

Entre lo oral y lo escrito: un estudio de corpus sobre los reformuladores *es decir*, *o sea* y *en plan* en español

MAR CAPILLA MARTÍN
Centro de Estudios de la Real Academia Española
mcapilla@rae.es

Resumen: Esta investigación tiene como objetivo identificar si existen diferencias contextuales en el uso de *es decir*, *o sea* y *en plan* entre los hablantes de español. Con la aparición de material procedente de Internet, existe un nuevo paradigma que ubica ciertos usos de los marcadores discursivos en la intersección entre lo oral y lo escrito. En este contexto, resulta pertinente un estudio que tome como referencia un corpus específico, como el CORPES XXI, caracterizado por una notable representación de este tipo de fuentes, así como por facilitar datos sobre la distribución tipológica y temática del texto. Los resultados del análisis revelan la presencia de diferencias determinadas por el tipo de texto en el que los hablantes emplean cada uno de los marcadores discursivos examinados en este estudio. En este sentido, el artículo se presenta como un trabajo que evidencia los nuevos intereses y enfoques en el análisis del discurso en la actualidad.

Palabras clave: lingüística de corpus, corpus oral, lenguaje coloquial, reformuladores, marcadores discursivos.

Between spoken and written language: a corpus study on the reformulators *es decir*, *o sea*, and *en plan* in Spanish

Abstract: This research aims to identify whether there are contextual differences in the use of *es decir*, *o sea*, and *en plan* among Spanish speakers. With the emergence of material from the Internet, a new paradigm has arisen that positions certain uses of discourse markers at the intersection between oral and written forms. In this context, a study that takes as a reference a specific corpus, such as CORPES XXI, characterized by a remarkable representation of this type of sources, as well as by providing data on the typological and thematic distribution of the text, is pertinent. The results of the analysis reveal the presence

of differences determined by the type of text in which speakers use each of the discourse markers examined in this study. In this sense, the article constitutes a work that highlights new interests and approaches in contemporary discourse analysis.

Keywords: corpus linguistics, oral corpus, colloquial language, reformulators, discourse markers.

1. Introducción

En el vasto campo de la lingüística contemporánea, el uso de corpus ha revolucionado la manera en que investigamos y comprendemos el lenguaje. El corpus lingüístico, ya sea oral o escrito, constituye el recurso más efectivo y natural para el análisis del lenguaje. Por ello, cualquier investigación lingüística debe fundamentarse en un conjunto amplio y bien delimitado de materiales que ofrezcan datos fiables (Briz 2012). Entre los diferentes tipos de corpus, los corpus que contienen textos escritos y orales ocupan un lugar destacado al capturar el lenguaje en diferentes medios, lo que permite un análisis profundo de fenómenos lingüísticos (Bolaños 2015; Rojo 2021). Los corpus lingüísticos se caracterizan por tres aspectos distintivos: un diseño, la recuperación selectiva de datos y la codificación textual. Por otro lado, también la calidad de los textos incluidos es un elemento fundamental, estrechamente ligada al proceso de codificación y anotación. En este sentido, resulta esencial subrayar la relevancia de la selección y codificación del material para garantizar la calidad de los datos que serán analizados en fases científicas posteriores.

Teniendo esto en cuenta, el presente artículo pretende explorar la importancia y relevancia de los corpus que contienen textos escritos y orales en la investigación lingüística del habla coloquial. A través de ejemplos y estudios de caso, se examinará la presencia y la naturaleza de tres marcadores discursivos propios de este tipo de lenguaje en diferente tipología y temática textual con el fin de arrojar luz sobre qué contexto lingüístico es más propenso a presentar estos marcadores. El trabajo se estructurará en tres secciones principales. En primer lugar, se presentarán las características fundamentales que debe tener un corpus, o subcorpus, compuesto por textos escritos y orales, así como los beneficios de su utilización para la investigación lingüística. En segundo lugar, se ofrecerá un resumen de los rasgos más relevantes de los reformuladores *es decir*, *o sea* y *en plan*, con el propósito de vincularlos a los datos extraídos del corpus empleado para nuestra investigación, el CORPES XXI. Finalmente, se analizará la información que este corpus

proporciona y su distribución tipológica y temática, con el fin de identificar en qué contextos, ya sean orales o escritos, se emplean estos marcadores del discurso.

2. Los corpus, o subcorpus, orales como fuente de investigación

El empleo de corpus lingüísticos en el campo de la lingüística ha transformado el estudio de la lengua en las últimas décadas. La lingüística de corpus ha subrayado la importancia de fundamentar la descripción lingüística en un estudio exhaustivo de esta en su uso natural, proporcionando una perspectiva de análisis más amplia y datos de frecuencia más precisos que facilitan la interpretación y la veracidad de los resultados. En consecuencia, podemos afirmar que el uso de corpus lingüísticos ha revolucionado la investigación del lenguaje (Rojo 2015, 2021). Los corpus lingüísticos son un conjunto de textos, tanto orales como escritos, obtenidos en condiciones naturales y pertenecientes a una lengua o variedad lingüística específica, que se almacenan en formato electrónico y se codifican con el propósito de ser sometidos a análisis. La elaboración de un corpus implica la integración de textos conforme a un diseño previamente establecido. En este sentido, cada corpus adopta una estructura general que responde a los objetivos para los cuales ha sido construido (Rojo 2021). En cuanto a su tipología, los corpus lingüísticos pueden ser compilados de diversas maneras y en variados tamaños, abarcando desde colecciones de textos de un único autor hasta extensas recopilaciones de textos de distintos géneros y épocas. Cada corpus se crea con unos objetivos específicos que lo diferencian de otros corpus, por lo que se pueden agrupar y clasificar según distintos criterios. Dado que nuestra investigación se centra en la tipología y temática textual, abarcando tanto textos escritos como orales, es necesario basarnos en corpus que distingan el medio en el que se encuentra el texto, ya sea escrito u oral. En este sentido, existen corpus de textos escritos, conocidos como *written corpora* o *text corpora*, que consisten en textos originalmente escritos. Por otro lado, tenemos los corpus que contienen textos orales, que consisten principalmente en la transcripción ortográficamente convencional de grabaciones. El tratamiento y análisis del corpus se efectúa a partir de dicha transcripción, denominada *texto hablado* (*spoken text*). La transcripción ortográfica se enriquece con diversos elementos que reflejan el proceso de producción del habla, adaptándose a los objetivos y aplicaciones del corpus. El objetivo es proporcionar una representación simbólica del uso oral espontáneo y natural de la lengua, sin incluir la variación fonética (Sinclair 1996).

En este contexto, es fundamental entender que la distinción entre oral y escrito suele estar asociada con la diferencia en los registros: los textos orales se vinculan comúnmente con el lenguaje coloquial, mientras que los textos escritos tienden a reflejar una orientación más literaria y cercana a la variedad estándar en el ámbito lingüístico correspondiente. A pesar de que esta asociación puede ser adecuada en gran medida, es importante reconocer que la naturaleza de la distinción se centra en el soporte material del texto y no en otras características. Una prueba de ello es la aparición y evolución de los medios electrónicos y las redes sociales, que ha dado lugar a nuevos géneros en los cuales el soporte escrito se combina frecuentemente con la lengua coloquial, como es el caso de los blogs y los tuits (Rojo 2021). Así, es importante destacar que, incluso en textos escritos, es posible identificar patrones propios de la oralidad. Esto se debe a que, en numerosos casos, dichos textos buscan emular la oralidad mediante el uso de diálogos, la transcripción de entrevistas y otros recursos similares.

En español, existe una amplia disponibilidad de corpus que contienen textos escritos como el CORDE, el CREA, el CORPES XXI, el CE o el CDH; y textos orales, entre los cuales destacan el corpus PRESEEA, Val.Es.Co., COSER, ESLORA, AMERESCO, COVJA o CREA, entre otros. Estos corpus, de distintas dimensiones y características, permiten obtener una visión general de fenómenos lingüísticos, proporcionando una base esencial para el estudio del habla espontánea y coloquial. Su relevancia es particularmente significativa para los análisis pragmalingüísticos y sociopragmáticos del español, facilitando la investigación en áreas como la atenuación, la cortesía o los marcadores discursivos (Briz y Carcelén 2019). Además, es relevante destacar la accesibilidad de estas herramientas, ya que están disponibles a través de plataformas en línea que permiten filtrar las búsquedas. Algunas de estas plataformas incluso facilitan la creación de subcorpus específicos, lo que posibilita la extracción de información más detallada y precisa, optimizando así su utilización en la investigación lingüística. Uno de estos corpus es el Corpus del Español del Siglo XXI (CORPES XXI), un corpus de referencia que incluye tanto un subcorpus oral como información textual escrita, lo cual permite la comparación de datos, enfocando el análisis en rasgos del habla espontánea y coloquial en diferentes medios. Este corpus se caracteriza por su rigor metodológico: los textos están cuidadosamente anotados con metadatos relevantes, como el tipo de texto, el país de origen, el año, el medio, el soporte, la tipología textual o el área temática. En el caso de los textos orales, se incorporan además datos sociolingüísticos específicos, como el sexo, la edad, el nivel educativo y la profesión de los hablantes. Asimismo, los textos están marcados con información morfosintáctica, lo que facilita la búsqueda y el análisis de palabras, frases o estructuras gramaticales. Por otro lado, la calidad del CORPES XXI radica en su representatividad, su actualización

constante y su diseño equilibrado, que garantiza la fiabilidad de los datos. Este enfoque asegura una selección rigurosa de los datos para garantizar una distribución equilibrada en relación con países, zonas geográficas, tipologías textuales y áreas temáticas. Debido a su diseño, codificación y a la calidad de los datos que ofrece, hemos seleccionado este corpus lingüístico para la investigación de tres partículas discursivas del español, *es decir*, *o sea* y *en plan*, con el objetivo de determinar los contextos más favorables para su uso. Con este enfoque, además, buscamos demostrar la utilidad de los corpus como herramientas para la investigación lingüística y, en particular, del habla coloquial.

3. Los reformuladores *es decir*, *o sea* y *en plan*

Como señala Portolés (1998), los marcadores del discurso no presentan la misma distribución en textos escritos y en conversaciones. De este modo, existen marcadores que son más frecuentes en el discurso oral y otros que, por el contrario, suelen limitarse al discurso escrito. Según este mismo autor, la diferencia principal en la distribución se debe al contexto del discurso oral y del discurso escrito. En la comunicación oral, los interlocutores tienen acceso directo a una abundante información contextual; mientras que el discurso escrito se caracteriza por una menor riqueza contextual, por lo que es necesario que se explicita lingüísticamente toda la información contextual necesaria para conseguir una comunicación efectiva (Portolés 1998). En el primer caso, se incluirían *es decir* y *o sea*, partículas vinculadas al lenguaje escrito; mientras que en el segundo caso, más propia de la oralidad, se encontraría *en plan* (Briz, Pons y Portolés 2008; Ciapuscio 2001). No obstante, como se ha señalado previamente, la aparición de material proveniente de Internet (blogs, redes sociales, etc.) ha dado lugar a nuevos géneros en los que el soporte escrito se combina frecuentemente con el lenguaje coloquial. En este sentido, será interesante investigar, a través de la documentación del corpus seleccionado, CORPES XXI, en qué contextos son comunes estos marcadores, aparentemente divididos en marcadores característicos del discurso escrito y/u oral.

En español, los conectores discursivos *es decir*, *o sea* y *en plan* se categorizan dentro del grupo de reformuladores, cuya función es introducir una nueva formulación en el discurso. La reformulación es una característica intrínsecamente vinculada a la oralidad conversacional. Esta se lleva a cabo, según Martín Zorraquino y Portolés (1999), de dos formas: reexpresando de manera más precisa lo mencionado anteriormente o expresando directamente las conclusiones que deberían derivarse del primer segmento, introduciendo un nuevo tema con un carácter conclusivo. En este sentido, aunque todos son clasificados como reformuladores, se evidencia un comportamiento contextual

diferenciado para *es decir*, *o sea* y *en plan*, dependiendo de la naturaleza del discurso o del texto, que será examinado detenidamente en el análisis de nuestro trabajo.

Los reformuladores explicativos introducen una idea que explica la anterior (NGLE § 30.13; Casado 1991; Portolés 1998). En concreto, existe una relación de equivalencia en distintos grados, lo que los hace intercambiables (Briz 2001). De manera particular, el marcador explicativo *es decir* se caracteriza principalmente por su función parafrástica, al expresar de manera diferente lo que se ha mencionado previamente, estableciendo una equivalencia entre ambas formulaciones. A través de esta operación reflexiva, el emisor ejerce un control metadiscursivo que demuestra su preocupación por el discurso y la comprensión del mismo (Portolés 1998; Núñez, Muñoz y Mihovilovic 2006).

- (1) Por ello, los cabildos y el ejército ocuparían las funciones de gobierno y administración de las instituciones del Imperio Español, **es decir**, Capitán General, Real Audiencia, Intendencia de Real Hacienda y Gobernaciones de las distintas provincias (CORPES XXI, *Rebeliones, alzamientos y movimientos preindependentistas en Venezuela* [Ven.] 2001).

Otros autores como Ciapuscio (2001) defienden que *es decir* también participa en procedimientos de reformulación no parafrástica, implicando un cambio en la perspectiva enunciativa (García Negroni 2009).

- (2) Si deseo ver obras de varios autores, en cada búsqueda tengo que cambiar la opción. **Es decir**, tengo que esforzarme el doble (CORPES XXI, *No me hagas pensar: Una aproximación a la usabilidad en la Web de Steve Krug* [Esp.] 2001).

Tiene un sentido parcialmente equivalente el conector *o sea*, cuando funciona como reformulador explicativo, que implica que el segmento donde aparece funciona como una explicación, clarificación o corrección de todo o parte de lo mencionado anteriormente.

- (3) Lo primero que lleva la papa es una semilla apropiada y, desde luego, no le puede faltar la nutrición, que en nuestro caso lo hacemos con biofertilizantes, **o sea**, productos biológicos logrados aquí mismo en la finca (CORPES XXI, *La papa, bien atendida, vale por dos* [Cuba] 2024).

El siguiente, y último, reformulador es *en plan*, que se posiciona como una alternativa a otros marcadores que se utilizan en discursos

más formales, como el ya mencionado *es decir*. Las funciones atribuidas a *en plan*, según los estudios existentes, incluyen el papel de atenuador, intensificador, reformulador, ejemplificador, introductor de estilo referido y prolongador (Méndez Orense 2016; Borreguero Zuloaga 2020). Para nuestro estudio comparativo con los marcadores *es decir* y *o sea*, nos enfocaremos en su función como reformulador explicativo. En este contexto, *en plan* actúa principalmente como un conector entre diferentes partes del discurso, facilitando la interpretación del mensaje por parte del interlocutor (Briz 2001). Al desempeñar el papel de reformulador explicativo, algunas autoras como Méndez Orense (2016: 135-136), plantean la posibilidad de que *en plan* haya empezado a desplazar a *o sea* como el reformulador más frecuente en el lenguaje coloquial. No obstante, *en plan* se distingue de *o sea* en que el primero de ellos funciona exclusivamente como reformulador explicativo y no como rectificativo, como ocurre con *o sea*.

Con el fin de investigar los contextos textuales y la extensión del uso de estos tres reformuladores con valores explicativos, ya sea en términos tipológicos y temáticos, procedemos a analizar el uso documentado de *en plan*, *o sea* y *es decir* en el corpus, y subcorpus, seleccionado para nuestra investigación: el CORPES XXI.

4. Resultados y análisis de *en plan*, *o sea* y *es decir* en el CORPES XXI

El análisis llevado a cabo a través de la documentación de corpus se ha consolidado como una metodología clave en la investigación lingüística. Este enfoque permite no solo identificar y catalogar patrones recurrentes en el uso del lenguaje, sino también entender las circunstancias y contextos en los que tales fórmulas son más propensas a surgir. En este apartado, se llevará a cabo un análisis exhaustivo de *es decir*, *o sea* y *en plan*, empleando documentación procedente de corpus, con el objetivo de establecer su naturaleza y determinar los contextos lingüísticos en los que se manifiestan con mayor frecuencia estos marcadores. Con el fin de alcanzar este objetivo establecido, se empleará como fuente de datos el CORPES XXI. Este corpus es un corpus de referencia con 25 millones de formas por año y una distribución de textos del 70 % provenientes de América y del 30 % de España. Los textos se seleccionan conforme a criterios específicos: el 90 % corresponde a la lengua escrita y el 10 % a la oral. Los materiales escritos se dividen en libros (40 %), publicaciones periódicas (40 %), material de Internet (7,5 %) y miscelánea (2,5 %). Geográficamente, asigna el 30 % de los textos a España y el 70 % a América, incluyendo por primera vez textos de Guinea Ecuatorial y Filipinas. La clasificación temática uniforme de los textos, organizada en ficción y no ficción, y su subdivisión en

áreas temáticas, junto con la categorización según género textual (novelas, relatos, noticias, reportajes, entrevistas, etc.), crea una tipología textual compleja que mejora la precisión en la recuperación de información para los investigadores. En resumen, el CORPES XXI es una herramienta representativa y equilibrada diseñada para el estudio del español contemporáneo, con una estructura meticulosa que garantiza la fiabilidad de los datos.

Atendiendo a la búsqueda de *es decir*, *o sea* y *en plan* en el CORPES XXI, encontramos la presencia de estos tres reformuladores en textos de distinta naturaleza: prensa y libros, orales e Internet. En particular, podemos hacer una distinción clave según el origen del texto: escrito u oral.

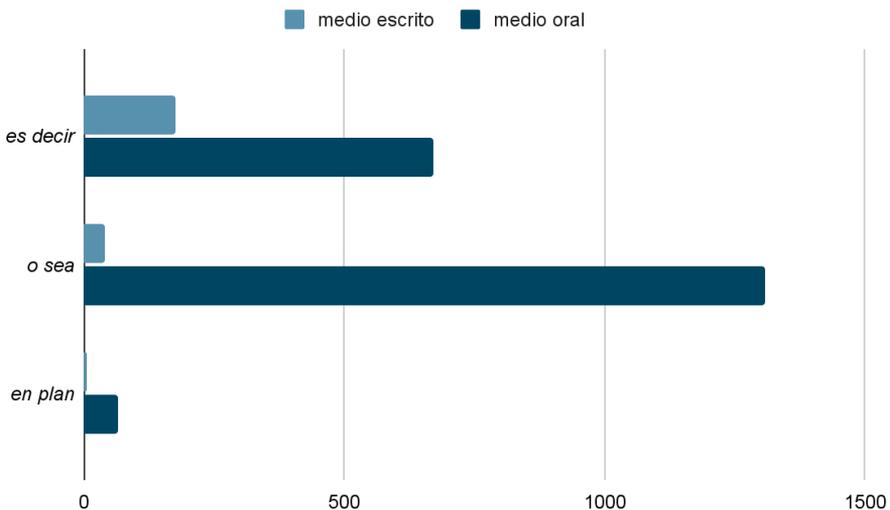


Gráfico 1. Distribución de *en plan*, *es decir* y *o sea* según el medio.

La información presentada en el Gráfico 1, y a continuación en la Tabla 1, muestra que estos reformuladores son característicos del lenguaje oral. Sin embargo, no podemos ignorar la presencia de estos elementos en textos escritos, donde, en algunos casos, se registra un número relevante de concordancias.

	Frecuencia normalizada	
	Escrito	Oral
<i>Es decir</i>	177,6	664,81
<i>O sea</i>	41,43	1353,41
<i>En plan</i>	4,44	69,81

Tabla 1. Distribución de *en plan*, *es decir* y *o sea* según el medio.

Al analizar con detalle cada marcador, observamos que, en el caso de *es decir*, la mayoría de los ejemplos provenientes de textos orales, como el ejemplo 1a, corresponden a autores con un alto nivel educativo. Esto sugiere que, aunque aparezca en textos orales, su uso tiende a estar asociado a discursos más formales, alejados de la coloquialidad propia de la oralidad. Del mismo modo, podría considerarse que el tipo de discurso oral puede ejercer influencia, dado que en el ámbito radiofónico, es común encontrar textos oralizados. En cuanto a los ejemplos de textos escritos, se pueden distinguir casos procedentes de Internet, como el ejemplo 3b, que proviene de un blog de cocina de una marca conocida, donde se infiere un uso más cuidado del lenguaje. Esta tendencia se ha observado en los últimos años en textos de Internet, como blogs, tuits o cuentas de Instagram, los cuales pertenecen a instituciones o cuentas verificadas con una importante cantidad de seguidores, donde el lenguaje es más cuidado.

- (3) a. eso no quiere decir que no haya méritos, **es decir**, evidentemente, el mérito tiene que estar y a partir de ahí (CORPES XXI, *Cadena SER: Las entrevistas de Aimar* | Baltasar Garzón, 2022).
- b. Siempre debemos bridar el pavo, **es decir**, amarrar las piernas y acomodar los alones para que no pierdan forma, así mismo, cubrir la orilla de las piernas para evitar que se quemem (CORPES XXI, *Thermomix Blog - Thermomix México*, 2023).

En particular, y en línea con el objetivo de esta investigación, los datos del CORPES XXI muestran una frecuencia normalizada mayor en textos orales (664,81 por millón) que en textos escritos (177,6 por millón). Como se observa en la Tabla 2 y la Tabla 3, en las cuales se presentan los tres casos tipológicos más frecuentes, el marcador *es decir* es característico de textos escritos formales y de textos orales en los que se emplea un lenguaje más formal y elaborado, mostrando características similares a las de los textos escritos.

Tipología textual	Frecuencia absoluta	Frecuencia normalizada
Académico	11 670	379,84
Manual de instrucciones	22	365,05
Libro de texto	122	333,16

Tabla 2. Distribución de *es decir* en textos escritos según la tipología textual.

Tipología textual	Frecuencia absoluta	Frecuencia normalizada
Tertulia	249	1 175,65
Debate	101	823,83
Discurso	40	736,87

Tabla 3. Distribución de *es decir* en textos orales según la tipología textual.

Algo similar ocurre con *o sea*, que se documenta con frecuencia en textos escritos, pero sobre todo orales. Así, el CORPES XXI evidencia una frecuencia normalizada de 1 353,41 por millón en textos orales y 41,43 por millón en textos escritos. En términos generales, este marcador se ha catalogado dentro del lenguaje hablado; sin embargo, su presencia en textos escritos proporciona información significativa. Se observa que en la mayoría de los ejemplos orales, así como en muchos de los escritos, se utiliza un lenguaje coloquial característico de la oralidad. Por ejemplo, en 4a se presenta una entrevista transcrita con un tono relajado y menos formal, y en 4b, un texto extraído de Internet muestra el reformulador en una entrevista de un blog, en la que el lenguaje se acerca a rasgos propios de la oralidad. Siguiendo a Brown y Gillman (1960), en textos orales en los que se da una situación en la que no existen relaciones de poder y sí de solidaridad entre los conversadores, y en la que hay además afinidad generacional y, quizá también, de sexo, se propicia el habla coloquial. Por consiguiente, en el ejemplo 4a la relación entre iguales, en este caso mujeres de edades similares, favorece la coloquialidad.

- (4) a. [...] esto ni me lo creo de que pude hacer yo ese trabajo / eehh / **o sea** no / **o sea** nunca nunca soñé que íbamos a poder a a lograr tantas cosas (CORPES XXI, *A solas: Sofía Vergara y Vicky Martín Berrocal* | *A SOLAS CON: Capítulo 11* | *Podium Podcast*, 2024).
- b. Con Armando fue muy curioso, me acuerdo que dije “quiero trabajar con Armando”, **o sea** yo quería hacer la obra con alguien con quien yo quisiera trabajar, hice una lista de tres actores con el perfil de personaje con los que quería trabajar y lancé mi primera carta, que era Armando (CORPES XXI, *El blog de FilminLatino*, 2023).

De esta manera, a partir de la documentación extraída del CORPES XXI y tal como se detalla en la Tabla 4, se concluye que *o sea* aparece frecuentemente en contextos comunicativos orales espontáneos, donde este marcador discursivo facilita al hablante iniciar sus intervenciones y organizar la estructura del discurso.

Tipología textual	Frecuencia absoluta	Frecuencia normalizada
Entrevista	5 139	3 099,27
Tertulia	520	2 455,18
Sorteos y concursos	4	1 416,43

Tabla 4. Distribución de *o sea* en textos orales según la tipología textual.

En cuanto a *en plan*, este marcador se diferencia notablemente de los contextos formales. Tal como se muestra en la Gráfica 1, se trata de una partícula predominantemente oral, aunque también se encuentra en textos escritos. En este sentido, el CORPES XXI registra una frecuencia normalizada de 4,44 por millón en textos escritos y 69,81 por millón en textos orales. Teniendo en cuenta su uso reciente y la cantidad de documentación ya recopilada en los corpus, es posible que este reformulador sea considerado uno de los más frecuente en el lenguaje coloquial (v. Tabla 5).

Tipología textual	Frecuencia absoluta	Frecuencia normalizada
Mensaje en redes sociales	2	32,97
Entrevista	363	28,69
Tertulia	4	19,04

Tabla 5. Distribución de *en plan* según la tipología textual.

Además de encontrarse en textos orales (5a), *en plan* aparece en textos escritos como prensa o libros, como se observa en 5b.

- (5) a. normalmente este tipo de personas tiene doble cara lo dice con tanta seguridad / y tanta frialdad que hasta yo misma dudo de mí misma y digo **en plan** puede ser que yo (CORPES XXI, *Luc Loren | Somos Estupendas: "LO PEOR FUE EL MALTRATO PSICOLÓGICO"* - Jessica Goicoechea #NoEstamosLocas, 2023).
- b. Te he traído una película —dice—. El extraño caso de Angélica, de Manoel de Oliveira. No la has visto, ¿verdad? —Dolores niega con la cabeza—. Da un poco de bajón, pero me ha recordado un poco a lo que está pasando, **en plan** un fotógrafo al que lo lían para unas fotos de difuntos. Lo que pasa es que el tío luego se enamora de la muerta (CORPES XXI, *Anoxia*, 2023).

En este contexto, lo más notable es la presencia de estos marcadores, que inicialmente se consideran propios de la lengua hablada, en textos escritos. Por lo tanto, resulta pertinente examinar su comportamiento en este medio, donde, *a priori*, se presume la existencia de un lenguaje más elaborado y formal. De acuerdo con las observaciones realizadas, procedemos a clasificar los datos recuperados del CORPES

XXI primero según la tipología textual (Gráfico 2) y, en segundo lugar, según la temática del texto (Gráfico 3). De acuerdo con la exposición de este estudio, se observa un mayor número de casos de *es decir* en textos de carácter formal con un estilo más elaborado, tales como textos académicos, libros de texto, noticias, reportajes, opiniones y prospectos, donde existe una planificación y se evita la espontaneidad. Entre *es decir* y *en plan*, se sitúa el reformulador *o sea*, que combina características de los dos extremos del espectro formal-coloquial. En textos escritos, este marcador se emplea habitualmente con un tono formal, en contextos donde el lenguaje es cuidado y planificado; mientras que, en textos orales o provenientes de Internet, parece favorecer un tono informal, la inmediatez, el diálogo y la espontaneidad, rasgos propios de la conversación. Por otro lado, *en plan* prevalece en textos de carácter informal, donde el lenguaje se aproxima más a la oralidad, como entrevistas, divulgación, blogs y redes sociales. Según Briz (1995, 2001), existen rasgos situacionales o coloquializadores que promueven el uso del registro coloquial. En relación con esto, en los ejemplos extraídos de la documentación del CORPES XXI, encontramos que factores como la igualdad entre los interlocutores, la relación de proximidad de los mismos y el marco discursivo familiar, determinado por el espacio físico y la relación específica de los participantes con ese espacio, entre otros, influyen en la preferencia por este registro.

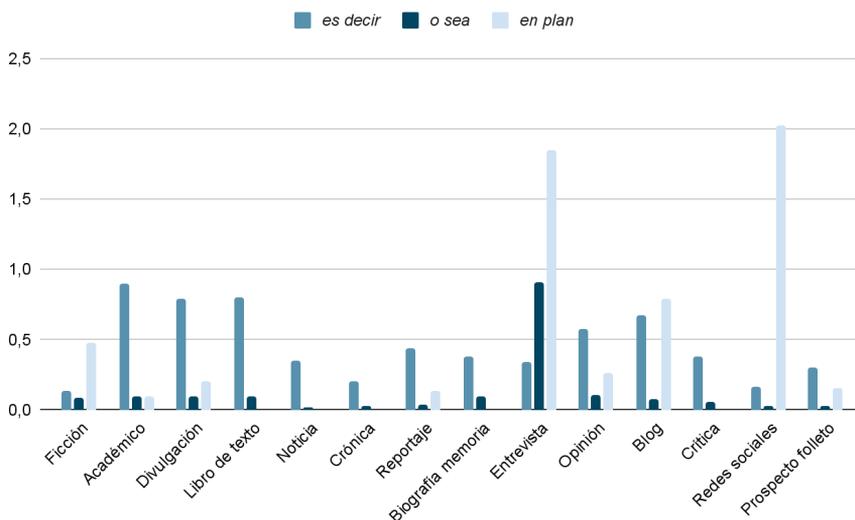


Gráfico 2. Distribución de *en plan*, *es decir* y *o sea* según la tipología textual.

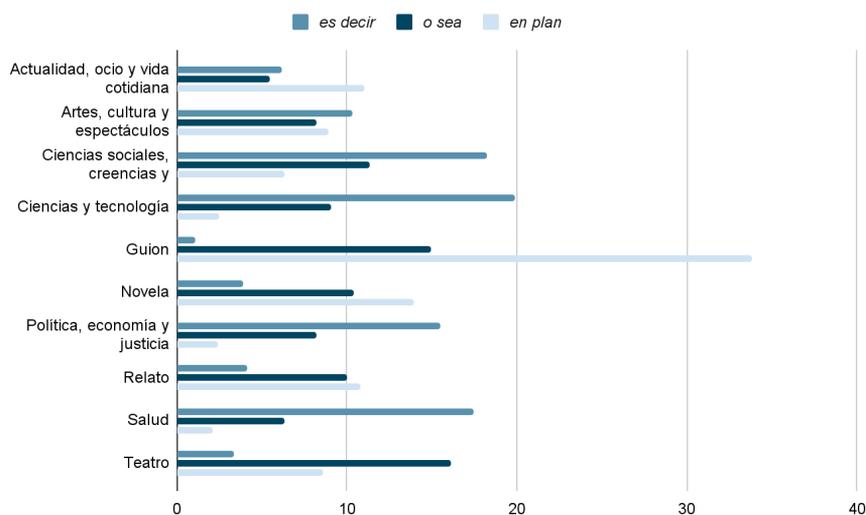


Gráfico 3. Distribución de en plan, es decir y o sea según la clasificación temática.

Considerando la información presentada en el Gráfico 2 y la información subsecuente del Gráfico 3, resulta notable la documentación proveniente de bloques como ficción para *en plan* y, en menor medida, *o sea*. Según lo planteado, en este tipo de documentación textual debería prevalecer el uso del reformulador *es decir*; sin embargo, al revisar los ejemplos, se descubre que en ficción los casos de *en plan* son, en realidad, casos en discursos que imitan la oralidad, ya que la mayoría aparecen en diálogos que simulan conversaciones entre personajes. En este tipo de textos, es fundamental la temática no especializada, la cotidianidad pretendida, la aparente ausencia de planificación, el tono informal, el dinamismo y el diálogo. Por lo tanto, este tipo de contenido textual puede ser considerado como un escrito coloquial que, en realidad, encubre una oralidad coloquial.

En este punto, resulta pertinente clasificar la información sobre estos tres marcadores, de acuerdo con los datos del corpus, en categorías de *+formal* o *-formal*, considerando las características típicas de cada tipo de texto. Con ello, se busca identificar qué reformulador predomina en cada contexto textual para determinar los rasgos distintivos de cada marcador y los tipos de textos en los que se emplean, según la información proporcionada por el CORPES XXI.

En este sentido, la Tabla 6 muestra, a través de la frecuencia normalizada —también presentada en el Gráfico 2—, que en textos generalmente formales (académicos, de divulgación, libros de texto, noticias, crónicas, reportajes, biografías, textos de opinión, críticas o prospectos) el marcador discursivo *es decir* es el más frecuentemente utilizado, marcando una considerable diferencia respecto a los otros

dos marcadores, *o sea* y *en plan*. En contraste, en textos de ficción, donde se incluyen diálogos, entrevistas, blogs o textos provenientes de redes sociales, el marcador *en plan* es el más utilizado, superando al marcador *es decir*. No obstante, se observa una excepción en los blogs, donde *es decir* también está presente en una proporción significativa. Esto se debe a que, en los blogs, las características de la escritura y la oralidad tienden a difuminarse. Así, se identifican blogs de marcas comerciales, instituciones o periódicos nacionales que se adhieren a normas de redacción establecidas por sus respectivos manuales de estilo; en contraste, también existen blogs que emplean un lenguaje más cercano a la oralidad y a la coloquialidad, lo que facilita el predominio del marcador *en plan*.

Tipología textual	+formal	-formal	Marcadores discursivos		
			<i>es decir</i>	<i>o sea</i>	<i>en plan</i>
Ficción	+	+	0,14	0,09	0,48
Académico	+	-	0,9	0,1	0,1
Divulgación	+	-	0,79	0,1	0,2
Libro de texto	+	-	0,8	0,1	0
Noticia	+	-	0,35	0,02	0
Crónica	+	-	0,2	0,03	0
Reportaje	+	-	0,44	0,04	0,14
Biografía memoria	+	-	0,38	0,1	0
Entrevista	+	+	0,34	0,91	1,85
Opinión	+	-	0,58	0,11	0,26
Blog	+	+	0,67	0,08	0,79
Crítica	+	-	0,38	0,06	0
Redes sociales	+	+	0,17	0,03	2,02
Prospecto folleto	+	-	0,3	0,03	0,16

Tabla 6. Distribución de *en plan*, *es decir* y *o sea* según una propuesta de clasificación textual.

5. Conclusiones

A partir del estudio realizado, podemos concluir que es fundamental contar con datos empíricos y fiables que hayan sido previamente codificados siguiendo criterios establecidos. La recuperación de textos, tanto orales como escritos, a través del uso de corpus en general, y de subcorpus orales en particular, permite desarrollar una investigación capaz de generar conclusiones significativas en el análisis lingüístico de marcadores discursivos como *es decir*, *o sea* y *en plan*. Así, sin una codificación previa del corpus utilizado, en este caso el CORPES XXI, cuyo diseño y codificación consideran tanto la tipología textual como la temática, no habría sido posible alcanzar las conclusiones que se presentan a continuación.

Este artículo ha presentado, en primer lugar, un análisis basado en corpus sobre los marcadores discursivos previamente mencionados. Los resultados indican que, aunque los tres marcadores son típicos de la oralidad, existe una clara diferenciación en su uso según el medio y la tipología textual en la que se emplean. De acuerdo con los datos obtenidos, *es decir* se caracteriza principalmente por su presencia en textos orales y, en menor medida, en textos escritos. En ambos casos, eso sí, es propio de discursos más formales, incluso en textos orales, que parecen tener rasgos propios frente a los escritos. El marcador *o sea*, por su parte, ocupa una posición intermedia entre *es decir* y *en plan*, ya que su campo de uso no parece estar claramente delimitado. En general, su frecuencia normalizada es similar en los distintos tipos de textos analizados, aunque destaca en el género de la entrevista, donde su frecuencia es mayor, seguida por los textos de opinión, que tienden a ser menos coloquiales y más formales. En contraste, el marcador *en plan* tiene un campo de uso bien definido, predominando en textos coloquiales cercanos a la oralidad, como blogs, entrevistas, redes sociales y textos de ficción con diálogos.

En cuanto al uso de estos marcadores según la clasificación temática de los textos, se observa una notable presencia de *es decir* en textos de no ficción que abordan temas como Salud, Política, economía y justicia, y Ciencias y tecnología. Por otro lado, *en plan* se concentra mayoritariamente en textos de ficción, como guiones, novelas, relatos y obras teatrales, así como en textos de no ficción relacionados con temas de Actualidad, ocio, vida cotidiana, y Artes, cultura y espectáculos. Cabe destacar también la presencia de *o sea* tanto en textos con rasgos formales como en aquellos más coloquiales. Como se ha señalado en este estudio, este marcador parece compartir características de ambos registros, reflejando la delgada línea que existe entre ellos. Por ello, la aparición de nuevas tipologías textuales híbridas, que combinan elementos propios de distintos medios orales y escritos, dificulta la categorización de estos marcadores en una sola clasificación. Teniendo esto en cuenta, consideramos relevante destacar que la principal diferencia en la distribución de estos marcadores está relacionada con el contexto *+formal* o *-formal*, más que con la distinción entre discurso oral y escrito.

En resumen, este artículo ha tenido como propósito analizar la frecuencia de uso y los contextos en los que se emplean los marcadores *es decir*, *o sea* y *en plan*, presentando los datos proporcionados por el CORPES XXI en relación con su distribución tipológica y clasificación temática. Con ello, además, se ha enfatizado la relevancia de incorporar datos de corpus, los cuales aportan mayor fiabilidad a las investigaciones gracias a la calidad y cantidad de información disponible. Finalmente, este trabajo busca ser una invitación a utilizar estos

datos como base para futuras investigaciones que profundicen en el comportamiento de los marcadores discursivos *es decir, o sea* y *en plan* teniendo en cuenta otras variables.

BIBLIOGRAFÍA

- AMERESCO = Albelda, Marta, y Estellés, María (dirs.), *Corpus Ameresco*. Disponible en: <https://corpusameresco.com>.
- Bolaños, Sergio (2015), «La lingüística de corpus: perspectivas para la investigación lingüística contemporánea», *Forma y Función*, 28 (1): 31-54. DOI: 10.15446/fyf.v28n1.49564.
- Borreguero Zuloaga, Margarita (2020), «Los marcadores de aproximación (en el lenguaje juvenil): esp. *en plan* vs. it. *tipo*», en Miguel Ángel Cuevas Gómez, Fernando Molina Castillo y Paolo Silvestri (coords.), *España e Italia: un viaje de ida y vuelta: studia in honorem Manuel Carrera Díaz*, Sevilla, Editorial de la Universidad de Sevilla: 53-78.
- Briz Gómez, Antonio (1995), «La conversación coloquial (materiales para su estudio)», en Antonio Briz (coord.), *Anejo XVI de la Revista Cuadernos de Filología*, Valencia, Universitat de València: 239-240.
- Briz Gómez, Antonio (2001), *El español coloquial en la conversación: esbozo de pragmatogramática*, Barcelona, Ariel.
- Briz, Antonio (2002), «Apuntes para la definición lexicográfica de *o sea*», en Bernhard Pöll y Franz Rainer (eds.), *Vocabula et vocabularia: études de lexicologie et de (méta)lexicographie romanes en l'honneur du 60e anniversaire de Dieter Messner*, Frankfurt am Main, Peter Lang: 45-52.
- Briz, Antonio, Salvador Pons, y José Portolés (coords.) (2008), *Diccionario de partículas discursivas del español*, [versión en línea]. Disponible en: www.dpde.es.
- Briz, Antonio (2012), «Los déficits de los corpus orales del español (y de algunos análisis)», en Tomás Eduardo Jiménez Juliá *et al.* (coords.), *Cum corde et in nova grammatica: estudios ofrecidos a Guillermo Rojo*, Santiago, Universidad de Santiago de Compostela: 115-137.
- 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, Bala Perdida/Instituto Cervantes: 189-217.
- Brown, Roger, y Albert Gilman (1960), «The pronouns of power and solidarity», en Pier Paolo Giglioli (ed.), *Language as social context: selected readings*, Nueva York, Penguin Brooks: 252-281.
- Camargo, Laura, y Ana María Grimalt (2022), «Nuevas y viejas funciones de *en plan*: estudio microdiacrónico en corpus orales y digitales del castellano de Mallorca en el siglo XXI», *Revista de Investigación Lingüística*, 25, e537931. DOI: 10.6018/rii.537931.
- Casado Velarde, Manuel (1991), «Los operadores discursivos *es decir*, *esto es*, *o sea* y *a saber* en español actual: valores de lengua y funciones textuales», *Lingüística Española Actual*, 13 (1): 87-116.
- Ciapuscio, Guiomar Elena (2001), «Los conectores reformulativos: el caso de *es decir*», en Elvira Narvaja de Arnoux y Ángela Di Tullio (comps.), *Homenaje a Ofelia Kovacci*, Eudeba: 157-172.
- CORPES XXI = Real Academia Española, *Corpus del Español del Siglo XXI (CORPES)*. Disponible en: <https://www.rae.es/corpes>. [Fecha de consulta: 8 de junio de 2024].
- COSER = Fernández-Ordóñez, Inés (dir.), *Corpus Oral y Sonoro del Español Rural (COSER)*. Disponible en: <https://coser.lllf.uam.es>.
- Costa Outeiro, Susana Luisa (2021), *Los marcadores del discurso en plan, o sea y rollo en WhatsApp: funciones textuales*, trabajo de Fin de Grado, Universidade da Coruña.
- COVJA = *Corpus oral de la variedad juvenil universitaria del español de Alicante*; integrado en el *Corpus oral para el estudio del lenguaje juvenil y del español hablado en Alicante*.
- CREA = Real Academia Española, *Corpus de Referencia del Español Actual (CREA)*. Disponible en: <https://www.rae.es/crea>.
- ESLORA = *Corpus para el estudio del español oral*. Disponible en: <http://eslora.usc.es>.
- Garcés, María Pilar (2008), *La organización del discurso: marcadores de ordenación y de reformulación*, Iberoamericana Vervuert.
- García Negroni, María Marta (2009), «Reformulación parafrástica y no parafrástica y *ethos* discursivo en la escritura académica en español: contrastes entre escritura experta y escritura universitaria avanzada», *Letras de Hoje*, 44 (1): 46-56.

- Martín Zorraquino, María y José Portolés (1999), «Los marcadores del discurso», en Ignacio Bosque y Violeta Demonte (coords.), *Gramática descriptiva de la lengua española*, Madrid, Espasa Calpe: 4052-4213.
- Méndez Orense, María (2016), «Valores pragmático-discursivos de la construcción lingüística *en plan*: ¿formación de un nuevo marcador?», *Philologia Hispalensis*, 30 (1): 123-144.
- NGLE = RAE y ASALE (2009), *Nueva gramática de la lengua española*, Madrid, Espasa.
- Núñez Paulina, Astrid Muñoz, y Estenka Mihovilovic (2006), «Las funciones de los marcadores de reformulación en el discurso académico en formación», *Revista Signos*, 39 (62): 471-492.
- Portolés, José (1998), *Marcadores del discurso*, Barcelona, Ariel.
- PRESEEA = *Proyecto para el estudio sociolingüístico del español de España y América*. Disponible en: <https://preseea.linguas.net>.
- Pons Bordería, Salvador (2013), «Un solo tipo de reformulación», *Cuadernos AISPI*, 2, 151-170.
- Pons Bordería, Salvador (2016), «Historia de *o sea*», *Boletín de la Real Academia Española*, 96 (313): 5-30.
- Rojo, Guillermo (2015), «Los corpus textuales del español», en Javier Gutiérrez Rexach (coord.), *Enciclopedia lingüística hispánica*, Londres, Routledge: 285-296.
- Rojo, Guillermo (2021), *Introducción a la lingüística de corpus en español*, Londres, Routledge.
- Sinclair, John McHardy (1991), *Corpus, concordance, collocation*, Oxford, Oxford University Press.
- Sinclair, John McHardy (2004), *Trust the text: language, corpus and discourse*, Londres, Routledge.
- Teubert, Wolfgang, y Anna Cermáková (2004), *Corpus linguistics: a short introduction*, London, Continuum.
- Val.Es.Co. = Pons Bordería, Salvador (dir.), *Corpus Val.Es.Co.* Disponible en: <http://www.valesco.es>.