compilaciones realizadas *ex profeso* para estos estudios (cfr. Pusch en prensa). Una situación análoga puede observarse en otras lenguas románicas, como el italiano y el español.

respetar los derechos de autor de los guionistas, es decir, «los autores del argumento, la adaptación y los del guión [sic] o los diálogos» (BOE, n.º 97, 22/04/1996, artículo 7), y ajustarse a los requisitos legales. La Ley de la Propiedad Intelectual permite incluir fragmentos de obras de naturaleza escrita, sonora o audiovisual en trabajos personales o en bases de datos, siempre y cuando su uso se destine a «fines docentes o de investigación» y se indique «la fuente y el nombre del autor de la obra utilizada». Así se indica en el artículo 32:

Artículo 32. Es lícita la inclusión en una obra propia de fragmentos de otras ajenas de naturaleza escrita, sonora o audiovisual (...), siempre que se trate de obras ya divulgadas y su inclusión se realice a título de cita o para su análisis, comentario o juicio crítico. Tal utilización sólo podrá realizarse con fines docentes o de investigación, en la medida justificada por el fin de esa incorporación e indicando la fuente y el nombre del autor de la obra utilizada. (BOE, n.º 97, 22/04/1996)

COLFI muestra los fragmentos de las transcripciones –aproximadamente cinco intervenciones– acompañadas de una cabecera en la que aparecen los guionistas de la obra en la que ha sido localizada la palabra o expresión buscada junto a otros datos de la obra audiovisual.

## 2.4. COLFI y las denominaciones de la «mímesis conversacional»

Varias son las denominaciones<sup>6</sup> que ha recibido la simulación de diálogos orales en la ficción: *mímesis de lo hablado* (Oesterreicher 1996), *mímesis de la oralidad* (López Serena 2007a), *oralidad fingida* (Brumme 2008a, 2008b), *oralidad simulada* (Mancera 2009), *oralidad ficticia* (Cano 2012; Kailuweit 2015; Kailuweit y Steffien 2023), *mímesis conversacional o verosimilitud conversacional* (Del Rey 2019) o *lengua fictiva realista* (Gancedo 2022). Salvo Brumme (2008a, 2008b), estos autores se referían a la ficción literaria, pero, por extensión, estas etiquetas podrían abarcar la ficción audiovisual, correspondiente a los filmes y series de televisión, y la visual gráfica, propia de las historietas gráficas, puesto que pretenden reflejar que se trata de simulaciones del habla real –no representaciones fidedignas–, que resultan de un proceso creativo y mediante las cuales el autor dota de verosimilitud el mundo ficticio creado.

Tras ciertas consideraciones, se optó por designarlo como *Corpus oral de coloquialidad ficticia*, con el acrónimo COLFI. A nuestro juicio, aunque los complementos especificativos posean su propia justificación, resultaría inapropiado incurrir en un exceso de nominalismo, atribuyendo mucho mayor peso a la forma o etiqueta que al contenido, que es el que mejor representa en este caso la base de datos.

El título elegido coincide parcialmente con algunas de las etiquetas mencionadas. Con el término *oral* se pretende subrayar que se ofrece la transcripción del texto finalmente representado en escena que llegó a la audiencia a través del canal fónico, que fue integrado en un formato audiovisual. El término clave es, sin duda, el binomio *coloquialidad ficticia*. El registro coloquial es el parámetro que ha guiado la selección de las obras de ficción filmica y televisiva, esto es, de textos creados y concebidos desde el prisma de la inmediatez comunicativa (Kock y Oesterreicher 2007 [1990]). *Coloquialidad* remite, por tanto, a la

<sup>6</sup> Vid. Brumme (2008a), que recoge y diserta sobre los términos que se habían venido utilizado para referirse a este fenómeno, a saber: *oralidad prefabricada*, *oralidad construida*, *oralidad ficticia o literaria* y *oralidad fingida*.

variedad diafásica predominante o a la variedad concepcional desde la que se construye el texto. Por su parte, *ficticia* pone el acento en que se trata de un universo discursivo imaginado mediante el que se caracterizan personajes y escenas que forman parte de narraciones, en este caso cinematográficas o televisivas, que crean la ilusión de realidad o autenticidad. Se prefirió *ficticia* a *fingida* porque, tratándose de voces con una base semántica ('inventado, falso') y etimológica<sup>7</sup> común, en nuestra opinión, el primero evoca con mayor suerte el universo de la ficción, empañado en el segundo por su uso frecuente con el sentido de ' fingir' o 'engañoso'. En el mismo sentido, Kailuweit y Steffien (2023) sostienen que es preferible hablar de «oralidad ficticia», pues «la oralidad no se utiliza con la intención de engañar, sino que forma parte de un mundo ficticio» (cfr. Kailuweit 2015). De hecho, la mayor recurrencia a *fingida* en la tradición española parece deberse a los volúmenes editados por Brumme (2008a, 2008b), quien años después muestra sus reservas ante el término (Brumme 2012 en Kailuweit 2015). En último lugar, ulteriores especificativos como *filmica* y *televisiva*, referidos al objeto de estudio, generaban un título excesivamente largo y están representados en el logotipo del corpus, de ahí que se decidiera prescindir de ellos.

## 2.5. Los archivos de COLFI

En COLFI los archivos contienen las transcripciones<sup>8</sup> del producto audiovisual que llega a la audiencia. Se descartó recopilar los subtítulos de las películas y/o series televisivas habida cuenta de que sufren bastantes alteraciones<sup>9</sup> con respecto al diálogo reproducido, tal como lo confirman estudios recientes, los cuales insisten en la dudosa fiabilidad de utilizar subtítulos sin revisión (Jucker 2021; Bednarek 2010, 2023). Asimismo, no se consideró adecuado<sup>10</sup> incluir los guiones de las obras seleccionadas principalmente porque ha predominado y predomina la concepción del guion como una herramienta de trabajo que guía la realización de una película (cfr. Pollaro 2011: 301-304). El guion<sup>11</sup> contiene la trama, las acciones, los personajes, la tensión dramática en una serie de diálogos encadenados en los que se evita comunicar información que se transmitirá por otros medios, sean visuales (distintos planos) o auditivos (música de fondo u otros sonidos). Se concibe como «un texto narrativo-descriptivo con vistas a su rodaje, o más exactamente a convertirse en filme» (Vanoye 1996 en Pollaro 2011: 292), de ahí su carácter «transitorio» (cfr. Sánchez-Escalona 2016: 12).

Durante el rodaje los diálogos cobran vida, se actualizan a veces con retoques e incluso se pueden extender o añadir pequeños diálogos. Berlanga, por ejemplo, tendía a enriquecer sus guiones sugiriendo nuevas escenas o intercambios con tal de plasmar mejor

<sup>7</sup> Ambos son términos derivados de *ingo*, -*ngere*: *ingido* es el participio del verbo *ingir*, mientras que en *ficticio* procede del adjetivo latino *ficticius*, que deriva del sustantivo *fiction*, -*onis*, y este del verbo *ingo*, -*ngere* más el sufijo latino *-tio*, *-tionis*.

<sup>8</sup> En 5 se indicará el sistema y convenciones de transcripción empleados.

<sup>9</sup> Ciertamente, la calidad varía según cumplan un propósito profesional, como los subtítulos para sordos o los subtítulos que ofrecen en la actualidad muchas plataformas –por ejemplo, Netflix–, o según el esmero del aficionado que los ha preparado. En cualquier caso, las limitaciones del espacio obligan a prescindir de elementos verbales o a recurrir a ciertas sustituciones repercutiendo en desajustes con el diálogo representado.

<sup>10</sup> La obtención de los guiones no estaba exenta de dificultades en bastantes casos.

<sup>11</sup> Los diálogos, junto a los recursos y técnicas previstos, quedarán articulados en el *Storyboard*, que servirá de base al director o directora y al resto de agentes involucrados (cfr. Stadler y McWilliam 2009).

su película imaginada; Azcona, el gran guionista español, prefería que se tomase como referente casi terminado la escritura guionizada (cfr. García-Aguilar 2018). Por su parte, los actores y actrices encarnan los personajes interpretando las interacciones de forma multimodal y, *por lo general*, ciñéndose al diálogo guionizado y a las indicaciones del/de la directora/a. No obstante, se realizan a veces escenas improvisadas bien sugeridas por los/as directores/as durante el rodaje, bien por iniciativa del actor o de la actriz. Azcona elogiaba a José Luis López Vázquez por su capacidad para interpretar papeles, que se reflejaba a veces en ese seguir actuando y alargando su turno después de que el director dijera «corten» (cfr. Sabanés 2022). La actriz Carmen Maura es reconocida por un estilo interpretativo que logra simular con notable eficacia la espontaneidad, lo que le concede el aval de directores para introducir pequeñas variaciones personales en sus intervenciones.

Esta tendencia experimenta un viraje en las producciones televisivas desde los noventa cuando se crean cadenas privadas y se multiplican las series de televisión. «Se trabaja a una velocidad que impide que a un actor se le pueda exigir mucho, aparte de que los textos no suelen estar tan elaborados» (Aguirre 2008 en González 2015: 379). A las condiciones laborales como la premura de tiempo y el menor presupuesto, se suma que el guion no suele ser obra de un autor, sino antes bien suele tratarse de un trabajo colaborativo por parte de un equipo de creadores que se distribuyen diferentes episodios (cfr. Stadler y McWilliam 2009: 174). Aunque este modo de rodar y producir obras audiovisuales fue visto como mala práctica, trajo consigo mayor improvisación y flexibilidad en la actuación (González 2016: 378-381): actores y actrices de la pequeña pantalla comenzaron a dar mayor colorido a sus intervenciones.

Chierichetti (2024) advierte que el equipo actoral de la serie televisiva *Patria* introdujo marcadores conversacionales, intensificadores y otros recursos sintácticos en su representación creando el efecto de mayor autenticidad y humanismo en sus personajes. La escalada en la improvisación alcanza cotas altas en la serie televisiva *Paquita Salas* (cfr. López Serena 2022). El actor protagonista, Brays Efe, revela que «en la dirección y los ensayos de la serie una de las frases más repetidas era que buscáramos nuestra propia expresión, una vez ya sabíamos lo que nuestro personaje quería lo de menos era si se expresaba exactamente como estaba escrito en el papel o no» (Efe, 2017 en López Serena 2022: 69). Este modo de dirigir e interpretar, gestado en los platós televisivos y del que los directores Javier Ambrossi y Javier Calvo son un exponente claro, ha traspasado al cine no solo de la mano de directores que dan el salto a este medio, sino también de directores de filmes que experimentan en su modo de rodar. Así, refiriéndose a las películas *La línea del cielo* e *Isla bonita*, Fernando Colomo comenta «son mis favoritas porque son las que rodé con menos medios, pero con más libertad. Y sin guion escrito...» (Tovar 2022).

### 3. CRITERIOS DE SELECCIÓN DE LAS OBRAS AUDIOVISUALES<sup>12</sup>

La selección de las películas y series de televisión que conforman el corpus se aborda desde una perspectiva holística, dado que muchos de los criterios admiten distintos grados y matices que impiden concebirlos como condiciones imprescindibles de manera aislada. Estos criterios<sup>13</sup> comprenden tanto aspectos externos a los diálogos como factores internos de naturaleza situacional y lingüística. En una primera instancia, se privilegia la consideración de los aspectos externos y situacionales internos, por su carácter más objetivo, mientras que los criterios lingüísticos se incorporan de manera complementaria<sup>14</sup>. Teniendo en cuenta la prevalencia de la lengua estándar en la producción audiovisual, la relativa o mayor presencia de marcas del registro coloquial, de sociolectos o de dialectos las convierte en candidatas más adecuadas para el corpus. El propósito fundamental de establecer esta guía metodológica era garantizar cierto grado de coloquialidad en los diálogos. La Figura 1 presenta el conjunto de aspectos externos e internos a los diálogos que se han sopesado.

---

<sup>12</sup> Una vez concluida la primera fase, se procurará incluir un número proporcional de palabras por década (criterio ulterior), con el objetivo de garantizar la representatividad del corpus y facilitar futuros estudios diacrónicos que requieran una masa textual equilibrada. Del mismo modo, se intenta mantener un equilibrio entre la cantidad de películas y series de televisión, al menos en aquellas décadas en las que pudiera ser factible.

<sup>13</sup> Gancedo (2022: 78-83) propone una serie de criterios para la selección de obras de teatro con vistas a analizar aspectos socioculturales desde una perspectiva histórica; se enmarca, por tanto, dentro de los estudios que toman como corpus las obras de ficción con tintes realistas. Nuestra propuesta coincide en cierta medida con los criterios planteados por Gancedo, los cuales se corresponden con los factores externos e internos de índole lingüística aunque con matices diferentes, y se añaden los factores internos situacionales.

<sup>14</sup> Aun siendo los factores más importantes, tomarlos como criterio básico implicaría cierta subjetividad –la identificación del registro coloquial no derivaría de un análisis sino de la percepción–, que conllevaría caer en la circularidad (Mc Enry y Wilson 1996: 14).

	Autor/es del guion
Aspectos externos a los diálogos	Lengua
	Historia
	Estilo
	Cantidad de diálogos
Aspectos internos a los diálogos:	Relación entre los interlocutores
	Marco comunicativo <sup>15</sup>
<i>Parámetros situacionales</i>	Temática
	Finalidad
	Número de participantes – dinámica conversacional
<i>Estrategias de verbalización y variación</i>	Marcas variacionales diafásicas
	Marcas variacionales diastráticas (y dialectales)

Figura 1. Aspectos externos e internos a los diálogos

### 3.1. Aspectos externos de los diálogos

Entre los aspectos externos a los diálogos, se encuentran el/los autor/es del guion, la lengua o versión de los diálogos (original/doblada), la historia y la cantidad de diálogos de la obra audiovisual.

#### a) Autores españoles o en colaboración con españoles

Las obras audiovisuales debían pertenecer a la producción del cine o televisión española con guiones escritos por españoles o en colaboración con españoles<sup>16</sup>. De este modo, se asegura que la obra se circunscriba a una comunidad sociocultural compleja, con normas y valores que pueden ser más o menos heterogéneos, pero que, a su vez, da por sentados una serie de presupuestos y presenta una seña de identidad. La mayor parte de las obras registradas fueron escritas por guionistas españoles, no obstante, que en casos *puntuales* no lo fueran no fue óbice para descartar las obras si sus diálogos eran representativos del registro coloquial. Se incorporaron dos películas del director italiano Marco Ferreri, que fueron coguionizadas con Rafael Azcona (*El pisito* 1958) y Leonardo Martín (*Los chicos* 1959), por tanto, se construyeron los diálogos con hablantes nativos. También se incluyó la película *El olivo* (2016), dirigida por Icíar Bollaín con guion del escocés Paul Laverty, quien llevó a cabo un trabajo previo de campo, desarrolló la historia con sus correspondientes diálogos, caracterizando a sus personajes y detallando pormenores de la

<sup>15</sup> No se incluye como aspecto interno «la relación entre los interlocutores y el referente comunicativo» (cfr. López Serena 2007b, 2024). Este criterio, contemplado por Koch y Oesterreicher (2007 [1990]), se manifestará a través de determinadas estrategias de verbalización, que manifestarán el campo referencial (por ej. déficticos).

<sup>16</sup> Esta restricción se debe a una cuestión práctica: el proyecto no era viable si se pretendía incluir una selección de la producción en español en los países de América, mal que nos pese.

trama; no obstante, tratándose de un trabajo colaborativo el de guionistas y directores<sup>17</sup>, no cabe duda de que la directora revisaría a fondo la escritura guionizada.

b) *Versión original en español europeo principalmente*

La versión original del guion debía ser el español europeo. La presencia de otros idiomas se circunscribe a préstamos o cambios de código en intervenciones y determinados diálogos. Asimismo, se dio cabida a algunas<sup>18</sup> películas con guion original escrito en una lengua cooficial, pero dobladas y comercializadas en español, con el fin de poder representar mínimamente el español de zonas bilingües<sup>19</sup>, como el español-catalán. Ciertamente, la productora asume la tarea de traducción en la que el guionista no puede ejercer control alguno, salvo excepciones<sup>20</sup>; no obstante, al tratarse de comunidades bilingües, la traducción es fruto de hablantes que conocen bien la realidad social y lingüística, y el doblaje<sup>21</sup> suele correr a cargo de los mismos actores.

c) *Obras ambientadas en el tiempo del estreno*

Las historias narradas debían estar situadas en la España de la década del estreno de la película o serie de televisión, lo que se apreciaba a través de la información verbal, (audio)visual y contextual. Este criterio excluía el cine fantástico, el *chorizo* wéstern o subgénero español de wéstern, los dramas históricos o las adaptaciones de obras literarias ambientadas en otras épocas, subgéneros que cosecharon filmes de gran calidad en la historia del cine español. Por el contrario, antepone los dramas y las (tele)comedias que relatan historias coetáneas, y también los documentales, los docudramas y otros formatos más híbridos, menos representados por su menor producción. Las obras se inscriben en la ficción fílmica y televisiva en su mayor parte, solo un porcentaje reducido corresponde a obras no ficticias (cerca del 7% en la versión actual).

Dentro del conjunto de obras de ficción seleccionadas se incluyen diversas adaptaciones literarias cinematográficas, como *Olvida los tambores* (1975), de Ana Diosdado; *Historias del Kronen* (1995), dirigida por Montxo Armendáriz y basada en la novela homónima de José Ángel Mañas; o *Una palabra tuya* (2008), de Ángeles González-Sinde, inspirada en la obra del mismo título de Elvira Lindo. Entre los documentales, constan *Queridísimos verdugos*, rodada en 1970 pero estrenada en 1977, del director Basilio Martín Patino, *Jóvenes del barrio* (1982) del colectivo Video-Nou y *Yo soy de mi barrio* (2002) de José Vicente Córdoba. En estos documentales se combinan entrevistas y conversaciones-tertulia que tienen lugar mientras realizan actividades gastronómicas, lúdicas o de esparcimiento. Los hablantes entrevistados proceden de áreas dialectales apenas representadas –Badajoz y Sevilla en el caso de los verdugos, el barrio barcelonés de

<sup>17</sup> Más si cabe habiendo compartido producciones en el pasado y siendo pareja en la vida real.

<sup>18</sup> En la versión actual solo cuatro películas de las sesenta y ocho proceden de guiones escritos originariamente en catalán.

<sup>19</sup> En la siguiente fase se explorarán las opciones con otras lenguas cooficiales, como el gallego y el vasco.

<sup>20</sup> Según información facilitada por *El films de la Rambla*, Ventura Pons sería quien tradujera al castellano la mayor parte de sus guiones, entre ellos el de *Puta miseria* (1987), que forma parte de COLFI en la versión actual.

<sup>21</sup> No sucede así en todo el equipo actoral de las películas *La orgía* (1978) o *La quinta del porro* (1980), dirigidas por Francesc Bellmunt, en las que se recurrió a dobladores profesionales en algunos casos.

Canyelles y el barrio madrileño de Vallecas respectivamente— y muchos de ellos emplean el sociolecto bajo. También se han incluido géneros híbridos de drama y comedia que intercalan secuencias de habla auténtica con secuencias narrativas (*Chicas de club* 1972, *To er mundo e güeno* 1982). En la trilogía de comedias que comenzó con *To er mundo e güeno*, el director sevillano Manuel Summers utilizó la técnica de la cámara oculta para grabar las bromas que gastaban una serie de actores y actrices camuflados a ciudadanos anónimos abusando de su bonhomía.

d) *Estética realista*

La película debe enmarcarse dentro de los parámetros estéticos y narrativos realistas. En la ficción filmica y televisiva, los diálogos reproducen en mayor o menor medida un guion previamente escrito a través del canal fónico y en un escenario que imita un espacio real con el fin de construir un mundo ficticio pero verosímil, en el que los personajes hablan y conversan de una forma que se aproxima o asemeja al habla real y cotidiana. Se trata, por tanto, de una «oralidad construida» que produce la «ilusión de realidad». Esta percepción se consigue al seguir las convenciones propias del realismo, al adherirse al «a complex code of what a culture at a given time agrees to accept as plausible, everyday, authentic» (Kozloff 2000: 47). Los intercambios verbales tienden a hacer uso de un lenguaje cotidiano, siguen pautas del comportamiento interaccional y sociopragmático, esto es, los personajes interaccionan como corresponde a su caracterización y a la relación interpersonal que mantienen con otros personajes, haciendo así una analogía de lo que sucedería en el mundo real (cfr. Richardson 2010: 106). Sin duda alguna, esta adhesión al código realista es variable en los distintos géneros, directores y/o guionistas y obras; por este motivo, se intenta detectar movimientos, tendencias y autores más proclives a esta estética (Vid. 3.3.).

e) *Abundante cantidad de diálogos*

El criterio de cantidad debe entenderse en relación con la cantidad de diálogos en una obra audiovisual (cfr. Kozloff 2000: 64-70). En las películas suele haber secuencias con música de fondo, voz en off o sonidos intradiegéticos que pueden extenderse varios minutos. Suele tratarse de secuencias narrativas en las que la acción se revela a través de la imagen, en las que se evidencia la primacía de lo visual en el cine. En el caso de las series, estos espacios suelen limitarse al inicio —con la música de cabecera— o a las transiciones, cuya duración oscila entre unos segundos y, como máximo, un par de minutos. Por tanto, este criterio resulta especialmente relevante en la selección de las películas, en las cuales se procuró que el tiempo de habla representara una proporción significativa de la duración total, entre un 70 % y un 75 %. Este criterio llevó a descartar películas que, *a priori*, podrían haber sido seleccionadas, como la obra neorrealista *Mi tío Jacinto* (1956), cuya escasez de diálogos justificó su exclusión.

### 3.2. Aspectos internos de los diálogos

Los aspectos internos de los diálogos comprenden tanto los parámetros situacionales como las marcas variacionales o estrategias de verbalización. Aunque se toma como base la propuesta del registro coloquial de Briz (1998, 2010), igualmente se han considerado los postulados de Koch y Oesterreicher (2007 [1990]) con respecto a la lengua hablada y a la

oralidad partiendo de la comparación trazada por López Serena (2007b) revisada recientemente (López Serena 2024).

a) *Relación de proximidad entre los interlocutores*

Se repara en las *relaciones entre los interlocutores*, es decir, entre los personajes protagonistas de la historia narrada. Se observa si predominan las relaciones funcionales jerárquicas referidas a entornos profesionales, que generan asimetrías en las interacciones (médico-paciente, profesor-alumno, jefe-empleado, etc.) o si, por el contrario, solo se localizan en secuencias determinadas, siendo mayoritarias las *relaciones funcionales de igualdad*. No se atiende a que pudiera haber una relación social de desigualdad derivada de rasgos sociales individuales –la edad, el sexo o el nivel de instrucción– o de dinámicas de grupo –en las que pesan la experiencia, el conocimiento, el liderazgo–, que originan roles epistémicos o deónticos, que pueden estar asociados a las normas y valores de las comunidades socioculturales (padre/madre-hijo/a, abuelo/a-nieto/a, etc.). Sí se tiene presente que los personajes protagonistas mantengan una *relación vivencial de proximidad*, por la que comparten saberes y experiencias que redundan en el *grado de familiaridad* con el que se comunican (cfr. Briz 1998; 2010; Koch y Oesterreicher 2007 [1990]). Así pues, se prefieren películas y series en las que los personajes protagonistas sean familiares, pareja o amigos/as o desarrolle una relación de cercanía a lo largo de la historia.

b) *Predominio de un marco comunicativo familiar*

El *marco comunicativo* debe ser preferentemente *familiar*, conocido por los hablantes y asociado a actividades cotidianas o no profesionales. A nuestro modo de ver y siguiendo a Koch y Oesterreicher (2007 [1990]), este criterio estaría asociado a las condiciones comunicativas de «grado de anclaje de los actos comunicativo en la situación», «campo referencial» de cercanía (*ergo-hic-nunc*) y de proximidad física de los interlocutores. Ciertamente, muchas secuencias tendrán lugar en las calles, en áreas recreativas o en locales, espacios con menor grado de familiaridad que una vivienda, pero sin el grado de formalidad exigido en una oficina, en un juzgado, etc.

c) *Temática de la vida cotidiana*

Las conversaciones deben versar sobre temas cotidianos y de la vida en general, esto es, no deben abordar, al menos con frecuencia, temas especializados.

La aplicación de los tres primeros criterios conduce a excluir series de televisión centradas en los avatares de la vida profesional como *Anillos de oro* (1983), *Periodistas* (1998) u *Hospital central* (2000), entre otras, y a inclinarse por series en las que desempeñan un papel crucial las relaciones familiares (*Pepa y Pepe* 1995, *Los Serrano* 2003-2008, *Familia* 2013), de amistad (*Al salir de clase* 1999, *Física y Química* 2008, *Por H o por B* 2020) o vecinales (*Aquí no hay quien viva* 2003-2006, *La que se avecina* 2007-). En algunas series televisivas lo laboral y lo familiar se entrelazan para resaltar la doble faceta profesional y personal de algunos de los personajes principales, siendo esta última la que suele cobrar mayor relieve. Así sucede en *Farmacia de Guardia* (1992-1994) o *Lleno por favor* (1993), en las que las acciones transcurren entre el comercio y la trastienda, primando

unas u otras según el episodio. Cuando se trata de oficios manuales o de menor cualificación, el habla podría asumir un tono más informal y podría ser de interés incluir episodios, como se observó, por ejemplo, en *Los camioneros* (1973-74), cuya historia gira alrededor de un camionero, del que se termina conociendo su vida personal, pero con quien sobre todo se comparten viajes transportando mercancías a distintos lugares de España.

En la misma línea, en las películas se intercalan con frecuencia secuencias situadas en ámbitos laborales como parte de la narración de la etapa de la vida de los personajes en la que se centra la historia. Es más, algunas películas neorrealistas reflejan la dificultad de encontrar trabajo o la miseria y degradación que reportan ciertos oficios, como sucede en las historias paralelas de los personajes de *Surcos* (1951), como Pepe, quien se integra en una banda de delincuentes. En ocasiones, justamente el propio entorno profesional se convierte en el escenario por el que circulan personajes marginales. Tal es el caso de *La vida alegre* (1987), cuya protagonista trabaja en un centro de enfermedades de transmisión sexual, o de las secuencias del cine quinqui en las que los policías entrevistan a los delincuentes. En estos contextos, la procedencia de estratos sociales bajos imprime un tono menos formal a los diálogos.

#### d) *Predominio del propósito socializador*

La mayor parte de los diálogos persiguen un supuesto y ficticio fin socializador (cfr. Briz 1998, 2010). En relación con las funciones del diálogo telecinemático<sup>22</sup> (Kozloff 2010; Bednarek 2010), se podría decir que deben prevalecer las funciones asociadas a la comunicación de la narración, esto es, los diálogos se crearon para indicar el anclaje a la diégesis (personajes, espacio, objetos), construir la causalidad de las acciones, caracterizar a los personajes o adherirse al código realista (cfr. Kozloff 2000). Por el contrario, las funciones vinculadas al efecto estético —como el empleo poético del lenguaje con fines humorísticos— no deberían explotarse de forma reiterada, tal como ocurre a menudo en las comedias de situación. Esta orientación resulta especialmente difícil en las comedias en las que a veces se busca provocar la risa por medio de la situación, de la imagen y, también, a menudo, del diálogo. Así, por ejemplo, en la película *¿Qué he hecho yo para merecer esto?* (1984), la abuela termina por imitar el habla cheli de su nieto adolescente en una especie de acomodación intergeneracional (1). En la telecomedia *Farmacia de Guardia*, son habituales los gags entre Lourdes y su exmarido Adolfo (2), recurso que aportaba un toque de humor a su superada separación.

(1) (después de verter Vichy en el vaso)

Abuela: a mí a mí/ ay cómo flipo con las burbujas (4'') ¿a ti no te enrollan? (2'')

(2) Adolfo: Lourdes (1'') te invito a cenar

Lourdes: ¿por qué?/ mejor dicho/ ¿para qué?

Adolfo: pero criatura/ si antes decías/ que pasabas hambre

<sup>22</sup> La nómina propuesta para el diálogo fílmico por Kozloff (2000) fue retomada y adaptada por Bednarek (2010) para la ficción televisiva.

La reiteración de este tipo de intercambios verbales termina por restarles verosimilitud; por ende, una presencia muy acusada conduciría a descartar la obra.

e) *Diálogos - dinámica conversacional*

En cuanto al número de participantes de los diálogos, se prefieren obras audiovisuales con abundantes diálogos en los que participen más de dos interlocutores. En las series tiende a haber repartos corales, que se traslucen con que los personajes aparecen en varias secuencias de forma alternada o intervienen a la vez en determinadas escenas –por ejemplo, en las series juveniles *Al salir de clase* (1997-1999) o *Física y Química* (2008)–. En cambio, en las películas sí suele haber protagonistas claramente definidos, con un rol destacado y una presencia reiterada en los diálogos. En estos casos, se observa si las intervenciones extensas de uno de los protagonistas son demasiado frecuentes, como ocurre en el teatro, situación que puede llevar a descartar la obra. Se prefiere prescindir de películas cuyos protagonistas monopolicen el habla con frecuencia, como sucede en *Celda 2011* (2009) con el personaje de Malamadre, cabecilla de los presos. Se parte de la premisa de que un mayor número de participantes y una distribución más equilibrada de los turnos de habla favorecen el intercambio de roles discursivos (hablante-oyente), reducen la presencia de pares dialógicos y aumentan la frecuencia de intervenciones reactivo-iniciativas, generando así una mayor dinámica conversacional.

f) *Presencia del registro coloquial en cierto grado*

Por lo que se refiere a la variación diafásica, se tiene en cuenta si el habla de los personajes es acorde a las situaciones comunicativas, en particular se pone atención en si hablan de forma coloquial cuando los parámetros corresponden a la inmediatez comunicativa. En el cine producido durante la dictadura de Franco, en especial en el denominado cine de barrio, se advierten notables discordancias de personajes variopintos que hablan con una asombrosa corrección lingüística y con oraciones perfectamente hilvanadas. En algunos casos estas discordancias concuerdan con la adhesión a los ideales del régimen, como se advierte en *La honradez de la cerradura* (1950), dirigida por Luis Escobar, en la que unos jóvenes esposos hablan y se comportan con decoro en cualquier contexto en que se encuentren. Se trata de un cine antirrealista desde el punto de vista lingüístico o muy alejado del ideal de realismo lingüístico. Sin embargo, en el cine disidente de la misma década, menos preocupado por ofrecer modelos de habla sino por reflejar la realidad vital y lingüística, sí se hace uso de vocabulario y construcciones del registro coloquial en cierta medida –*Surcos* (1951), *Felices Pascuas* (1954), *El pisito* (1958), *Los chicos* (1959), *Plácido* (1961), entre otras–.

En la somera revisión del registro de las obras audiovisuales, se atiende sobre todo a la esfera lingüística, no tanto a la conversacional ligada a la planificación o espontaneidad del discurso, que viene a ser la piedra de toque de la oralidad ficticia (cfr. Rossi 2011; Bublitz 2017). En su análisis de las comedias neorrealistas, Rossi (2011: 26-44) detecta que los diálogos cinematográficos difieren sobre todo por la escasa presencia de cambios de tópico, autocorrecciones, repeticiones, interrupciones, solapamientos, reinicios, y la menor presencia de marcadores pragmáticos; según Rossi, «film dialogues present a high degree of coherence, cohesion and conciseness, bearing traces of the (written) screenplay» (Rossi 2011: 26). En la misma línea, Bublitz (2017) señala que los fenómenos verbales y no

verbales ligados a las funciones de planificación, reparación y gestión de turnos son mucho menos frecuentes que en el diálogo espontáneo. Por consiguiente, si una obra audiovisual presenta estos rasgos conversacionales<sup>23</sup> en algunas o varias de sus secuencias se convierte en candidata para ser incluida en el corpus. En los siguientes extractos de diálogos, a los solapamientos, se suman fragmentos pronunciados de forma ininteligible –representados con el doble paréntesis vacío en su interior–, fragmentos de transcripción dudosa –representados con el doble paréntesis–, a modo de susurro –reflejado con los signos º( )º– o de forma enfática. Estas secuencias relatan encuentros casuales entre sus personajes (3, 4), llegadas a locales con muchos bullicio (5) o discusiones vivas entre los jóvenes interlocutores que gritan (6) y se interrumpen como podría suceder en la vida real.

- (3) Chispa: [º(mira qué declaración)º]  
 Luis: [vamos a tomarnos] un vaso Fernando  
 Fernando: ¿te acuerdas la [cantidad de (( ))]? (RISAS)  
 Chispa: [si yo fuera] la federación de ciclismo/ ya verías tú cómo (*Los chicos*, 1959)
- (4) Felipe: [nada nada no es nada]  
 Julia: [gracias/ muchas gracias]/ <obs t="dirigiéndose a Calisto">¿y usted no me va a dar nada para mis pobrecitos? [¿para (( ))]</obs>  
 Julián: [dele/ dele una limosnita] (*Se vende un tranvía*, 1959)
- (5) M3: qué animada va la partida  
 Teresa: [¿qué tal?]  
 H3: [bueno/ venga] vamos (( ))  
 Mariona: no has venido a buscarme  
 Miki: mi padre/ que no me deja en paz  
 H3: [((no no no))]  
 Mariona: [¿por qué no le mandas a paseo?] (*Tuset Street*, 1968)
- (6) Miguel: [(( )) SE OYE] FUERA  
 Rosa: oye/ dijo que iba arreglar los grifos y no los está arreglando  
 Paco: ¡no/ no los estoy arreglando/ por una sencilla razón!/ porque no sé/ [yo soy pintor]  
 Rosa: [¡vaya hombre!]  
 Quique: pues si no sabes te ((lo inventas))/ [porque aquí entró todo el mundo sin saber nada/ y mira cómo está la casa/ ¿vale?]) (*Sentados al borde de la mañana, con los pies colgando*, 1979)

#### g) *Marcas de la variación sociolectal y dialectal*

Se nota como rasgo positivo que el habla refleje la pertenencia a un grupo social o la procedencia geográfica, es decir, la potencial presencia de la variación sociolectal y dialectal. Con respecto a la ficción audiovisual, se ha observado que la presencia de la variación lingüística suele reducirse a ciertos marcadores, léxicos o fonéticos, que se reiteran para caracterizar a los personajes e insertarlos en coordenadas sociales y regionales (cfr. Androusovopoulos 2012; Planchenault 2017). Al respecto, Planchenault (2017: 265) señala que «writers and performers display non-standard phonological and lexical variables (...) in order to index social, regional and ethnic traits as well as associated values». Suele tratarse de una forma estilizada del habla que busca diferenciarse de la variedad estándar y construir a los personajes como representantes de los «Otros» (cfr.

---

<sup>23</sup> Vid. el análisis de episodios de la serie Paquita Salas en Pérez Béjar y Méndez Orense (2022).

Coupland, Thøgersen y Mortensen 2016; Planchenault 2017). Coupland, Thøgersen y Mortensen (2016: 249) consideran que «others' voices are (...) displayed and framed for local, creative, sociolinguistic effect». En general, se hace uso de las construcciones más fácilmente reconocibles para los espectadores; ahora bien, como señala Androusovoulos (2012: 146):

Some literary representations of dialect have been assigned a high degree of accuracy and detail (...). Familiarity with the represented dialect is a relevant factor here, as is the author's or director's representational strategy with its orientation to realism, social-criticism, or deliberate exaggeration.

Aunque la familiaridad con los sociolectos y dialectos sea clave en la calidad de la estilización, en relación con la ficción audiovisual, también incide la tendencia imperante del momento según respalde la lengua estándar o se decante más hacia el realismo lingüístico (cfr. Coupland, Thøgersen y Mortensen 2016; Moreno 2022). Aunque la estética realista y el realismo lingüístico pueden rastrearse a lo largo de la historia del cine y la ficción televisiva (Vid. 3.3.), ha sido en las últimas décadas cuando se ha producido un avance notable, sobre todo en el tratamiento de los dialectos (ver 3.4.).

### **3.3. La estética realista en géneros, movimientos, autores y obras**

La adhesión a la estética realista es variable en los distintos géneros, directores y/o guionistas y obras; por este motivo, en la preselección de las obras audiovisuales se intenta detectar géneros, movimientos y autores más proclives a esta estética. En la escala de mayor realismo con distintos grados y matices, se situarían los documentales cinematográficos o televisivos, seguidos de los docudrama, y pasando a la ficción, los dramas y las comedias de carácter social en el cine o en la televisión –en este último medio se incluirían las comedias de situación y también los programas de tele-realidad– (cfr. Stadler y Mcwilliam 2009: 188; Dynel 2011).

Los distintos movimientos y/o corrientes realistas que han surgido a lo largo de la historia del cine han dado lugar a producciones que se ajustan más a la estética realista al hacer uso de determinadas técnicas que permiten incrementar la dosis de autenticidad. En este sentido, las películas del neorealismo español, muy influenciado por el italiano, no solo trataban de problemas de la sociedad española del momento, como la inmigración del campo a la ciudad, el desempleo, el mercado negro, etc., sino también incorporaban formatos más próximos a la realidad, tales como: escenarios reales, diálogos de mucha frescura en los que los guionistas trataban de reproducir las conversaciones de la calle (*El pisito* 1958, *Los chicos* 1959, *Se vende un tranvía* 1959, *Plácido* 1961), un ritmo de elocución más natural en la interpretación o incluso la superposición de voces, propia de muchas secuencias de las películas de García Berlanga (*Plácido* 1961). No obstante, como advierte Álvarez-Peyreire (2011), incluso en las películas tildadas de realistas se advierte el «sesgo artístico» en la elección ciertos elementos lingüísticos, dicho con sus palabras, «*artistic choices exploiting (and inducing) a secondary semiotic value of linguistic items can be found*» (Álvarez-Peyreire 2011: 57).

El «cine de barrio» abarca películas españolas de carácter comercial —en su mayoría comedias— producidas entre las décadas de 1950, 1960 y 1970, aunque en ocasiones también incluye títulos de los años 80 y 90. Estas obras ofrecen un retrato de la sociedad española de su tiempo, reflejando aspectos como los roles de género, la oposición entre lo

urbano y lo rural, la influencia de la religión o el auge del turismo, al mismo tiempo que profundizan en sus convencionalismos y arquetipos (Huerta y Pérez Morán 2012). Como señala Huerta (2012: 39), «se apuesta por relatos que se alimentan argumentalmente de la actualidad y se desarrollan con un tono generalmente cómico y costumbrista, con unos diálogos coloquiales y una estética inspirada en el clasicismo estadounidense». Ahora bien, debido a la influencia de la ideología de la lengua estándar (Vid. 3.4), los diálogos no siempre presentan el grado de coloquialidad esperable. Por ello, en la conformación del corpus resulta necesario seleccionar aquellos largometrajes que contengan una cantidad significativa de diálogos y cuyos personajes empleen un lenguaje lo más cercano posible al registro coloquial. Pueden incluirse en este grupo algunas películas dirigidas por Pedro Lazaga guionizadas por Pedro Masó como *Los chicos del Preu* (1967) o *Novios 68* (1967), entre otras.

Entre finales de los setenta y principios de los ochenta alcanzó notable éxito el cine quinqui, que incorporó en su reparto a chicos procedentes de barrios marginales, como José Luis Manzano o José Luis Fernández «el Pirri» en las películas dirigidas por Eloy de la Iglesia (*Navajeros*, *El Pico*), y delincuentes, como Ángel Fernández Franco «el Torete» (*Perros callejeros*, *El Torete*). Por otra parte, durante los años de la Transición española, se rodaron bastantes películas con actores y actrices noveles, con escasos recursos y premura de tiempo, circunstancias que potenciaron el efecto realista. Así se aprecia en *Sentados al borde de la mañana, con los pies colgando* (1978), dirigida por Antonio José Betancor –con guion de Javier Moro y protagonizada por Miguel Bosé– y en la emblemática *Tigres de papel* (1977), de Fernando Colomo, elogiada por su naturalidad, que hacía sospechar cierta improvisación por parte de los actores; sin embargo, el director y guionista reconoce que

*Tigres de papel* es de las películas menos improvisadas que he hecho. Todo estaba escrito a excepción de algunas frases de Concha Gregori, que hacía un personaje secundario. Incluso me di cuenta de que el guion era demasiado largo, algo que no le gustaba a Joaquín Hinojosa, porque le quitaba frases. (Tovar 2022)

Una impresión<sup>24</sup> similar causó *La orgía* (1978), guionizada en catalán por Juanjo Puigcorbé y Francesc Bellmunt, aunque fue sobre todo comercializada en español. Al igual que la anterior, la historia partía de vivencias reales, y los diálogos, inspirados en las experiencias de los jóvenes guionistas, reproducían con verosimilitud el habla coloquial –con largas conversaciones o turnos largos, interrupciones, cambios de tópico, etc.–.

En la década ochenta se redujo notablemente la producción española como consecuencia de la Ley Miró; además, fueron motivos recurrentes la recuperación de la historia –centrada en la República, la Guerra civil y la posguerra– y las adaptaciones literarias (Riambau 2023). Aun así, contamos con un repertorio modesto de dramas sociales y más cuantioso de comedias —críticas con la España profunda (*Qué he hecho yo para merecer esto* 1984), urbanas o asociadas a la *Movida madrileña* (*A tope* 1984)— que permite acercarse y escuchar el habla de la época o algunas de sus variedades. Se incluyen aquí las películas de Manuel Summers, tanto la trilogía rodada con cámara oculta —*To er mundo é güeno* (1982), *To e'r mundo é i'mejó!* (1982), *To e'r mundo é ¡demasiado!*— como las películas sobre el grupo de música de Hombres G —*Sufre mamón* (1987) o *Suéltate el pelo* (1988)—; en estas últimas se pretendía reproducir el modo de hablar y los diálogos de los jóvenes integrantes del grupo.

<sup>24</sup> Exceptuando el doblaje, particularmente de algunos actores, que resta naturalidad.

En la década de los noventa continúan engrosando su filmografía directores de generaciones anteriores, a la par que comienzan su trayectoria un número muy nutrido de directores y directoras, cifra que se acrecentará en el siglo XXI. Aunque la comedia urbana seguirá ocupando un lugar indiscutible en el cine español –con filmes de Miguel Albaladejo, Manuel Gómez-Pereira, Joaquín Oristrell o Álex de la Iglesia, entre otros–, el drama social se consolida como uno de los géneros más representativos de las décadas posteriores. Su vocación de reflejar la realidad social convierte a muchas de estas producciones en candidatas para formar parte del corpus. Este repertorio comprende películas como *Historias del Kronen* (1995) de Montxo Armendáriz, *La buena estrella* (1997) de Ricardo Franco, *Barrio* (1998) o *Los lunes al sol* (2002) de Fernando León de Aranoa, *Solas* (1999) de Benito Zambrano, *El Bola* (2002) de Achero Mañas, *Te doy mis ojos* (2003) de Icíar Bollaín, *Siete bolas de billar francés* (2007) Gracia Querejeta, entre otras. El relato fílmico de estas obras se sustenta en buena medida en los diálogos, que se combinan con los sonidos ambientales, con la música de fondo y, de forma especial, con las imágenes, con las que se potencia la tensión dramática.

### 3.4. Realismo lingüístico, representación de la diversidad lingüística y mimesis de variedades

En las décadas cincuenta y sesenta, predominan las películas españolas en las que personajes de clases populares que cometan delitos emplean un estilo de habla estándar que se diferencia poco del de otros personajes, como se observa en *Los ladrones somos gente honrada* (1956), *Los tramposos* (1959) o *Atraco a las tres* (1963). No sucede así en las películas del género quinqui de finales de los setenta y de los ochenta, en las que se advierte una estilización del argot<sup>25</sup> juvenil marginal o de la delincuencia (*Perros callejeros* 1977; *Navajeros* 1980, *Los últimos golpes del Torete* 1980; *Colegas* 1982; *El Pico* 1983, *Perras callejeras* 1985, etc.). De este género, destacan los directores y guionistas José Antonio de la Loma y Eloy de la Iglesia, que llevaron a cabo un notable trabajo de campo que culminó con la incorporación de jóvenes o delincuentes de barrios marginales en el reparto de actores (Vid. 3.3.). Esta tendencia se mantendrá en las producciones audiovisuales de las décadas posteriores, aunque el grado de realismo lingüístico o de representación del argot variará según el propósito de crítica o denuncia social, la competencia del guionista con respecto al argot y la participación de actores no profesionales. Un ejemplo especialmente logrado de esta tendencia se encuentra en *El truco del manco* (2008), una película que destaca por su aproximación realista al habla de los sectores marginales urbanos y por la autenticidad que transmite en la caracterización de sus personajes.

La diversidad lingüística corre pareja a la presencia de otros idiomas y, en particular, de los dialectos. En las obras audiovisuales, se fueron utilizando algunos extranjerismos, especialmente anglicismos, conforme se iban adoptando y extendiendo en el idioma (*show*, *hándicap*, *footing*, etc.), sobre todo en áreas como la música, como se aprecia en *A 45 revoluciones por minuto* (1969). Con frecuencia se han empleado para caracterizar a jóvenes estudiantes, más proclives a su uso (*Los chicos del Preu* 1967) y/o de clase alta (en la serie *Tres eran tres* 1974). En contraste, su presencia en las series de televisión

<sup>25</sup> La censura obstaculizó la transmisión del argot y de otras esferas del léxico por su visión moralizante.

contemporáneas resulta mucho más marcada —como ocurre en *La que se avecina*—, en sintonía con las tendencias actuales vinculadas a los procesos de globalización.

Menos habituales eran los cambios de código en las intervenciones, los cuales servían para caracterizar a los personajes extranjeros —como en *El verdugo* (1963) o *Abuelo made in Spain* (1968)— o a personajes españoles de perfil snob, como Mery en *El pisito* (1958). Este tratamiento contrasta con el de producciones más recientes, que muestran a españoles comunicándose en otros idiomas con normalidad —como las profesionales residentes en Alemania en *El olivo* (2016)—, en consonancia con la realidad sociolingüística actual.

Peor suerte corrieron las lenguas cooficiales, prácticamente ausentes en las películas anteriores a finales de los setenta, en aras de preservar la hegemonía del español como lengua nacional, ideal de la dictadura franquista. Su modesta presencia comenzará de la mano de directores vascos<sup>26</sup> y catalanes quienes lo emplearán para caracterizar personajes y situaciones recurriendo a voces y expresiones más o menos comunes, como el vasco en *El Pico* (1983) de Eloy de la Iglesia o *La muerte de Mikel* (1984) de Imanol Uribe, o el valenciano en *Puta miseria* (1987) de Ventura Pons, doblada al español. Décadas más tarde, siguen sin reproducirse apenas diálogos de bilingüismo real, salvo algunas excepciones —por ejemplo, en *As bestas* 2022—.

En cuanto a la representación dialectal, se impuso la lengua estándar en la producción española<sup>27</sup>, al igual que ocurrió en otros países europeos imbuidos por la ideología de la lengua estándar (Coupland, Thøgersen y Mortensen 2016). Se tendía a representar el habla urbana y correcta como modelo a imitar. El lenguaje rural, sin embargo, se usa estilizado para representar al ingenuo o cateto de pueblo creando así un efecto cómico por su falta de habilidades o por contraste con otros personajes más resueltos. Personajes como Orencio, interpretado por José Luis Ozores en *Ahí va un recluta* (1960), o Saturnino, encarnado por José Sacristán en *Novios 68* (1967), ilustran este tipo de habla caracterizada por el uso de apócope, junto a una articulación y pronunciación particular, rasgos con los que singularizan y ridiculizan estos personajes. Sin duda, el máximo exponente de esta estilización lingüística puede observarse en los personajes interpretados por Paco Martínez Soria en *La ciudad no es para mí* (1967) o *Abuelo made in Spain* (1969), cuyas intervenciones reproducen un aragonés estilizado no exento de significado ideológico sobre lo rural (7). Efectivamente, «a speech style is always an abstraction from the dynamic process of styling language, and that it is always freighted with socially interpretive, metapragmatic meaning» (Coupland, Thøgersen y Mortensen 2016: 33).

(7) Marcelino: es que (2") <obs t="se refiere al cordero">(1")¿a'(dónde vas tú lucerico?</obs> (RISAS) (2") ¿y su madre? (*Abuelo made in Spain* 1969)

En las últimas décadas del siglo XX, fueron entrando en escena personajes procedentes de distintas zonas del territorio español no tanto con una intención

---

<sup>26</sup> Vid. Martínez (2023) para una revisión de la presencia del euskera en los diálogos de las películas producidas en el País Vasco en los años ochenta, las cuales optaron por tratamientos distintos con el fin de reflejar la realidad lingüística.

<sup>27</sup> Refiriéndose a las producciones americanas en español, Moreno (2022) señala que desde los sesenta se dejó de hacer uso de un acento más neutro para mostrar acentos más locales a fin de lograr cierta autenticidad.

humorística sino antes bien con el fin de representar otras variedades dialectales. Así ocurre en *Farmacia de Guardia* con la sevillana Queen; será su pronunciación y algunos rasgos léxicos los que indicarán su origen mediante un estilo dialectal de nuevo simplificado. En este caso habría una organización asimétrica de los personajes, «protagonists typically align to mainstream legitimate language, and code-contrasts come to accentuate, and often iconise character contrasts» (Androutsopoulos 2012: 148). En otras series parece haber mayor equilibrio debido a que los personajes de habla dialectal son precisamente protagonistas, como vemos en *Chicas de hoy en día*, serie en la que las protagonistas, Nuri y Charo residen en Madrid, pero proceden de Barcelona y Sevilla respectivamente.

El realismo lingüístico de corte dialectal alcanza una cota más alta en las series y películas ambientadas en zonas geográficas marcadas dialectalmente y menos habituales en la pantalla. Estas obras se han incrementado en la última década. Con respecto al andaluz, y citando solo algunos largometrajes, pueden mencionarse *El Niño* (2014), *Techo y comida* (2015), *Mi querida cofradía* (2018), *Mi gran despedida* (2020) o *Sevillanas de Brooklyn* (2021). La mayor presencia de dialectos ha venido acompañada de una representación más verosímil: bien los personajes son interpretados por actores o actrices originarios de la zona, bien reciben el asesoramiento de especialistas que les ayudan a reproducir con mayor fidelidad el acento correspondiente (cfr. Moreno 2022), como en el caso de la andaluza *Malaka* (2019) o de la gallega *El desorden que dejas* (2020). Este cambio de paradigma fue más temprano en el español de América utilizado en las producciones españolas desde las últimas décadas del siglo XX, cuando se comenzó a optar por interpretaciones más auténticas a cargo de nativos, sin necesidad de imitar acentos. Además, se ha ampliado de forma notable la representación de personajes procedentes de distintos países de América –ya no son solo de origen argentino–, en consonancia con el flujo migratorio. Se refleja bien en *Flores de otro mundo* (1999), con los personajes de Patricia y Milady, interpretados por la dominicana Lissette Mejía y la cubana Marilyn Torres.

En definitiva, como señala Moreno (2022), el paradigma de la «ideología de la autenticidad» no solo sigue vigente sino que está en auge, como muestra la mayor presencia de personajes con perfiles dialectales y, además, una más cuidada estilización y pronunciación.

#### 4. METADATOS DE OBRAS AUDIOVISUALES Y DE LOS PERSONAJES

Cada archivo va acompañado de una ficha técnica que está organizada en dos partes. La primera parte contiene metadatos de las obras audiovisuales, a saber: el código del archivo, el tipo de formato audiovisual, la fecha de estreno, el título de la película, el número y título del episodio, el/la director/a, el//la guionista, el género, los temas principales de la película, la duración en minutos y modos de acceso para su visionado. En la Figura 2 aparecen en negrita aquellos campos que constituyen motores de búsqueda, mientras que están escritos sin negrita los que solo se ofrecen en la cabecera.

Archivo	Formato audiovisual	Fecha	Título	Número del episodio	Título del episodio
Director/a/es/as	Guionista/s	Género	Temas	Duración	Acceso

Figura 2. Ficha de los metadatos de la obra

El nombre del archivo sirve para identificar la obra audiovisual, sea película o serie televisiva (P o S) y el año del estreno, por ejemplo 1P1951. Para la anotación del género, se ha consultado la página web de *Filmfaffinity*, una web que proporciona la ficha técnica de películas y series de televisión, y se revisa conforme a la clasificación propuesta por Sánchez Noriega (2002). Por último, se decidió incluir el modo de acceso (DVD o plataforma) para facilitar su localización y visionado por parte del usuario; de esta manera, aunque el corpus no ofrezca las secuencias audiovisuales, el usuario podría acceder a ellas si fuera de su interés o quisiera analizarlas.

La segunda parte corresponde a los metadatos de los personajes, que ofrecen una caracterización aproximada de las variables sociales y geográficas, de su relación con otros personajes con los que mantienen muchos intercambios comunicativos y del actor o actriz que los ha interpretado (Vid. Figura 3).

nombre del personaje	
sexo	hombre / mujer
grupo de edad	0-12   13-17   18-34   35-64   >65
nivel sociocultural	alto / medio / bajo
ocupación	tipo de ocupación <i>profesión específica</i>
	tipo de zona (urbana, semiurbana, periurbana, rural)
	provincia
zona geográfica	país
relación entre personajes	
actor/actriz	

Figura 3. Ficha de los metadatos de los personajes

Los campos concernientes a las variables sociales y geográficas, a excepción de la profesión específica, pueden ser filtrados en las búsquedas; su resultado cuantitativo también se proporciona en las búsquedas de *frecuencia completa*. Los datos de la relación entre personajes o del equipo actoral solo se muestran en la cabecera que acompaña el contexto de la concordancia (Vid. Figura 4). La relación entre personajes da cuenta de las relaciones familiares y de los lazos de amistad de los personajes protagonistas y secundarios, o en su defecto, se anota que se trata de desconocidos. El reparto de actores y actrices se comprueba en los créditos o en otras fuentes disponibles.

Archivo: IPI951	Formato audiovisual: película	Fecha: 1951
Título: Surcos	Número de episodio: no aplica	Título de episodio: no aplica
Director: José Antonio Nieves Conde	Guionista: José Antonio Nieves Conde, Gonzalo Torrente Ballester, Natividad Zaro. Idea: Eugenio Montes	Género: drama social
Temas: posguerra, inmigración, pobreza	Duración: 99	Acceso: FlixOlé
Nombre: Pili	Sexo: mujer	Grupo de edad: 18-34
Nivel sociocultural: bajo	Tipo de ocupación: trabajadores no cualificados	Profesión: vendedora ambulante
Tipo de zona geográfica: urbana	Provincia: Madrid	País: España
Relación entre personajes: pariente y novia de Pepe, hija de Engracia	Actor/actriz: María Asquerino	

Figura 4. Cabecera ofrecida en la parte inferior al contexto de la concordancia

Las variables sociales y geográficas se obtienen de las escenas en las que se desarrollan los filmes o series, esto es, atendiendo a aspectos de los personajes –su apariencia, a qué se dedican, dónde viven, cómo visten, etc.– y del lugar en el que transcurren las acciones –menciones de la ciudad, edificios emblemáticos, matrículas de los coches, etc.–. En todos los campos se ha contado con la opción de «no indicado» cuando no se disponía de indicios suficientes, lo que ocurre sobre todo con personajes que tienen una aparición puntual y son interpretados por actores de reparto.

Varios campos requieren clasificaciones de las variables, entre ellos el grupo etario, el nivel sociocultural, el tipo de profesión o el tipo de zona geográfica. Se intentó tomar clasificaciones empleadas en otros corpus y adaptarlas si era necesario. Para el grupo de edad, resultaba especialmente complicado porque los datos observables suelen ser si estudia en el colegio, en el instituto o en la universidad, si trabaja o está jubilado/a, así como las relaciones familiares (madre, padre, hijo/a, nieto/a, etc.) y la apariencia física. Esta limitación condujo a establecer categorías solo parcialmente comunes a otros corpus: 0-12 | 13-17 | 18-34 | 35-64 | >65.

Se descartó incluir el nivel de estudios por ser una categoría bastante restringida que precisa de información específica y se optó por anotar el nivel sociocultural, categoría más amplia y sincrética, cuya asignación combina datos observables como la profesión, el barrio, la vivienda, la indumentaria, el ocio y la forma de hablar. Se distinguió entre nivel alto, medio y bajo. Para el tipo de ocupación, se recurrió a la clasificación nacional de ocupaciones establecida por el INE (BOE, n.º 306, 10/12/2010). Se mantuvieron la mayor parte de las categorías, se incorporó la distinción de directivos según el grado de responsabilidad (1, 4 y 5 en el listado) –procedente de PRESEA– y se añadieron tipos de no ocupación, como estudiantes, jubilados/as o sin ocupación –aquí caben tanto los/as parados/as como las «amas de casa» tradicionales–.

1. *Altos cargos*: «directores de empresas y de Administración Pública
2. *Profesionales universitarios*: «técnicos y profesionales científicos e intelectuales»
3. *Artistas creativos e interpretativos*
4. *Gerentes y medianos empresarios*
5. *Técnicos de apoyo, pequeños empresarios*: «técnicos y profesionales de apoyo»
6. *Trabajadores en restauración, comercios*
7. *Trabajadores en protección y seguridad*

8. *Obreros especializados*: «trabajadores cualificados en agricultura y pesca», «trabajadores cualificados en industrias», «operadores y montadores de instalaciones»

9. *Trabajadores no cualificados*

+ 10. *Estudiantes 11. Jubilados 12. Sin ocupación 13. No indicado*

Con respecto a la zona geográfica, se distingue según procedan de grandes ciudades (urbana), ciudades o capitales pequeñas de provincias (semiurbana), barrios periféricos marginales (periurbana) o pueblos y aldeas (rural). Este dato resultaba fácil de incorporar y podía orientar la interpretación de los resultados en casos particulares, como el habla de los delincuentes en el cine quinqui o el estereotipo de personajes rurales, entre otros. La mayor parte de los personajes proceden de zonas urbanas en consonancia con la ambientación de las películas y las series de televisión.

También se anotaron el país y, en el caso de España, las provincias. Los datos referidos al país corren parejos a la tasa de movilidad internacional, que era esporádica y vacacional en los años sesenta y setenta, permanente aunque minoritaria desde finales de los noventa, y más asentada y común en las últimas décadas. En cuanto a las provincias, un porcentaje muy alto de películas y series de televisión están situadas en Madrid; por ello, podría ser de interés revisar si el habla de otras provincias está caracterizada de forma realista o, sin embargo, como parece previsible, se emplea un español europeo neutro con una mínima diferenciación dialectal –salvo en algunas producciones de las últimas décadas–.

Al tratarse de personajes de ficción, los datos de variables sociales y geográficas son orientativos, por una parte, y no auténticos, por otra. No servirían para llevar a cabo estudios sociolingüísticos al uso; antes bien, permiten examinar cómo se han caracterizado los personajes (cfr. Culpeper 2001; Bednarek 2010, 2023), es decir, qué marcas lingüísticas han seleccionado los guionistas por ser más representativas de un grupo social o generación, o qué estereotipos están transmitiendo o consolidando. Se explota el habla como un recurso, en combinación con la apariencia y el comportamiento, del que se sirven para presentar a personajes verosímiles que recuerden a personas de la vida real. Al respecto, Dynel (2011) señala que

Repetitive patterns in people's verbal and nonverbal expression lead to stereotypes or simply prevailing means of communication, which foster expectations as to how people communicate, e.g. depending on their gender and social background. These default conventions should be mirrored in contemporary film discourse to promote viewers' perception of reality on the screen. Therefore, the language of popular culture texts may also reinforce stereotypical beliefs and attitudes. (Dynel 2011: 47)

Con respecto a la perspectiva dialectológica, dada la estandarización habitual en la ficción fílmica y televisiva (Vid. 3.4.), y el predominio del lenguaje urbano, los datos geográficos se prestan a limitados fines, tales como estudiar los estereotipos de personajes rurales o extranjeros (cfr. Planchenault 2017), explorar el argot de la delincuencia en las películas de cine quinqui (personajes que habitan en zonas periurbanas) o comprobar si el fenómeno lingüístico era percibido como eminentemente urbano (frecuencia alta en zonas urbanas). A pesar del fenómeno de normalización lingüística experimentado en las últimas décadas (cfr. Briz 2022; Autor 2024), varias producciones cinematográficas y televisivas recurren a la diversidad dialectal como una vía para dotar de mayorrealismo sus historias

(cfr. Moreno 2022; Androutsopoulos 2012) (Vid. 3.4.). Las obras seguidoras de esta corriente podrían ser de mayor utilidad para estudios dialectales, tomando las reservas pertinentes.

## 5. DE LA TRANSCRIPCIÓN Y ETIQUETADO A LA BÚSQUEDA DE UNIDADES

Se puede recurrir a distintos materiales en bruto y procedimientos para obtener transcripciones de películas y series de televisión, tales como guiones, subtítulos, subtítulos para sordos, transcripción manual y, recientemente, transcripción automática (opción que reportaba bastantes fallos cuando se inició la recopilación de los archivos). Sea cual sea el material de base (guiones o subtítulos), se precisa una revisión para comprobar la correspondencia del texto escrito con respecto al audio de la obra representada (Vid. 2.4.). Después de explorar las distintas opciones (dificultades de acceso, inconvenientes para su procesamiento y/o coste), se tomó la decisión de transcribir los archivos de forma manual. En la actualidad no se realiza la transcripción desde cero sino tomando como base la transcripción automática obtenida con el programa Whisper.

En la transcripción, se emplean todos los signos del sistema de transcripción del Grupo Val.Es.Co. (Briz et al. 2002), salvo el signo de sucesión inmediata (§) y el de la aspiración (H), que no se han marcado (Vid. Tabla 1), y las intervenciones de los hablantes, que se han indicado de otra forma.

H:	<i>Intervención de un hablante</i>
[ ]	solapamientos
/	pausas fónicas menores a un segundo
-	autointerrupciones y reinicios
°( )°	fragmento pronunciado con intensidad baja
((siempre))	transcripción dudosa
(( ))	fragmento indescifrabla
(2")	intervalo de segundos o minutos / sin etiquetar corresponde a silencio
pa(ra)	reconstrucción de una unidad léxica pronunciada de forma incompleta
(en)tonces	ídem
pa(ra)'(e)l	reconstrucción de fenómenos de fonética sintáctica entre palabras
PESADO	pronunciación marcada o enfática
(RISAS)	momentos puntuales que aparecen antes o después de los enunciados
(TOSES)	ídem
(GRITOS)	ídem
aa	alargamientos vocálicos (si se produce en una vocal con tilde, la primera vocal llevará la tilde)
ss	alargamientos consonánticos (salvo el caso de los dígrafos <i>ll</i> , <i>rr</i> , y los grupos consonánticos <i>nn</i> y <i>cc</i> )
¿?	interrogaciones
¡!	exclamaciones
cursiva	reproducción o imitación de emisiones de estilo directo

Tabla 1. Signos de transcripción de Val.Es.Co.

Como se puede observar, se indican fenómenos propios de la oralidad, como solapamientos, reinicios, alargamientos o susurros, y se reduce el uso de signos ortográficos, como el de la mayúscula a las siglas y pronunciaciones enfáticas, y la mayúscula inicial, a los nombres propios.

Para la identificación de los hablantes, se anota el nombre propio del personaje utilizado por otros personajes («Mariana», «Juan»), se añade la forma de tratamiento («Don Pedro», «Doña María») o se emplea la profesión que desempeña o el cargo que ocupa («Comisario») si los personajes se dirigen habitualmente de este modo. En el caso de los personajes ocasionales cuyo nombre se desconoce, se emplea la identificación<sup>28</sup> de H (hombre) o M (mujer) seguida de un número; del mismo modo se procederá en los documentales en los que se entreviste a ciudadanos desconocidos o no identificados (H1, H2, M1, M2, M3...).

Asimismo, la transcripción en Word incorpora cinco etiquetas dobles señaladas en el Protocolo de trabajo de equipos del corpus Ameresco, basadas en la propuesta de PRESEEA (2008), que presenta un formato compatible con el lenguaje XML (Carcelén y Uclés 2019). Las etiquetas de los extranjerismos y de las siglas portan atributos entrecerrillados en los que se anota la pronunciación que se escucha: <extranjero t="lait">light</extranjero> o <siglas t="equis te">XT</siglas>.

<cita> </cita>	Reproduce el estilo directo marcado también en cursiva
<entre_risas> </entre_risas>	Pronunciado entre risas
<sic> </sic>	Marca un error fonético o gramatical
<extranjero t=" " > </extranjero>	Señala el uso de extranjerismos
<siglas t=" " > </siglas>	Marca el uso de siglas

Tabla 2. Etiquetas dobles utilizadas

Al tratarse de obras audiovisuales, se estimó pertinente añadir dos etiquetas más para marcar información multimodal o contextual, atributos que aparecen entrecerrillados.

1) *Etiqueta de observación simple*: recoge información multimodal o contextual relevante para entender la intervención de un hablante. Se han incluido en esta etiqueta las onomatopeyas. Los ejemplos de (8) han sido tomados de diferentes obras audiovisuales.

- (8) Miguel: <obs t="asiente">mm</obs>  
 Rosario: <obs t="habla por teléfono">¿sí?/ ah Cosme/ no no no no dime/ ¿qué?</obs>  
 Juan: <obs t="refiriéndose al reloj">dámelo</obs>  
 Carlos: <obs t="ruega silencio">¡chiss!</obs>

También señala la presencia de música, ruidos u otros sonidos.

<obs t="música de fondo">(2'30')</obs>  
<obs t="suena el timbre de la puerta">(1')</obs>

2) *Etiqueta de observación compleja*: se emplea para indicar información que afecta a más de una intervención, a saber: acciones simultáneas al habla, el espacio en el que transcurre la acción y/o sonidos de fondo que pudiera haber (tráfico, canciones o música de fondo.).

<obs\_inicio t="jugando a las cartas"/>  
<obs\_fin/>

<sup>28</sup> El recurso de utilizar H y M enumerados para distinguir hombres y mujeres se tomó del corpus MESA.

Cuando hay más de una observación compleja, se utiliza un identificador para cada una.

- (9) <obs\_inicio id="n20" t="en el salón comedor cenando"/> <obs\_inicio id="n21" t="varias conversaciones a la vez"/>  
 Sergi: toda la vida juntos  
 Alfonso: ¿tus padres?  
 Sergi: sí  
 H5: Alfonso/ <obs\_fin ref="n21"/> Alfon- -Alfonso ¿tú a qué te dedicas? (8 citas 2008)

Las etiquetas dobles, incluidas las de observaciones, se visualizan por medio de pequeños cuadros de notas al poner el cursor en la intervención coloreada de amarillo –etiquetada– (Vid. Figura 5).

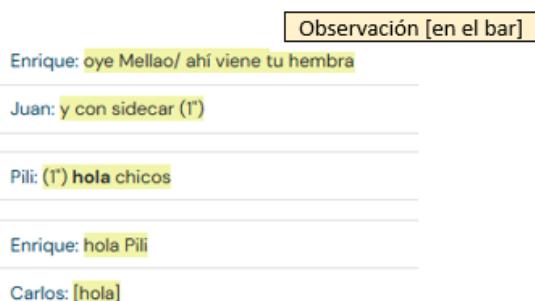


Figura 5. Contexto de la palabra buscada con la etiqueta.

Los archivos transcritos y etiquetados son validados por el equipo de informáticos, que notifican los errores detectados para su correcto procesamiento. Una vez se han corregido los archivos, se vuelcan en la base de datos<sup>29</sup> de la página web y se realizan búsquedas para comprobar su funcionamiento. El formulario de búsqueda<sup>30</sup> incluye diferentes selectores y campos agrupados en siete áreas que actúan sobre la búsqueda de diferentes formas (Vid. Figura 6).

<sup>29</sup> En la versión actual consta de 95 documentos que computan por 752 165 palabras ortográficas, sin incluir las palabras de idiomas distintos del español y las palabras cortadas. De los 95 documentos, 68 corresponden a películas y 27 a episodios de series de televisión, que ascienden a 588 071 y 164 094 respectivamente.

<sup>30</sup> La página web incluye una Guía de consulta del Corpus en la que se explican en detalle las áreas de búsqueda con las opciones que ofrecen.

Figura 6. Interfaz de la Búsqueda en COLFI

## 6. CONSIDERACIONES FINALES

En los apartados anteriores se han explicado los criterios de selección de las obras audiovisuales que integran COLFI, se han descrito los metadatos de las obras y de los personajes que han configurado la base de datos y se han indicado las convenciones de transcripción que se han implementado. En este último apartado se reflexiona acerca de la idiosincrasia de los diálogos telecinemáticos para poner de relieve las ventajas y limitaciones que presenta el corpus.

Los diálogos telecinemáticos se caracterizan por una dualidad intrínseca: por un lado, mimetizan las conversaciones auténticas para dotar de verosimilitud el relato; por otro, se producen en unas condiciones muy diferentes, determinadas por la planificación o ausencia de espontaneidad y por su finalidad de construir o recrear una historia ficticia destinada a una audiencia (cfr. López Serena 2007a; Álvarez-Peyre 2011; Rossi 2011; Bublitz 2017; Jucker 2021). Esta dualidad origina una serie de semejanzas y diferencias que conviene tener presentes al utilizar estas obras como material de análisis.

Su voluntad de simular diálogos reales conduce a emplear construcciones idiomáticas en contextos pragmáticos aceptables. Las elecciones lingüísticas de los guionistas ponen de relieve con frecuencia usos y posibilidades del lenguaje (cfr. Álvarez-Peyre 2011). Estas ocurrencias resultan cruciales para aproximarse al lenguaje de las décadas anteriores a los noventa, de las que no se dispone de corpus conversacionales de materiales auténticos, puesto que pueden posibilitar reconstrucciones sobre diversos usos o cambios lingüísticos<sup>31</sup>, que se deberán interpretar con la ayuda de fuentes metalingüísticas o completar con otros métodos de investigación (cfr. Llopis y Jansegers 2024; Llopis y Pons 2024; Llopis 2025). Asimismo, con respecto a la sincronía, como advierte Álvarez-Peyre

<sup>31</sup> En varios trabajos sobre cambio y difusión se ha utilizado COLFI bien de forma complementaria junto a otros corpus (Autor 2024 (ed.), Autor en prensa), bien como corpus de estudio único (Autor 2025a).

(2011: 66-67), estas muestras proporcionan «a pool of linguistic structures acceptable for theoretical research», «good specimens for the study of the system of language and its use», las cuales «may be used to formulate quantitative hypotheses, especially regarding linguistic structures». Además, las obras audiovisuales ponen a disposición una mayor variedad de contextos, muchos de los cuales no se documentan en las entrevistas y conversaciones grabadas; por tanto, proporcionan un repertorio más completo de las funciones pragmáticas y comunicativas. Incluso se ha comprobado que los recursos lingüísticos con funciones expresivas se utilizan con mayor frecuencia, pues constituyen mecanismos especialmente eficaces para transmitir emociones, intensificar la tensión dramática en determinadas escenas y caracterizar a los personajes (cfr. Quaglio 2009 sobre los intensificadores en *Friends*; cfr. López Serena 2007a).

En cambio, la dinámica conversacional pierde autenticidad por la escasa presencia de solapamientos, interrupciones, cambios de tópico, autocorrecciones, titubeos, etc. La reproducción de un guion previo, así como la intención de que el discurso sea inteligible y no genere tedio o desatención conducen a limar estos fenómenos de la comunicación espontánea (cfr. López Serena 2007a; Rossi 2011; Bublitz 2017). Estos rasgos conversacionales, unidos a los fónicos y prosódicos presentes en la actuación o en la postsincronización, conforman un lenguaje cinematográfico propio que no se puede tomar como copia fiel del habla real (cfr. Kozloff 2000; Rossi 2011).

Aun reconociendo estas limitaciones de la mimesis, sostendemos que los diálogos telecinemáticos no deben entenderse como un bloque homogéneo, puesto que algunos géneros, corrientes, autores y obras se aproximan más a los parámetros situacionales y a las estrategias verbales propias de la inmediatez comunicativa, mientras que otros se alejan notablemente. En COLFI se intenta identificar e incorporar aquellas producciones que mejor simulan la conversación coloquial, sin obviar que son recreaciones ficticias.

## REFERENCIAS BIBLIOGRÁFICAS

- ALVAREZ-PEREYRE, Michael (2011): «Using film as linguistic specimen: Theoretical and practical issues», en Roberta Piazza, Monika Bednarek y Fabio Rossi (eds.), *Telecinematic Discourse: Approaches to the Language of Films and Television Series*. Amsterdam: John Benjamins (Pragmatics & Beyond New Series 211), pp. 47–67.
- AMERESCO = Marta Albelda Marco, María Estellés Arguedas (eds.) (2010-): *AMERESCO. América y España. Español coloquial*. <https://esvaratenuacion.es/ameresco>.
- ANDROUTSOPoulos, Jannis (2012): «Introduction: Language and society in cinematic discourse», *Multilingua*, 31, 2-3, pp. 139-154.
- BEDNAREK, Monika A. (2010): *The Language of Fictional Television: Drama and Identity*. London: Continuum.
- BEDNAREK, Monika A. (2023): *Language and Characterisation in Television Series. A corpus-informed approach to the construction of social identity in the media*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- BERTOLOTTI, Virginia, Concepción COMPANY (2024): «Corpus históricos del español. Avances y tareas pendientes». *Studia linguistica romanica*, 12, pp. 1-17. <https://doi.org/10.25364/19.2024.12.1>.
- BRIZ GÓMEZ, Antonio (1998): *El español coloquial en la conversación. Esbozo de pragmagramática*. Barcelona: Ariel.
- BRIZ GÓMEZ, Antonio (2010): «El registro como centro de la variedad situacional. Esbozo de la propuesta del grupo Val.Es.Co. sobre las variedades diafásicas», en Irene Fonte Zarabozo y Lidia

- Rodríguez Alfano (eds.), *Perspectivas dialógicas en estudios del lenguaje*. México: Editorial de la Universidad Autónoma Metropolitana-Iztapalapa, pp. 21-54.
- BRIZ GÓMEZ, Antonio (2022): «Tres hipótesis explicativas del cambio dialectal: el caso del habla de la comarca de Requena-Utiel (Valencia, España)», en Carmen Díaz (ed.), *Studia philologica in honorem José Antonio Samper*. Madrid: Arco-Libros, pp. 173-184.
- BRIZ GÓMEZ, Antonio y Marta Albelda Marco (2009): «Estado actual de los corpus de lengua española hablada y escrita: I+D», en Instituto Cervantes (ed.), *El español en el mundo: anuario del Instituto Cervantes*. Madrid: Instituto Cervantes, pp. 165-226.
- BRIZ, Antonio y Andrea CARCELÉN GUERRERO (2019): «El futuro iberoamericano del español: la investigación del español oral y en español», en Instituto Cervantes (ed.), *El español en el mundo: anuario del Instituto Cervantes*. Madrid: Instituto Cervantes, pp. 189-217.
- BRUMME, Jenny (2008a): *La oralidad fingida: obras literarias: descripción y traducción*. Madrid: Iberoamericana/Francfort del Meno.
- BRUMME, Jenny (2008b): *La oralidad fingida: descripción y traducción. Teatro, cómic y medios audiovisuales*. Madrid: Iberoamericana/Francfort del Meno.
- BUBLITZ, Wolfram (2017): «Oral features in fiction», en Miriam A. Locher y Andreas H. Jucker (eds.), *Pragmatics of Fiction*. Berlin: De Gruyter, pp. 235–264.
- CAES = Guillermo Rojo, Ignacio Palacios (eds.) (2022): *Corpus de aprendices de español L2*. <http://galvan.usc.es/caes>.
- CALDERÓN CAMPOS, Miguel y Gael VAAMONDE DOS SANTOS, G. (2020): «Oralia diacrónica del español: un nuevo corpus de la edad moderna», *Scriptum Digital*, 9, pp. 167–189. <https://doi.org/10.5565/rev/scriptum.108>
- CANO AGUILAR, Rafael (2012): «Diálogo y oralidad ficticia en las Crónicas medievales», en Victoria Béguelin-Argimón, Gabriela Cardone y Mariela de la Torre (eds.), *En pos de la palabra viva: huellas de la oralidad en textos antiguos. Estudios en honor al profesor Rolf Eberenz*. Bern: Peter Lang. pp. 351-370.
- CARCELÉN, A. y G. UCLÉS (2019): «Diseño y construcción de un corpus oral multidialectal. El corpus Ameresco», *Normas*, 9, pp. 17-36. doi: <https://doi.org/10.7203/Normas.v9i1.16007>
- CdE = Mark Davies (ed.) (2016): *Corpus del español*. <https://www.corpusdelespanol.org/>.
- CDH = Instituto de Investigación Rafael Lapesa de la Real Academia Española: *Corpus del Nuevo diccionario histórico*. <https://apps.rae.es/CNDHE/view>
- CHARTA = Pedro SÁNCHEZ-PRIETO BORJA (ed.) (2011-): *Corpus hispánico y americano en la red*. <https://www.redcharta.es>.
- CHIERICHETTI, Luisa (2024): «*Patria*, del guion a la serie: la oralidad representada y escenificada en la reescritura telecinematográfica», *Artifara*, 24.2., pp. 221-238.
- CODEA+ 2022 = Pedro SÁNCHEZ-PRIETO BORJA (ed.) (2022): *CODEA+ 2022. Corpus de documentos españoles anteriores a 1900*. <https://www.corpuscodea.es>.
- COLA = Annette Myre Jørgensen (ed.) (2023): *Corpus oral de lenguaje adolescente*. <https://blogg.hiof.no/colam-esp>.
- C-ORAL-ROM= CRESTI, E. y MONEGLIA, M. (eds.) (2005): CORAL-ROM Integrated Reference Corpora for Spoken Romance Languages. Amsterdam, John Benjamins.
- CORDIAM = Concepción Company Company, Virginia Bertolotti (eds.) (2016-): *Corpus diacrónico y diatópico del español de América*. <https://www.cordiam.org/>.
- CorDisDial = Santiago Del Rey Quesada (ed.) (en preparación): *Corpus del discurso dialógico en la historia de las lenguas romances*. Universidad de Sevilla.
- CORLEC= Marcos Marín, F. (dir.): *Corpus Oral de Referencia de la Lengua Española Contemporánea*. <http://www.lllf.uam.es/ESP/Corlec.html>.
- CORMA= Enghels, Renata, Fien De Latte, Linde Roels, Nele Van Den Driessche, María Elena Azofra Sierra (2024): *El Corpus Oral de Madrid (CORMA)*. Gante: Universidad de Gante.
- COPRES XXI = Real Academia Española (ed.) (2023-): *Corpus del español del siglo XXI*. <https://www.rae.es/corpes/>.

- COSER = Fernández-Ordóñez, Inés (ed.) (2005-): *Corpus oral y sonoro del español rural*. <http://www.corpusrural.es>.
- COUPLAND, Nikolas, Jacob THØGERSEN y Janus MORTENSEN (2016): «Introduction: Style, media and language ideologies», en Jacob Thøgersen, Nikolas Coupland y Janus Mortensen (eds.), *Style, Media and Language Ideologies*. Oslo: Novus Press, pp. 11–49.
- COVJA= AZORÍN, Dolores (2002) (ed.): *Corpus del español hablado en Alicante*. Alicante: Publicaciones de la Universidad de Alicante.
- CREA = Real Academia Española (ed.) (2008): *Corpus de referencia del español actual*. <https://corpus.rae.es/creanet.html>.
- CULPEPER, Jonathan (2001): *Language and Characterisation: People in Plays and other Texts*. London: Longman Pearson Education.
- DEL REY QUESADA, Santiago (2019): «Variantes de la *oralidad elaborada* en la segunda mitad del siglo XIX: dos traducciones coetáneas de *Los cautivos de Plauto*», *Oralia* 22(2), pp. 283-326.
- DEL REY QUESADA, Santiago y Elena CARMONA YANES (2024): «Los corpus digitales en el proyecto *DiacOralEs*», *Studia linguistica romanica*, 12, pp. 86-106. <https://doi.org/10.25364/19.2024.12.5>.
- DYNEL, Marta (2011): «"You talking to me?" The viewer as a ratified listener to film discourse», *Journal of Pragmatics*, 43(6), pp. 1628–44.
- ESLORA = Vázquez Rozas, Victoria (ed.) (2007-): *Corpus para el estudio del español oral*. <http://eslora.usc.es>.
- GANCEDO RUIZ, Marta (2022): «De nuevo, reflexiones metodológicas sobre el empleo del teatro como corpus para los estudios de pragmática histórica», *SOPRAG*, 10 (1), pp. 70-88.
- GARCÍA AGUILAR, Alberto (2018): «*El cochecito*, de Rafael Azcona: El guion cinematográfico como obra literaria», *Revista Latente*, 16, pp. 83-95.
- GONZÁLEZ GOSÁLBEZ, Rafael (2016): *Actores españoles en primera persona: el oficio de cómico en sus testimonios*. Tesis Doctoral. Alacant: Universitat d'Alacant.
- HUERTA FLORIANO, Miguel Ángel (2012): «Contexto cinematográfico del tardofranquismo: síntesis de las cualidades estético-narrativas (1966-1975)», en Miguel Ángel Huerta Floriano y Ernesto Pérez Morán, *El cine de barrio tardofranquista. Reflejo de una sociedad*. Madrid: Biblioteca Nueva, pp. 33-40.
- HUERTA FLORIANO, Miguel Ángel y Ernesto PÉREZ MORÁN (eds.) (2012): *El «cine de barrio» tardofranquista. Reflejo de una sociedad*. Madrid: Biblioteca Nueva.
- HumCor= Repede, Doina (2024-): *Corpus Oral Multimodal de Humor en Español*. <https://humcor.snl.es/>
- HumText= Repede, Doina (2024-): *Corpus de humor escrito*. <https://humtext.snl.es/>
- JUCKER, Andreas H. (2021): «Features of orality in the language of fiction: A corpus-based investigation», *Language and Literature: International Journal of Stylistics*, 30(4), pp. 341–60.
- JUCKER, Andreas y Daniela LANDERT (2023): «The diachrony of im/politeness in American and British movies (1930–2019)», *Journal of Pragmatics*, 209, pp. 123–141. <https://doi.org/10.1016/j.pragma.2023.02.020>.
- KAILUWEIT, Rolf y Steffien DOMINIQUE (2023): «Acomodación fonética en conversaciones entre protagonistas argentinos y españoles en las películas *Truman* (2015) y *Bar «El Chino»* (2003)», en Rolf Kailuweit, Sandra Schlumpf, Eva Staudinger (eds.), *Migración, pluricentrismo y acomodación: nuevas perspectivas desde la lengua española*. Freiburg: Nomos, pp. 251-277.
- KAILUWEIT, Rolf (2015): «Voces de inmigrantes. La literaricidad potenciada de Roberto Arlt», en Rolf Kailuweit, Volker Jaeckel y Ángela di Tullio (eds.), *Roberto Arlt y el lenguaje literario argentino*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert, pp. 87–102.
- KOCH, Peter y Wulf Oesterreicher ([1990] 2007): *Lengua hablada en la Romania: español, francés, italiano*. Madrid: Gredos (Biblioteca Románica Hispánica, Estudios y ensayos, 448). [Versión española de Araceli López Serena del original alemán *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. Tübingen: Max Niemeyer, 1990].

- KOZLOFF, Sarah (2000): *Overhearing Film Dialogue*. Berkeley: University of California Press.
- LLOPIS CARDONA, Ana (2024): «Culturas juveniles, contracultura y procesos de difusión en el español coloquial: reflexiones previas y contribuciones», *Spanish in Context*, 21, 1-22.
- LLOPIS CARDONA, Ana y Marlies JANSEGERS (2024): «Del rollo como sustantivo comodín contracultural al rollo aproximador en el español coloquial actual», *Spanish in Context*, 21, 159-189.
- LLOPIS CARDONA, Ana y Salvador PONS BORDERÍA (2024): «Fases y factores socioculturales en la difusión de tío/tía como vocativos: “juvenilización” del español coloquial actual», *Spanish in Context*, 21, 23-50.
- LLOPIS CARDONA, Ana (2025): «La reconfiguración de la expresión del acuerdo en el registro coloquial en la segunda mitad del siglo XX», en Andrzej Zieliński y Silvia Iglesias Recuero (eds.), *Estudios de (des)cortesía verbal en la historia del español: actos de habla, modulación del discurso y formas de tratamiento*. Wilmington: Vernon Press, pp. 121 - 155.
- LÓPEZ SERENA, Araceli (2007a): *Oralidad y escrituralidad en la recreación literaria del español coloquial*. Madrid: Gredos.
- LÓPEZ SERENA, Araceli (2007ba): «El concepto de ‘español coloquial’: vacilación terminológica e indefinición del objeto de estudio», *Oralia. Análisis del discurso oral*, 10, pp. 161-191.
- LÓPEZ SERENA, Araceli (2022): «Discursividad y mimesis de la oralidad en *Paquita Salas*. Tras las huellas de la improvisación dialógica en la ficción televisiva contemporánea», en Víctor Pérez Béjar y María Méndez Orense (eds.), *Perspectivas integradas para el análisis de la oralidad*. Sevilla: Editorial Universidad de Sevilla, pp. 57-84.
- LÓPEZ SERENA, Araceli (2024): «La modelización de la situación comunicativa en la teorización sobre la variación lingüística», *Círculo de Lingüística Aplicada a la Comunicación*, 97, pp. 1-14. <https://dx.doi.org/10.5209/clac.93487>
- MANCERA RUEDA, Ana (2009): «La oralidad simulada en la narrativa contemporánea», *Verba*, 36, pp. 419-436.
- MARTÍNEZ, Josu (2023): «El euskera en el cine de los años 80: entre la falta de apoyo, la falta de realismo y la condena de la diglosia», *Zer* 28(55), pp. 241-265.
- MC ENERY, Tony y Andrew WILSON (1996): *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- MORENO FERNÁNDEZ, Francisco (2022): «De Hollywood a Netflix. Las variedades del Español en el cine y la televisión», en Instituto Cervantes (ed.), *Nuevo mundo. El impacto del audiovisual en la difusión internacional de la lengua y de la cultura en español*, pp. 349-386.
- ODE = Calderón Campos, Miguel, María T. García Godoy (2019-): *Oralia diacrónica del español*. <http://corpora.ugr.es/ode>.
- OESTERREICHER, Wulf (1996): «Lo hablado en lo escrito. Reflexiones metodológicas y aproximación a una tipología»”, en Thomas Kotschi, Wulf Oesterreicher y Klaus Zimmermann (eds.), *El español hablado y la cultural oral en España e Hispanoamérica*. Frankfurt am Main: Iberoamericana Vervuert, pp. 317-340.
- PCFD= Pavesi, Maria et al. (2005- ): *Pavia Corpus of Film Dialogue*. Pavia: Universidad de Pavia.
- PÉREZ BÉJAR, Víctor y María MÉNDEZ ORENSE (eds.) (2022): *Perspectivas integradas para el análisis de la oralidad*. Sevilla: Editorial Universidad de Sevilla.
- PLANCHENAUFT, Gaëlle. (2017): «Doing dialects in dialogues: Regional, social and ethnic variation in fiction», en Miriam A. Locher y Andreas H. Jucker (eds.), *Pragmatics of Fiction*. Berlin: De Gruyter Mouton, pp. 265-296.
- POLLAROLO, Giovanna (2011): «El guion cinematográfico, ¿texto literario?», *Lexis*, 35 (1), pp. 289-318.
- PONS BORDERÍA, Salvador (2023): «Del corpus Val.Es.Co. 3.0 a los futuros corpus diacrónicos orales: perspectivas de futuro», *Scriptum Digital*, 12, pp. 39-53.
- Post Scriptum = Centro de lingüística da Universidade de Lisboa (ed.) (2014): *Post Scriptum. Archivo digital de escritura cotidiana en Portugal y España en la Edad moderna*. <http://teitok.clul.ul.pt/postscriptum/index.php>.

- PRESEA = Moreno Fernández, Francisco, Ana Cestero Mancera (eds.). *Proyecto para el estudio sociolingüístico del español de España y de América*. <https://preseea.uah.es>.
- PUSCH, Claus (en prensa): «Les interrogatives dans l'oralité fictionnelle médiatisée. Une analyse de données du français québécois»
- QUAGLIO, Paulo (2009): *Television dialogue: The sitcom Friends vs. natural conversation*. Amsterdam/Philadelphia: John Benjamins.
- Real Decreto 1591/2010, de 26 de noviembre, por el que se aprueba la Clasificación Nacional de Ocupaciones 2011. *Boletín Oficial del Estado*, n.º 306, 10 de diciembre de 2010.
- Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia. *Boletín Oficial del Estado*, n.º 97, 22 de abril de 1996.
- RIAMBAU, Esteve (2023): «El periodo socialista (1982-1992)», en Roman Gubern et al., *Historia del cine español*. Madrid: Cátedra, pp. 399-454.
- RICHARDSON, Kay (2010): *Television Dramatic Dialogue. A Sociolinguistic Study*. Oxford: Oxford University Press.
- ROJO, Guillermo (2024): «El futuro de los corpus de referencia», *Studia linguistica romanica*, 12, pp. 18-33. <https://doi.org/10.25364/19.2024.12.2>.
- Rossi, Fabio (2011): «Discourse analysis of film dialogues: Italian comedy between linguistic realism and pragmatic non-realism», en Roberta Piazza, Monika Bednarek and Fabio Rossi (eds.), *Telecinematic Discourse: Approaches to the Language of Films and Television Series*. Amsterdam: John Benjamins, pp. 21-46.
- SABANÉS, Diego (2022): «Elogio de lo imperfecto. Berlanga y Azcona hablan sobre la comedia, la confusión de los géneros y la sociedad española», *Factoría del guion*, <https://factoriadeguion.es/elogio-de-lo-imperfecto/>
- SÁNCHEZ NORIEGA, José Luis (2002): *Historia del Cine. Teoría y géneros cinematográficos, fotografía y televisión*. Madrid: Alianza editorial.
- SÁNCHEZ-ESCALONILLA, Antonio (2016): *Del guion a la pantalla: lenguaje visual para guionistas y directores de cine*. Barcelona: Ariel.
- STADLER, Jane y Kelly MCWILLIAM (2009): *Screen media. Analysing film and television*. New York: Routledge.
- SydTV= Bednarek, Monika (2018-): *Sydney Corpus of Television Dialogue*. <https://www.syd-tv.com/>
- TOVAR, Julio (2022): «Entrevista a Fernando Colomo “El único problema de rodar es el guion”», *Jot Down Cultural Marazin*, <https://www.jotdown.es/2022/01/fernando-colomo/>
- TV Corpus/Movies Corpus= Davies, Mark (2021-): TV Corpus/Movies Corpus. <https://www.english-corpora.org/>
- Val.Es.Co. = BRIZ, Antonio y Grupo Val.Es.Co. (2002): *Corpus de conversaciones coloquiales*. Madrid: Arco Libros.
- Val.Es.Co. 3.0. = Grupo Val.Es.Co. (ed.) (2024): *Corpus Val.Es.Co. 3.0*. <http://www.valesco.es>.