

# El corpus ESLORA de espa1ol oral: un recurso para la investigaci3n en ELE

Victoria VÁZQUEZ ROZAS (Autor para correspondencia)

Universidade de Santiago de Compostela (Espan1a)

[victoria.vazquez@usc.es](mailto:victoria.vazquez@usc.es)

<https://orcid.org/0000-0001-8155-669X>

Marta BLANCO

Universidade de Santiago de Compostela (Espan1a)

[marta.blanco@usc.es](mailto:marta.blanco@usc.es)

<https://orcid.org/0000-0002-8635-6950>

**Resumen:** En esta contribuci3n se presenta el corpus *ESLORA* (<https://eslora.usc.es/>) y se describen, mediante ejemplos concretos, algunas de sus posibilidades de explotaci3n como recurso para la investigaci3n en espa1ol lengua extranjera (ELE). En primer lugar (§ 1.1), se contextualiza el proyecto y se resumen sus objetivos generales; en el § 1.2 se presentan las caracter1sticas de composici3n y dise1o del corpus, con especial atenci3n a aquellos aspectos directamente relacionados con sus posibilidades de explotaci3n para la investigaci3n lingüística; y en el § 1.3 se exponen las opciones que ofrece la interfaz de consulta para recuperar la informaci3n. La segunda parte incluye la descripci3n de varios casos pr1cticos de utilizaci3n del corpus para el estudio de fen3menos relevantes en la ense1anza y aprendizaje de ELE. El § 2.1 est1 dedicado a dos mecanismos derivativos, los afijos *-ito/a/os/as* y *super-*; el § 2.2 trata aspectos de la combinatoria sint1ctica de los verbos *ser* y *estar*; el § 2.3 se centra en el uso y valores de las part1culas discursivas; por 1ltimo, en el § 2.4 se recoge un ejemplo de trabajo colaborativo entre docentes e investigadores de acuerdo con la metodolog1a de investigaci3n-acci3n, que atiende a la expresi3n del futuro en espa1ol.

**Palabras clave:** explotaci3n de corpus; espa1ol oral; ELE

## Catal1n:

*El corpus ESLORA d'espanyol oral: un recurs per a la investigaci3n en espanyol llengua estrangera*

**Resum:** En aquesta contribuci3n es presenta el corpus *ESLORA* (<https://eslora.usc.es/>) i es descriuen, mitjanant exemples concrets, algunes de les seves possibilitats d'explotaci3n com a recurs per a la recerca en espanyol llengua estrangera (ELE). En primer lloc (§ 1.1), es contextualitza el projecte i se'n resumeixen els objectius generals; al § 1.2 es presenten les caracter1stiques de composici3n i disseny del corpus, amb especial atenci3n a aquells aspectes directament relacionats amb les seves possibilitats d'explotaci3n per a la investigaci3n lingüística; i al § 1.3 s'exposen les opcions que ofereix la interfície de consulta per recuperar la informaci3n. La segona part inclou la descripci3n de diversos casos pr1ctics d'utilitzaci3n del corpus per a l'estudi de fen3mens rellevants en l'ensenyament i l'aprenentatge d'ELE. El § 2.1 est1 dedicada dos mecanismes derivatius, els afixos *-ito/a/os/as* i *super-*; el § 2.2 tracta aspectes de la combinatoria sint1ctica dels verbs *ser* i *estar*; el § 2.3 se centra en l'ús i valors de les part1cules discursives; finalment, al § 2.4 es recull un exemple de treball col·laboratiu entre docents i investigadors d'acord amb la metodologia de recerca-acci3n, que at3n a l'expressi3n del futur en espanyol.

**Paraules clau:** explotaci3n de corpus; espanyol oral; ELE

**English:****ESLORA corpus of spoken Spanish: a resource for the research into Spanish as a foreign language (ELE)**

**Abstract:** This paper introduces the ESLORA corpus (<https://eslora.usc.es/>) and describes, by means of specific examples, some possibilities of its use as a resource for the research into Spanish as a foreign language (ELE). First (section 1.1), the project is contextualized and its main goals are summarized. In section 1.2, we introduce the structure and design of the corpus, paying special attention to those aspects that are immediately related to its use in linguistic research. In section 1.3, we present the possibilities the interface offers in gathering linguistic information. The second part of this contribution concerns the description of several practical cases of how the corpus can be used in ELE teaching and learning. Section 2.1 deals with two derivative mechanisms, the affixes *-ito/a/os/as* and *super-*. Section 2.2 is about certain features of the combinatorial syntax of *ser* and *estar*. Section 2.3 focuses on the use and the functions of discourse particles. Finally, in section 2.4, we present an example of teamwork between teachers and researchers in accordance with the action research methodology, which is about the way future tense is expressed in Spanish.

**Keywords:** corpus use; spoken Spanish; Spanish as a foreign language

## Introducción

La aportación que brindan la lingüística de corpus, como disciplina, y los corpus, como recurso, a la enseñanza-aprendizaje de lenguas extranjeras es hoy en día un hecho indiscutible. En la actualidad, se dispone de numerosos corpus de distinto tipo (oral, escrito, general, especializado, multilingüe, de aprendientes, entre otros), con un diseño que permite la recuperación de la información textual y contextual que contienen y con unas opciones de acceso cada vez menos restrictivas, todas ellas condiciones muy favorables para que el ámbito de la didáctica de lenguas se beneficie de su uso.

Este trabajo se centra en la utilidad de los corpus aplicada a la enseñanza del español como lengua extranjera, atendiendo a la vertiente investigadora. El recurso utilizado, el corpus ESLORA de español oral (<https://eslora.usc.es/>), responde a un diseño concreto que determina las opciones de uso y explotación en cuanto al área específica de estudio (nivel fónico, morfosintáctico, léxico y pragmático-discursivo, incluyendo la perspectiva sociolingüística); por otra parte, el tratamiento y anotación que han recibido los materiales que lo integran es de crucial importancia para su aprovechamiento, pues condiciona la información que puede ser objeto de búsqueda y recuperación, algo que resulta decisivo desde la perspectiva del corpus como fuente de datos para la investigación lingüística.

Siguiendo a Leech (1997), se pueden establecer tres líneas de confluencia entre corpus lingüísticos y enseñanza de lenguas. La primera remite al uso indirecto de los corpus como fuente de datos, la segunda se relaciona con el uso directo de los corpus en la práctica docente, mientras que la tercera hace referencia al desarrollo de corpus orientados a un aprovechamiento didáctico. De las tres líneas señaladas nos centraremos en la primera de ellas y aportaremos algunos ejemplos concretos de uso de ESLORA de acuerdo con ese enfoque.

## 1. El corpus ESLORA

### 1.1. Contexto y objetivos

El corpus *ESLORA* es un recurso integrado por entrevistas semidirigidas y conversaciones espontneas audiograbadas en Galicia entre 2007 y 2015. En su versi3n 2.1 (junio de 2022), la transcripci3n de los materiales, que se presenta alineada con el audio, alcanza 768 023 palabras ortogrficas, lematizadas y anotadas morfosintcticamente.

El corpus fue creado para poner a disposici3n pblica una muestra del espaol hablado en Galicia en condiciones adecuadas para su uso y explotaci3n en la investigaci3n lingüística. Con anterioridad no existía un corpus de materiales orales accesibles y representativos del espaol de esta comunidad; por una parte, debido a las dificultades intrínsecas a la elaboraci3n de un corpus oral, frente a la relativa rapidez con que se compilan actualmente los corpus escritos; y por otra, por la tradicional desatenci3n a las variedades del espaol en contextos bilingües, producto de la “ideología del estándar” (Lippi-Green 2012), que favorece a un patr3n normativo monolingüe.

Aunque el espaol convive con el gallego en Galicia desde finales del s. XV, cuando se impuso como lengua de la administraci3n del reino de Espaia, su expansi3n se aceler3 desde mediados del siglo XX, con la generalizaci3n de la escolarizaci3n obligatoria (en espaol) y la migraci3n progresiva del campo a los núcleos urbanos. El resultado es que actualmente el 70 % de los gallegos tiene el espaol como lengua habitual en algunas situaciones de comunicaci3n o en todas ellas<sup>1</sup>. A pesar de este uso extendido, el espaol de Galicia est3 escasamente documentado. La mayor parte de las referencias y descripciones disponibles est3n basadas en la observaci3n personal de los gramáticos y en textos en espaol (en su mayoría literarios) de autores gallegos. Adem3s, las menciones al uso del espaol en Galicia adoptan en general un enfoque prescriptivista que refuerza los estereotipos negativos hacia la variedad.

Es preciso subrayar que *ESLORA* naci3 como parte del macroproyecto PRESEEA, con el que comparte el objetivo de contribuir al conocimiento de la variaci3n geogrfica del habla en el ámbito hispánico. En ese marco, pretende asimismo documentar el uso oral atendiendo a la variaci3n social, lo que condiciona, como veremos, la distribuci3n de los materiales y las opciones necesarias para la recuperaci3n de los datos (cf. Barcala *et al.*, 2018; Vázquez Rozas *et al.*, 2020, para obtener informaci3n detallada sobre la construcci3n y características del corpus).

### 1.2. Diseo del corpus: composici3n, codificaci3n y anotaci3n

El diseo del corpus est3 determinado por los objetivos que persigue su elaboraci3n. En *ESLORA* los objetivos indicados en el § 1.1 se plasman en su composici3n y en la anotaci3n de los textos transcritos, un proceso que facilita su consulta posterior.

---

<sup>1</sup>Según los datos del IGE de 2018, el 24 % de los encuestados declara usar solo espaol, el 23 % más espaol que gallego, el 22 % más gallego que espaol, el 30 % solo gallego y el 1 % otras lenguas.

ESLORA contiene materiales obtenidos a través de dos tipos distintos de interacción, la conversacional coloquial y la entrevista semidirigida. Permite así el estudio de la variación asociada al registro y al género (Fernández Sanmartín, 2018, 2023). Además, siguiendo la configuración común a PRESEEA, el subcorpus de entrevistas está distribuido de forma equilibrada entre hablantes mujeres y hombres, tres grupos etarios (de 20 a 34, de 35 a 54, y de más de 54 años) y tres niveles educativos (con estudios primarios, medios y universitarios). Las grabaciones de entrevistas se completaron con las respuestas de los hablantes a un cuestionario con información sociolingüística y actitudinal. Las conversaciones, por su parte, llevan asociadas fichas que recogen los datos sociolingüísticos básicos y la información contextual necesaria para interpretar adecuadamente las situaciones de comunicación.

Los textos del corpus se enriquecieron mediante diferentes tipos de codificación, que dan acceso a la información que se considera relevante. Saber qué se ha codificado y con qué criterios es imprescindible para un buen aprovechamiento de los materiales. La Figura 1 recoge los tipos de codificación que incluye la versión 2.1 de ESLORA.

Figura 1

*Tipos de codificación del corpus ESLORA*



Se establece una primera distinción entre codificación no lingüística y codificación lingüística. La codificación no lingüística tiene un componente extratextual y otro intratextual. Se considera extratextual la indicación de los metadatos, que se introduce en la cabecera de los documentos en formato XML y abarca, entre otras informaciones, el tipo de discurso –conversación o entrevista– y las características sociolingüísticas y roles de los hablantes, es decir, edad, sexo, nivel educativo y, para las entrevistas, si es informante o entrevistador. La codificación no lingüística intratextual señala ciertos elementos o condiciones de los textos que se consideran relevantes tanto para su adecuada interpretación como para la recuperación de la información mediante las opciones de consulta. Las marcas incluidas

en ESLORA se refieren a la asignación de turnos de habla, solapamientos, alternancia de código, nombres propios anonimizados, risas, siglas, segmentos de discurso referido, palabras cortadas, fragmentos ininteligibles, etc<sup>2</sup>.

La codificación lingüística, por su parte, enriquece los textos con información léxica y gramatical, identificando los lemas y las unidades multipalabra y asignándoles la categoría y subcategoría gramatical que les corresponde. A las categorías gramaticales tradicionales se añaden las indicaciones de pausas y silencios, junto con los signos de puntuación de interrogación y admiración. En ESLORA, la anotación morfosintáctica se ha llevado a cabo con el etiquetador XIADA adaptado para textos orales en español y testado mediante su aplicación a un amplio corpus de entrenamiento<sup>3</sup>.

### 1.3. Recuperación de la información: la interfaz de consulta

El motor de búsqueda de ESLORA (<https://eslora.usc.es/search>) admite la combinación de los diferentes tipos de codificación presentados en el § 1.2 y facilita la construcción de búsquedas complejas sin requerir el conocimiento detallado de la larga lista de etiquetas de codificación y anotación empleadas en su elaboración. No obstante, para rentabilizar todas las posibilidades que ofrece la interfaz de consulta, es preciso conocer previamente (i) qué tipo de búsquedas pueden hacerse, (ii) qué tipo de resultados se obtienen, y (iii) qué opciones de exportación hay.

(i) La **Búsqueda** puede ser ortográfica o gramatical, y puede referirse a elementos únicos o a varias unidades contiguas o distantes. Si elegimos la opción de Elementos gramaticales podremos combinar las indicaciones de clases y subclases de palabras y categorías morfológicas con información léxica (lema) y ortográfica. Todos los términos de búsqueda, sean etiquetas morfosintácticas o lemas, son compatibles con los caracteres comodín ? y \*. El sistema de consulta incluye además los operadores booleanos OR (|) y NOT (!), que enriquecen notablemente el filtrado de los datos.

(ii) Los **Resultados** pueden obtenerse en forma de concordancias, frecuencias e inventarios. Las Concordancias KWIC (*Key Word in Context*) dan acceso al contexto ampliado y a los fragmentos de audio. Las Frecuencias proporcionan información cuantitativa general, tanto en términos absolutos como relativos (frecuencia normalizada por millón de formas), así como los datos parciales de los diferentes subcorpus virtuales derivados de los metadatos del corpus. El sistema permite, de esta manera, comparar la frecuencia relativa de uso de un elemento o combinación de elementos entre hablantes jóvenes y mayores, mujeres y hombres, en entrevistas y conversaciones, etc. La opción Inventarios ofrece la relación de elementos gramaticales o lemas distintos que cumplen las condiciones de la búsqueda, junto con el dato de frecuencia total y normalizada de cada uno.

<sup>2</sup> V. [https://eslora.usc.es/guide\\_oral\\_tags](https://eslora.usc.es/guide_oral_tags)

<sup>3</sup> Sobre XIADA v. <http://corpus.cirp.gal/xiada/>. El etiquetario morfosintáctico completo de ESLORA puede consultarse en [https://eslora.usc.es/guide\\_tags](https://eslora.usc.es/guide_tags).

Para ilustrar algunas de las opciones de Búsqueda y Resultado, en la Figura 2 vemos cómo obtener la lista de adjetivos registrados en el corpus con el sufijo *-ísimo/a/os/as*.

Figura 2  
Captura de la interfaz de consulta de ESLORA 2.1

The screenshot shows the search interface of ESLORA 2.1. The search criteria are: Corpus: Cualquiera; Tipo: Elem. gramaticales; Sensibilidad: Acentos, Mayúsculas. The result options are: Tipo de resultado: Inventarios; Ordenación: Total; Agrupar: Elemento gramatical. The search query is '\*ísim?|\*ísim?s' and the result shows 12 selected tags. The interface also includes filters for Age, From, To, Role, Level of studies, Sex, Speaker, and Search in. The search results are displayed as 1 to 50 of 132 (281) results, with a 'Selección de columnas' section for selecting columns to display.

La captura de pantalla de la Figura 2 muestra enmarcadas en rojo la opción de Elementos gramaticales en Tipo de búsqueda y la cadena *\*ísim?|\*ísim?s* en la casilla Elem. gramaticales junto con la selección de la categoría Adjetivo en el menú desplegable de la casilla Etiqueta (la aplicación indica el número de etiquetas seleccionadas, que depende de las variantes morfológicas de cada categoría). Otra manera de hacer la misma búsqueda es poner en la casilla de Etiqueta **A\***, formada por la A inicial de todas las etiquetas de adjetivo seguida del comodín \* para incluir todas las subcategorías y formas morfológicas de esta clase de palabras (v. *supra* nota 3 para la lista completa de etiquetas morfosintácticas).

En cuanto a la configuración de los resultados, enmarcados en verde, la Figura 2 presenta activada la opción de Inventarios, con Ordenación Total (de mayor a menor frecuencia) y agrupados por Elemento gramatical. En la parte inferior de la imagen se seleccionan los datos cuantitativos que interese obtener en cada caso (Selección de columnas), como la frecuencia absoluta y relativa y los metadatos disponibles (no activados en el ejemplo).

Los resultados se visualizan en la interfaz de consulta. Siguiendo con el mismo ejemplo, la Figura 3 representa las primeras 20 líneas de la relación obtenida con las condiciones de búsqueda de la Figura 2. La Figura 4 recoge las primeras líneas de la misma búsqueda y resultado de Inventarios, pero con Ordenación por Coincidencia

(alfabética, no Total) y agrupación por Lema (no por Elemento gramatical).

**Figura 3**

Lista parcial del inventario de elementos gramaticales de los adjetivos en  
–ísimo/a/os/as ordenados por frecuencia

1	<b>buenísimo</b>	24 / 12
2	<b>buenísima</b>	15 / 10
3	<b>carísimo</b>	12 / 9
4	<b>riquísimo</b>	11 / 5
5	<b>chulísimo</b>	7 / 2
6	<b>malísimo</b>	7 / 5
7	<b>malísima</b>	6 / 4
8	<b>buenísimos</b>	6 / 6
9	<b>grandísimo</b>	6 / 4
10	<b>riquísima</b>	5 / 3
11	<b>rarísimo</b>	5 / 4
12	<b>guapísimo</b>	5 / 4
13	<b>buenísimas</b>	5 / 4
14	<b>clarísimo</b>	5 / 5
15	<b>carísimos</b>	5 / 3
16	<b>contentísima</b>	4 / 2
17	<b>incomodísimo</b>	4 / 3
18	<b>graciosísimo</b>	3 / 3
19	<b>cariñosísima</b>	3 / 3
20	<b>facilísimo</b>	3 / 3

**Figura 4**

Lista parcial del inventario de los lemas de los adjetivos en  
–ísimo/a/os/as ordenados alfabéticamente

1	*	24 / 6
2	<i>aburrido</i>	1 / 1
3	<i>agobiado</i>	1 / 1
4	<i>alto</i>	5 / 5
5	<i>bajo</i>	2 / 1
6	<i>barato</i>	2 / 1
7	<i>bello</i>	1 / 1
8	<i>bonito</i>	1 / 1
9	<i>borracho</i>	1 / 1
10	<i>bueno</i>	50 / 25
11	<i>burra</i>	1 / 1
12	<i>cacho</i>	1 / 1
13	<i>cariñoso</i>	4 / 3
14	<i>caro</i>	17 / 11
15	<i>chulo</i>	8 / 2
16	<i>chungo</i>	3 / 3
17	<i>claro</i>	5 / 5
18	<i>colgado</i>	1 / 1
19	<i>cómodo</i>	1 / 1
20	<i>complicado</i>	2 / 2

(iii) La opción **Descargar** de la interfaz de consulta facilita la exportación de los resultados de Concordancias e Inventarios cuando se desea exportar los datos obtenidos en un formato directamente integrable en una base de datos o en una hoja de cálculo. La aplicación genera un fichero con campos delimitados por comas (.csv) o bien en formato de libro de Excel (.xlsx). Las opciones de descarga son especialmente útiles para organizar y filtrar los resultados de frecuencias absolutas y normalizadas de Inventarios en función de los parámetros que estructuran el corpus (metadatos), que hemos visto que se seleccionan en la parte inferior de la pantalla de búsqueda (v. *supra* Figura 2). Con estas opciones la aplicación aporta las funcionalidades de un diccionario de frecuencias dinámico que incluye las frecuencias absolutas y normalizadas de los elementos léxicos o gramaticales que interesen al usuario, tanto para todo el corpus como para los subcorpus que se pueden definir mediante los parámetros de tipo de texto y características de los hablantes.

## 2. El corpus ESLORA y la investigación en ELE

El corpus ESLORA constituye un recurso versátil como fuente de datos para investigaciones destinadas a la enseñanza y aprendizaje de ELE. Por una parte, da acceso al contexto lingüístico –sonoro y en transcripción ortográfica– de las unidades léxicas y gramaticales y sus combinaciones tal como se producen en un discurso oral informal. Por otro lado, proporciona información cualitativa y cuantitativa sobre el uso de dichas unidades y combinaciones por parte de hablantes de diferentes características sociales.

En los siguientes apartados se presenta una muestra de las múltiples posibilidades de explotación del corpus en el campo de los estudios sobre ELE. En el § 2.1, tomando como ejemplos el sufijo *-ito/a/os/as* y el prefijo *super-* se ilustran las ventajas que ofrece ESLORA para investigar el uso los procedimientos derivativos en el discurso oral. A continuación (§ 2.2), mostraremos algunas posibilidades de explotación sintáctica del corpus a través de las combinaciones de *ser* y *estar* con atributos adjetivos. El § 2.3 se centra en el uso y valores de las partículas discursivas y, finalmente, el § 2.4 describe un ejemplo de trabajo colaborativo entre docentes e investigadores sobre la expresión del futuro en español.

### 2.1. Mecanismos derivativos. El sufijo *-ito/a/os/as* y el prefijo *super-*

La lematización y anotación morfosintáctica de los materiales de ESLORA permite abordar los fenómenos morfológicos integrando aspectos léxicos, gramaticales y pragmático-discursivos. Además, el diseño del corpus da acceso a la distribución social de los procedimientos de formación de palabras, que revela en algunos casos la correlación existente entre ciertas características de los hablantes y sus elecciones lingüísticas. Ambas perspectivas se ilustran a continuación mediante los afijos *-ito/a/os/as* y *super-*.

*-ito/a/os/as* es el afijo diminutivo más frecuente y general en español y constituye, por tanto, un elemento del sistema que los aprendientes de esta lengua han de incorporar a su competencia morfológica. Tanto la diversidad de contenidos expresados por el diminutivo como las diferencias de uso constatadas entre dialectos, incluido el contraste con otros sufijos apreciativos, justifican el interés del tema en el campo de ELE (v. los trabajos con datos de corpus de Criado de Diego y Andión Herrero, 2016; Malaver, 2019; Hu, 2020). Ciertamente, las descripciones gramaticales generales (NGLE § 9) aportan información sobre los diferentes matices de contenido que manifiesta el diminutivo (tamaño o extensión, atenuación, ironía, intensificación) y sobre su empleo con diferentes clases de palabras, mayoritariamente sustantivos (*casa – casita*) y adjetivos (*pequeño – pequeñito*), pero también con algunos adverbios (*cerca – cerquita*) y pronombres (*todo – todito*). La bibliografía presta asimismo atención a la mayor extensión del diminutivo en variedades americanas, en las que se registra con una amplia gama de adverbios (*ahorita, despuesito*), con gerundios (*andandito*), interjecciones (*adiosito*) y ocasionalmente con posesivos (*tuyito*) y numerales (*unito*). Se

indica también que estos sufijos presentan variantes morfológicas dependientes de la base léxica (*cafecito*, *pueblecito*, pero también *pueblito*) y que algunas unidades o subclases se resisten a la combinación con el diminutivo, como ocurre con los adjetivos de relación (*anual*, *marino*, *familiar*).

Los estudios de corpus sobre diferentes variedades geográficas del español aportan una base empírica consistente al conocimiento de la sufijación apreciativa (Manjón Cabeza, 2016; Malaver y Paredes, 2020). En esta línea, es útil completar con datos de otros dialectos, en nuestro caso del español de Galicia, la información disponible sobre el sufijo *-ito*, puesto que conocer con más detalle su distribución efectiva en el discurso es esencial para facilitar el aprendizaje de ELE. La aportación novedosa de ESLORA, gracias a su interfaz de consulta, es dar acceso inmediato a datos cuantitativos y cualitativos del uso real de este y otros sufijos en el habla. Mediante las opciones de búsqueda se identifican rápidamente las unidades léxicas que aparecen con *-ito/a/os/as*, ordenadas por frecuencia o alfabéticamente, y también categorizadas según la clase de palabras, el género gramatical o el número. Además, cabe observar la distribución de los diminutivos en función de las características de los hablantes o el tipo de discurso.

La Figura 5 corresponde a las condiciones de búsqueda que recuperan las unidades léxicas registradas en el corpus con el sufijo en cuestión:

**Figura 5**  
Condiciones de búsqueda de elementos gramaticales con sufijo *-ito/a/os/as*

Elem. gramaticales			
<input type="text" value="*ito *ita *itos *itas"/>	<input type="text" value="!V*"/>	<input type="text" value="!*ito!*ita!*itos!*itas"/>	<input type="text" value="Palabra ortográfica"/>
372 etiquetas seleccionadas			

El uso del operador OR (|) junto con el comodín \* en la casilla de Elemento gramatical (la primera de la izquierda) en combinación con el operador NOT (!) en la casilla del Lema (la tercera por la izquierda) permite excluir cualquier lema cuya terminación coincida formalmente con el sufijo buscado (como *bonito*, *hábito*, *requisito*, etc.). Además, la búsqueda se restringe a las clases de palabras que no sean verbos, de ahí que en la casilla de Etiqueta (segunda por la izquierda) se utilice el operador de negación seguido de la etiqueta V y el comodín \*. Debe notarse que esta búsqueda deja fuera formas de participio sufijadas, que requieren una búsqueda específica con la etiqueta de participio (VP\*) en la casilla Etiqueta (o su elección en el menú desplegable), que según muestra la Figura 6, se completa excluyendo ciertos participios irregulares terminados en *-ito/a/os/as* (como *adscrito*, *escrito*, *frito*, *suscrito* y variantes) mediante el operador de negación y las cadenas *\*scrit\** y *frit\** en la casilla de Palabra ortográfica (la casilla de la derecha).

Figura 6

Condiciones de búsqueda de participios con el sufijo derivativo *-ito/a/os/as*

Elem. gramaticales

\*ito|\*ita|\*itos|\*itas    VP\*    Lema    !\*scrit!\*frit\*

6 etiquetas seleccionadas

Volver    Limpiar    Buscar

Resultados 1 a 18 de 18

Descargar

ienes razón meter te luego a el frío por la noche puede ser un poco <pausa/> hay que pasar di muy abrigadita ¿ no ? tendrías que ir

043\_hab1    había que andar bien <pausa/> bien abrochaditos bien vestiditos con la ropa planchadita <pausa/>

.C010\_hab2    además es de miedo y así estamos agarraditos <pausa/> es más román bueno romántica no es po

.038\_hab1    sí siempre ya dice mi hija ella <pausa/> a <pausa/> la tienen cantidad de arregladita y no gasta nada <pausa/>

La Figura 6 muestra también las primeras líneas de las concordancias resultantes de la búsqueda de participios sufijados, que proporcionan el contexto (ampliable) y dan acceso al audio de cada fragmento, una información necesaria para determinar la función comunicativa de cada ocurrencia. Por otra parte, como ya indicamos en § 1.3 supra, la interfaz de ESLORA facilita la obtención de los resultados cuantitativos agrupados por lemas, categorías o subcategorías gramaticales, según la elección del usuario. Si, para visualizar y exportar los resultados de la búsqueda recogida en la Figura 5 supra, escogemos en Resultado la opción de inventarios, con Ordenación Total (por frecuencia) y agrupamiento por Lema y Elemento gramatical (en ese orden), y pedimos valores tanto absolutos como relativos en la Selección de columnas, obtendremos automáticamente los datos organizados como en la Tabla 1, que muestra las filas de los ítems de mayor frecuencia (de un total de 1367 resultados).

Tabla 1

Lista parcial de lemas y elementos gramaticales registrados con el sufijo *-ito/a/os/as*, junto con su frecuencia absoluta y relativa (datos ordenados por frecuencia)

Lema	Elemento gramatical	Total: Valor absoluto	Total: Valor relativo (por millón)
un poco	un poquito	252	339
pequeño	pequeñito	59	79
pequeño	pequeñita	55	74
poco	poquito	46	61
pequeño	pequeñitos	30	40
cerca	cerquita	27	36
pequeño	pequeñitas	25	33
cosa	cositas	24	32
año	añositos	21	28
casa	casita	18	24
poco	poquitos	15	20
fulano	fulanita	15	20
joven	jovencita	14	18

rato	ratito	14	18
casa	casitas	11	14
trozo	trocito	11	14
trozo	trocitos	8	10
cosa	cosita	8	10
corto	cortito	7	9
chaval	chavalito	7	9
chaval	chavalitos	7	9
chapa	chapita	7	9
cigarro	cigarrito	7	9
vaso	vasito	6	8
fresco	fresquito	6	8
año	añito	6	8
rollo	rollito	6	8
grupo	grupito	6	8
pueblo	pueblecito	6	8
malo	malita	5	6

En cambio, si para la misma búsqueda de la Figura 5 escogemos los resultados de inventarios ordenados por Lema\_1 y agrupados por Categoría, Lema y Elemento gramatical (en ese orden), obtendremos una lista alfabética de los lemas con sus variantes flexivas agrupados por clase de palabras. La Tabla 2 muestra un fragmento de los resultados que incluye parte de la lista de lemas adjetivos (desde *pobre*), seguidos de los adverbios y los determinantes.

Tabla 2

Lista parcial de categorías, lemas y elementos gramaticales registrados con el sufijo *-ito/a/os/as*, con su frecuencia absoluta y relativa (ordenación alfabética)

Categoría	Lema	Elemento gramatical	Total: Valor absoluto	Total: Valor relativo (por millón)
Adjetivo	pobre	pobrecita	3	4
Adjetivo	pobre	pobrecitos	2	2
Adjetivo	raso	rasita	1	1
Adjetivo	redondo	redonditas	1	1
Adjetivo	redondo	redonditos	1	1
Adjetivo	relleno	rellenita	1	1
Adjetivo	sano	sanita	2	2
Adjetivo	sano	sanitos	1	1
Adjetivo	sencillo	sencillo	2	2
Adjetivo	solo	solita	1	1
Adjetivo	tranquilo	tranquilita	4	5
Adjetivo	tranquilo	tranquilito	4	5
Adjetivo	urbano	urbanita <sup>4</sup>	1	1
Adjetivo	viejo	viejecita	1	1
Adverbio	cerca	cerquita	27	36
Adverbio	despacio	despacito	1	1
Adverbio	en pequeño	en pequeñito	3	4
Adverbio	enseguida	enseguidita	2	2
Adverbio	mucho	muchito	1	1
Adverbio	pronto	prontito	4	5
Adverbio	rápido	rapidito	1	1
Adverbio	temprano	tempranito	3	4
Adverbio	un poco	un poquito	186	250
Determinante	poco	poquita	5	6
Determinante	poco	poquitas	2	2
Determinante	poco	poquito	15	20
Determinante	poco	poquitos	15	20

La recuperación de las palabras del corpus prefijadas con *super-* se realiza con opciones de búsqueda similares a las ya vistas para el sufijo diminutivo:

<sup>4</sup> El etiquetador morfosintáctico interpreta *urbanita* como diminutivo de *urbano*. El error ha de corregirse añadiendo el término al lexicón del sistema.

Figura 7

Condiciones de búsqueda de los elementos prefijados con *super-*

Elem. gramaticales

super\*      Etiqueta      !super\*      Palabra ortográfica

En este caso nos encontramos con un número reducido de ocurrencias (72) de adjetivos (*superchunga*, *superinteresante*), participios (*superarregladas*) y adverbios (*superbién*, *supercerca*, *superfluidamente*). Para los estudios de ELE resulta relevante identificar las bases léxicas preferidas por los hablantes nativos para la expresión prefijada. En la relación de lemas abundan los términos que denotan valoraciones subjetivas de actitudes y comportamientos: *superchunga* (4), *superarregladas* (3), *superinteresante* (3), *superguay* (3), *superamables* (2), *superbién* (2), *superpin-up* (2), *superespecial* (2), *supertal* (2), *superfea* (2), *superagradable* (2), *superchachi*, *superinfantil*, *superunidas*, *superestrecha*, *superfácil*, *superamable*, *superdivertidos*, *superpronto*, *superextraño*, etc.

Además de los inventarios y las concordancias, la interfaz de ESLORA proporciona información sobre la distribución sociolingüística de las unidades del corpus. Si elegimos como Resultado la opción de Frecuencia completa, obtenemos los valores cuantitativos absolutos y relativos del prefijo *super-* tal como muestra la figura 8:

Figura 8

Captura de los Resultados de Frecuencia completa de los elementos con el prefijo *super-*

Grupo de edad			Papel				
	Resultados	Documentos	Frec. norm.		Resultados	Documentos	Frec. norm.
<19	0 / 82	0 / 1	0 / millón	Audiencia	0 / 2210	0 / 24	0 / millón
19-34	54 / 282 396	16 / 70	191 / millón	Entrevistador	3 / 65 313	3 / 53	46 / millón
35-54	13 / 226 717	5 / 38	57 / millón	Informante	69 / 675 037	23 / 83	102 / millón
>54	5 / 232 563	5 / 29	21 / millón				
Desconocido	0 / 802	0 / 19	0 / millón				

  

Sexo			Nivel de estudios				
	Resultados	Documentos	Frec. norm.		Resultados	Documentos	Frec. norm.
Hombre	5 / 291 197	4 / 52	17 / millón	Universitarios	54 / 362 988	15 / 81	149 / millón
Mujer	67 / 451 363	22 / 81	148 / millón	Medios	6 / 180 681	5 / 25	33 / millón
				Primarios	12 / 196 386	6 / 19	61 / millón
				Desconocidos	0 / 2505	0 / 21	0 / millón

  

Corpus			
	Resultados	Documentos	Frec. norm.
Conversaciones	32 / 112 097	9 / 30	285 / millón
Entrevistas	40 / 630 463	16 / 53	63 / millón

Comprobamos así que la distribución de frecuencias es muy desigual: *super-* es empleado sobre todo por mujeres jóvenes con estudios universitarios mientras que apenas se registra en el habla de hombres, de personas del grupo de edad más avanzada y de estudios primarios y medios.

## 2.2. Combinatoria sintáctica. Atributos con *ser* y *estar*

Como ejemplo de la explotación de ESLORA para investigaciones sintácticas de especial interés en ELE mostraremos de qué forma el corpus puede ampliar nuestro conocimiento de algunas construcciones atributivas con *ser* y *estar* (v.

también Vázquez Rozas y Blanco, 2022). Es sabido que ambos verbos en su uso copulativo se combinan con adjetivos y participios, y que sus respectivas preferencias léxicas están condicionadas por la función caracterizadora de la construcción con *ser* frente al valor de estado episódico que expresa *estar*.

Las gramáticas suelen describir el contraste entre las construcciones con referencias a la casuística de los elementos en función de atributo. Señalan la asociación de *ser* con los adjetivos de relación (*¿Esa empresa es pública?*, ESLORA) y en general con aquellos que expresan estados no resultativos (*el trabajo es importante pero no es lo que me hace feliz*, ESLORA), frente a la preferencia de *estar* por los adjetivos de origen participial (*La gente que juega al golf está encantada y enganchadísima*, ESLORA). Las descripciones destinadas a aprendientes de español aportan asimismo explicaciones para el contraste de las combinaciones de ambos verbos con un mismo adjetivo (*María es/está feliz; La entrada es/está libre...*), además de indicaciones sobre algunas diferencias entre dialectos, en particular acerca de la mayor extensión de *estar* en variedades americanas (v. también NGLÉ § 37.9). Sin embargo, no es tan habitual que los manuales, e incluso estudios más específicos, basen las descripciones en un análisis sistemático del uso registrado en corpus, especialmente en corpus orales.

El corpus ESLORA no está anotado sintácticamente, pero ofrece posibilidades de recuperación de información sintáctica a través de su interfaz de búsqueda. Junto a las funcionalidades mostradas en los apartados anteriores, podemos realizar búsquedas por elementos contiguos o próximos para así modular el contexto de identificación de las unidades que participan en una construcción. En el caso que nos ocupa, teniendo en cuenta que los adjetivos o participios en función atributiva no se sitúan necesariamente justo a continuación del verbo, resulta útil el tipo de búsqueda de Elem. gramaticales próximos, que nos permite variar la distancia entre ambos constituyentes. Si en la casilla Distancia elegimos 1 obtendremos 1732 resultados, como vemos en la Figura 9<sup>5</sup>.

---

<sup>5</sup> Hay que tener en cuenta que no todos los casos resultantes de la búsqueda son construcciones atributivas, ya que algunos representan estructuras pasivas. Para facilitar el filtrado de los casos, conviene hacer búsquedas independientes para adjetivos y participios, pues solo estos últimos entran en ambas estructuras.

Figura 9

Búsqueda de la combinación del verbo *ser* seguido inmediatamente de un adjetivo o participio y resultados en forma de concordancias con ordenación por Palabra\_1

Elem. gramaticales próx.

Elemento gramatical Etiqueta ser Palabra ortográfica +

78 etiquetas seleccionadas

Distancia 1

Elemento gramatical A\*|VP+ Lema Palabra ortográfica +

18 etiquetas seleccionadas

Volver Limpiar Buscar

Resultados 1 a 50 de 1732

← 1 2 ... 35 →

1 Ir a la página Descargar

▶ SCOM_M31_038_hab1	no sea tanto que tú estás <pausa/> eh él está <ininteligible/> fastidiado <pausa_larga/>
▶ SCOM_H22_024_hab1	era era la verdad es que era encantador bueno yo creo que no terminé Medicina en su vida porque llevó más años en casa <pausa/>
▶ SCOM_M23_004_hab1	las afueras y eso <pausa/> uff <pausa/> yo no sé si sería capaz <pausa/> el tener que depender continuamente de el coche <pausa/> para venir <pausa/>
▶ SCOM_M21_041_hab2	porque si ella te era muy buena en su oficio también soy buena en el mío <pausa_larga/>
▶ SCOM_M33_005_hab1	hacer ciertas cosas <pausa/> y en un año <pausa/> es capaz de superar <pausa_larga/>
▶ SCOM_IH13_C010_hab2	bueno depende si son gente guay sí pero si son viejos no <ininteligible/> <ininteligible/>
▶ SCOM_H32_035_hab1	hoy es impensable pero <pausa/> de aquella pues era según paráramos a la señora María que tenía un pue
▶ SCOM_M12_020_hab1	re aquí <pausa/> es libre <pausa/> no no la gente no es tolerante para nada <pausa/>
▶ SCOM_E_07	y hacer un documental ¿ te resulta difícil <pausa/> es complicado ? <pausa/>
▶ SCOM_IM13_C008_hab1	eres capaz
▶ SCOM_H31_046_hab2	pues como eran pobres <pausa_larga/>

Si con la misma búsqueda de la Figura 9 aumentamos la distancia a 2 obtenemos 1548 casos, en gran parte representativos también de la construcción buscada, como (1a-e). Pero entre los resultados se registran algunas otras combinaciones que no responden a la estructura atributiva *ser* + adjetivo / participio, pues en (1f), (1g) y (1h) el núcleo del atributo es un sustantivo (respectivamente *cargo*, *Yugoslavia* y *carne*).

1. a. la gente que la gente era muy abierta
- b. después era todo abierto hacia la hacia hasta arriba hasta el techo
- c. era yo aficionada
- d. el cardenal no era muy aficionado a tener servicio
- e. antes era todo ajardinado
- f. era un alto cargo en una / en Madrid
- g. de aquella aún era la antigua Yugoslavia
- h. era carne fresca

A medida que aumentamos la distancia entre verbo y adjetivo / participio, las concordancias muestran una menor proporción de la estructura buscada. Con distancia 3 se obtienen 877 resultados, pero la estructura en cuestión es claramente minoritaria. A modo de ejemplo, véase la relación de (2), en la que solo (2f) y (2g) representan cláusulas con *ser* como predicado junto a un adjetivo o participio como núcleo de una frase adjetiva en función de atributo (nótese que las pausas también se consideran en el cómputo de la distancia).

2. a. era un tío abierto
- b. era una madera africana
- c. pues era / ciudadano argentino
- d. era muy muy buena niña
- e. pues esto era del famoso tirachinas
- f. para mí antes era mucho más difícil
- g. era / muy espabilado

La búsqueda por proximidad puede resultar útil en las distancias cortas, pero produce muchos resultados no deseados al aumentar la separación entre los elementos buscados, ya que no permite restringir los elementos que aparecen entre las dos unidades fijadas por el usuario. Para un filtrado más preciso es necesario condicionar las características del material situado entre el verbo y el núcleo atributivo: podemos eliminar de los resultados casos como (1f-h) si impedimos que el elemento intermedio entre verbo y adjetivo / participio sea un determinante o un sustantivo. La interfaz de consulta permite aplicar tal restricción usando el tipo de búsqueda por Elem. gramaticales y abriendo y rellenando con las condiciones oportunas las filas de casillas necesarias, como muestra la Figura 10:

**Figura 10**

*Búsqueda del verbo ser seguido de un elemento que no sea determinante ni sustantivo y de adjetivo o participio*

**Elem. gramaticales**

Elemento gramatical	V*	ser	Palabra ortográfica	-	+
78 etiquetas seleccionadas					
Elemento gramatical	ID*IN*	Lema	Palabra ortográfica	-	+
294 etiquetas seleccionadas					
Elemento gramatical	A* VP*	Lema	Palabra ortográfica	-	+
18 etiquetas seleccionadas					

Con la restricción indicada para el elemento intermedio, el número de casos se reduce de 1548 a 1126, de lo que se infiere que más de 400 ocurrencias del resultado inicial de distancia 2 no correspondían a la construcción pretendida. Ahora, con un filtrado más preciso, podemos visualizar y descargar los inventarios de las combinaciones más frecuentes, como el de la Figura 11, que corresponde a la ordenación por frecuencia (Total) de los lemas, o bien de alguna de las otras opciones de ordenación y agrupación que ofrece la interfaz (en la Figura 12 vemos una muestra de la ordenación alfabética por Lema\_3, el adjetivo o participio, con agrupación de Lema y Elemento gramatical):

Figura 11

Lista parcial del inventario de lemas de *ser* + elemento no determinante y no sustantivo + adjetivo / participio ordenados por frecuencia

1	<i>ser mucho bonito</i>	40 / 24
2	<i>ser mucho bueno</i>	36 / 24
3	<i>ser mucho difícil</i>	26 / 18
4	<i>ser mucho joven</i>	23 / 17
5	<i>ser mucho importante</i>	21 / 17
6	<i>ser mucho grande</i>	19 / 13
7	<i>ser más fácil</i>	19 / 16
8	<i>ser mucho pequeño</i>	17 / 13
9	<i>ser mucho fácil</i>	16 / 11
10	<i>ser mucho malo</i>	14 / 10
11	<i>ser mucho divertir</i>	14 / 2
12	<i>ser más pequeño</i>	12 / 10
13	<i>ser mucho agradable</i>	12 / 6
14	<i>ser más difícil</i>	11 / 8
15	<i>ser mucho raro</i>	10 / 8
16	<i>ser mucho complicar</i>	10 / 5
17	<i>ser mucho mayor</i>	9 / 9
18	<i>ser tanto fácil</i>	9 / 9
19	<i>ser mucho distinto</i>	8 / 5
20	<i>ser mucho interesante</i>	8 / 6

Figura 12

Lista parcial del inventario de lemas y elementos gramaticales de *ser* + elemento no determinante y no sustantivo + adjetivo / participio ordenados alfabéticamente por el adjetivo / participio

8	<i>ser es mucho muy abierto abierta</i>	1 / 1
9	<i>ser era mucho muy abrir abierta</i>	2 / 1
10	<i>ser era todo todo abrir abierto</i>	1 / 1
11	<i>ser eran más más abrir abiertos</i>	1 / 1
12	<i>ser es un poco un poco absurdo absurdo</i>	1 / 1
13	<i>ser es tanto tan aburrido aburrída</i>	1 / 1
14	<i>ser es más más accesible accesible</i>	1 / 1
15	<i>ser ser aceptable aceptable</i>	1 / 1
16	<i>ser es adecuado adecuado</i>	1 / 1
17	<i>ser fue un poco un poco adelantado adelantada</i>	1 / 1
18	<i>ser es como como adictivo adictivo</i>	1 / 1
19	<i>ser era yo yo aficionado aficionada</i>	1 / 1
20	<i>ser somos mucho muy aficionado aficionados</i>	1 / 1
21	<i>ser era mucho muy aficionar aficionado</i>	3 / 2
22	<i>ser es mucho muy aficionar aficionado</i>	2 / 2
23	<i>ser es mucho muy agobiante agobiante</i>	1 / 1
24	<i>ser fue mucho muy agobiante agobiante</i>	1 / 1
25	<i>ser es un poco un poco agotador agotador</i>	2 / 1
26	<i>ser es mucho muy agradable agradable</i>	7 / 5
27	<i>ser es nada nada agradable agradable</i>	1 / 1

Si nos interesa comparar las combinaciones vistas del verbo *ser* con el uso de *estar* en estructuras semejantes, es sencillo obtener la correspondiente información del corpus con búsquedas similares. Las Figuras 13 y 14 reproducen las capturas de parte de los listados obtenidos<sup>6</sup>.

<sup>6</sup> Se podrían refinar más las búsquedas eliminando la posibilidad de que el segundo elemento sea una preposición, para evitar casos como *estar + de pequeño*.

Figura 13

Lista parcial del inventario de lemas de estar + elemento no determinante y no sustantivo + adjetivo / participio, ordenada por frecuencia

1	<i>estar mucho contento</i>	11 / 10
2	<i>estar bien hacer</i>	6 / 5
3	<i>estar mucho unido</i>	5 / 2
4	<i>estar mucho meter</i>	5 / 4
5	<i>estar mucho rico</i>	4 / 3
6	<i>estar mucho cómodo</i>	3 / 3
7	<i>estar todo limpio</i>	3 / 3
8	<i>estar más tranquilo</i>	3 / 3
9	<i>estar mucho orgulloso</i>	3 / 3
10	<i>estar más controlar</i>	3 / 2
11	<i>estar más cómodo</i>	3 / 3
12	<i>estar de pequeño</i>	2 / 1
13	<i>estar todo lleno</i>	2 / 2
14	<i>estar mucho ocupar</i>	2 / 2
15	<i>estar mucho satisfacer</i>	2 / 1
16	<i>estar mucho controlar</i>	2 / 2
17	<i>estar totalmente descolocar</i>	2 / 1
18	<i>estar mucho oscuro</i>	2 / 2
19	<i>estar mucho pendiente</i>	2 / 2
20	<i>estar tremendo</i>	2 / 2

Figura 14

Lista parcial del inventario de lemas y elementos gramaticales de estar + elemento no determinante y no sustantivo + adjetivo / participio, ordenada alfabéticamente por el adjetivo / participio

5	<i>estar está también también abandonado abandonada</i>	1
6	<i>estar están así así abierto abiertas</i>	1
7	<i>estar estoy abrir abierto</i>	1
8	<i>estar está tanto tan absorber absorbida</i>	1
9	<i>estar estaban tanto tan acojonar acojonados</i>	1
10	<i>estar están bien bien acondicionado acondicionadas</i>	1
11	<i>estar está tanto tan acostumar acostumbrada</i>	1
12	<i>estar estaba bastante bastante acostumar acostumbrada</i>	1
13	<i>estar estaba más más acostumar acostumbrado</i>	1
14	<i>estar estaba mucho muy acostumar acostumbrada</i>	1
15	<i>estar estoy tanto tan acostumar acostumbrada</i>	1
16	<i>estar estaba en en activo activo</i>	1
17	<i>estar estaba mucho muy adaptar adaptado</i>	1
18	<i>estar estoy yo yo adaptar adaptada</i>	1
19	<i>estar estaba adscribir adscrita</i>	1
20	<i>estar estar a a afiliado afiliado</i>	1
21	<i>estar estaba estar estaba agarrar agarrado</i>	1
22	<i>estar está allí allí aglomerar aglomerada</i>	1
23	<i>estar estén así así agobiado agobiados</i>	1
24	<i>estar está menos menos agobiar agobiados</i>	1

Desde cualquiera de estos listados, podemos consultar automáticamente las concordancias correspondientes, que a su vez nos dan acceso al contexto extendido junto con la anotación morfosintáctica y al audio de cada fragmento. Las Figuras 15 y 16 visualizan, respectivamente, las concordancias de la estructura buscada con las combinaciones de lemas *estar mucho contento* y *estar bien hacer* de la lista de la Figura 13.

Figura 15

Concordancias del inventario de lemas *estar mucho contento*

1	▶ SCOM_M31_038_hab1	<b>sí estaba muy contento</b> <pausa/> sí <pausa/> y las <pausa/> las enfermeras dice que fabulosos
2	▶ SCOM_H22_021_hab1	el lado de la feria de ganados <pausa/> que también <b>estaba muy contento</b> <pausa_larga/>
3	▶ SCOM_IM13_C015_hab1	y <b>están muy contentos</b>
4	▶ SCOM_M11_044_hab2	:h profesoras <pausa/> <b>y genial</b> con ellas <pausa/> <b>y estoy muy contenta</b> <pausa_larga/>
5	▶ SCOM_M12_020_hab1	yo <b>estoy muy contenta en la</b> panadería <pausa/> sí sí <silencio/>
6	▶ SCOM_M23_018_hab1	<b>bien muy bien estoy muy contenta</b>
7	▶ SCOM_IM13_C016_hab3	<b>sí sí estoy muy contenta</b>
8	▶ SCOM_M22_019_hab1	<b>estoy muy contenta</b> <pausa/>
9	▶ SCOM_M23_004_hab1	no sé <pausa/> <b>estoy contenta</b> ¿ eh ? que yo aquí <b>estoy muy contenta</b> <pausa/>
10	▶ SCOM_M12_036_hab1	<b>pero el horario en sí estoy yo estoy muy contenta</b> <pausa/>
11	▶ SCOM_H22_021_hab1	<b>sí la verdad es que</b> estoy muy bien <b>estoy muy contento</b> <pausa_larga/>

Figura 16

Concordancias del inventario de lemas *estar bien hacer*

1	▶ SCOM_M31_045_hab1	<pausa/> <vacilación/> <pausa/> seis meses aparte que no <b>está bien hecho</b> no <pausa/> mire <ruido tipo="carraspeo"/> <pausa/> usted ahora <si
2	▶ SCOM_H13_014_hab1	y si no <b>está bien hecho</b> <b>pués</b> <pausa/> lo hablas con tus compañeros con tus jefes se mejora
3	▶ SCOM_H33_002_hab1	e decir otra cosa o sea oye pues no si está bien hecho pues <b>está bien hecho</b> <silencio/>
4	▶ SCOM_H33_002_hab1	nosotros tenemos que decir otra cosa o sea oye pues no si <b>está bien hecho</b> pues está bien hecho <silencio/>
5	▶ SCOM_IM33_C001_hab1	pero en las otras <b>estaban bien hechas</b> y estas no <pausa/> claro
6	▶ SCOM_M33_005_hab1	ra <pausa/> pero eran cómodos <pausa/> eran <pausa/> si <b>estaban bien hechos</b> ¿comprendes? si estaban <pausa/> renovados claro <pausa/> €

La comparación entre las estructuras atributivas con *ser* y con *estar* se beneficia asimismo del análisis de las concordancias correspondientes a aquellos adjetivos o participios que se combinan con ambos verbos, fácilmente identificables confrontando la relación de los inventarios combinatorios de cada verbo. Al disponer del contexto real de uso de las unidades es posible valorar cuantitativa y cualitativamente la funcionalidad del contraste entre *ser* y *estar* con atributos como *abierto*, *bueno*, *cansado*, *cómodo*, *fácil*, *feliz*, *jodido*, *libre*, *mayor*, *rico*, *seguro*, *soltero*, y otros registrados con ambos verbos en ESLORA.

Además de un mejor conocimiento de la combinatoria léxico-sintáctica de las cláusulas atributivas en el discurso oral, hay otros aspectos del uso de estas estructuras que pueden ser analizados en profundidad a partir del corpus, ya que este proporciona, por ejemplo, información precisa sobre la distribución de las formas del paradigma verbal, un dato relevante para la planificación didáctica de la enseñanza del uso de las formas verbales según su valor temporal y modal.

### 2.3. Uso y valores de las partículas discursivas

La interacción oral es una de las actividades comunicativas de la lengua que se debe contemplar en el proceso de enseñanza-aprendizaje de español/LE. La adquisición de esta destreza requiere que el aprendiente desarrolle una serie de habilidades no solo lingüísticas, sino también cognitivas, sociales, discursivas y pragmáticas.

Entre los numerosos elementos verbales y no verbales propios de la comunicación oral se encuentran las denominadas partículas discursivas, cuyo dominio por parte de los aprendientes suele constituir un área de dificultad. Las opciones que ofrece la interfaz de consulta de ESLORA permiten recuperar información cuantitativa (Figuras 17 y 18) y cualitativa (Figura 19) sobre el uso y valores de estos elementos que puede resultar de enorme utilidad para la investigación en ELE.

De acuerdo con el etiquetario morfosintáctico empleado, el acceso a estas unidades requiere que seleccionemos en la opción Tipo, Elementos gramaticales, y en Etiqueta (Categoría), Interjección. Si nos interesa saber cuáles son esas unidades y con qué frecuencia se emplean, debemos seleccionar al mismo tiempo en la opción Tipo de resultado, Inventarios, en la opción Ordenación, Total, y en la opción Agrupar, Lema, como se muestra en la Figura 17.

Figura 17

Búsqueda de las interjecciones ordenadas por frecuencia y resultado parcial en forma de Inventario

Búsqueda		Resultado	
<b>Corpus</b>	Cualquiera	<b>Tipo de resultado</b>	Inventarios
<b>Tipo</b>	Elem. gramaticales	<b>Ordenación</b>	Total
		<b>Agrupar</b>	Lema

Resultados 1 a 120 de 120 (31 684)

1	<i>bueno</i>	6092 / 83
2	<i>claro</i>	3961 / 82
3	<i>mmm</i>	3880 / 78
4	<i>eeh</i>	3704 / 81
5	<i>eh</i>	3249 / 81
6	<i>hm</i>	2012 / 79
7	<i>ah</i>	1389 / 82
8	<i>vamos</i>	753 / 71
9	<i>hombre</i>	665 / 69
10	<i>vale</i>	589 / 75

La aplicación también nos devuelve las unidades por orden alfabético, simplemente cambiando en la opción Ordenación, Total por Lema 1. Esta ordenación permite comparar el comportamiento de unidades similares, como por ejemplo *eeh*, *eemm*, *eh*, *ehm*.

Figura 18

Búsqueda de las interjecciones por orden alfabético y resultado parcial en forma de Inventario

Búsqueda		Resultado	
<b>Corpus</b>	Cualquiera	<b>Tipo de resultado</b>	Inventarios
<b>Tipo</b>	Elem. gramaticales	<b>Ordenación</b>	Lema 1
		<b>Agrupar</b>	Lema

34	<i>eeh</i>	3704 / 81
35	<i>eemm</i>	346 / 53
36	<i>eh</i>	3249 / 81
37	<i>ehm</i>	46 / 22

Además, la aplicación facilita el acceso a las concordancias desde las listas generadas pinchando sobre los valores numéricos (resaltados en rojo en las Figuras 17 y 18). Las concordancias dan acceso al contexto ampliado y a los fragmentos de audio, a partir de los cuales se podrían extraer datos sobre valores y usos (posición, sintaxis, prosodia y entonación) de las partículas discursivas, como se muestra en la Figura 19.

Figura 19

Captura parcial de las concordancias y del contexto ampliable del ejemplo 20

19. SCOM\_M23\_004\_hab1 a por completo a raíz de ahí ya <pausa/> mmm es otra <pausa/> otro planteamiento de vida ¿ eh ? <pausa/> y <pausa\_larga/>

20. SCOM\_H31\_043\_hab1 ya digo cambió mucho ¡ eh ! <pausa/> ya te digo me me gusta que cambie <pausa/> para bien <pausa\_larga/>

21. SCOM\_H31\_046\_hab2 y bueno me podía trab eh trabajar como ellos yo era el como si fuera el comodín <pausa\_larga/>

^ <ruido tipo="chasquido boca"/> <silencio/>

^ no sé <pausa/> no sé <pausa/> sí es verdad <pausa\_larga/>

^

ya digo cambió mucho ¡ eh ! <pausa/> ya te digo me me gusta que cambie <pausa/> para bien <pausa\_larga/>

ya	digo	cambió	mucho	¡	eh	!	<pausa/>	ya	te	digo	me	me	gusta	que	cambie	<pausa/>	para	bien	<pausa_larga/>
W	VIP1S	VIS3S	W	Q	I	Q	ETQ_PAUSA	W	PY2ESO	VIP1S	PY1ESO	PY1ESO	VIP3S	C	VSP3S	ETQ_PAUSA	X	W	ETQ_PAUSA
ya	decir	cambiar	mucho	¡	eh	!		ya	te	decir	me	me	gustar	que	cambiar		para	bien	

^ y pero bueno <pausa\_larga/>

^ sin hacer daño <pausa\_larga/>

## 2.4. La expresión del futuro

Una forma de aprovechamiento de los corpus como fuente de información responde a la obtención de muestras reales de lengua en las que sustentar la descripción de determinadas estructuras gramaticales, elementos léxicos (palabras, colocaciones) y unidades discursivas (marcadores, conectores). Ahora bien, el empleo de los corpus como fuente de información se puede abordar también desde la colaboración entre los agentes implicados en el proceso de enseñanza-aprendizaje de idiomas en la línea de la denominada *investigación-acción* (*action research*) (Kemmis y McTaggart, 1988; McNiff, 2013 [1988]; McNiff y Whitehead, 2006).

La muestra de explotación de ESLORA que se recoge en este apartado combina las dos aproximaciones antes mencionadas. Es el resultado de la fase inicial de una experiencia de colaboración entre docentes de ELE e investigadoras de lengua española y se centra en la expresión del futuro en español (*cf.* Blanco *et al.*, 2021).

El primer paso de la investigación consistió en buscar un área de estudio de interés común que permitiera compartir el conocimiento y combinar la experiencia de las investigadoras con la experiencia de los docentes. Se optó por trabajar con la expresión del futuro en español, concretamente con el futuro morfológico (FM), atendiendo al desequilibrio observado entre el tratamiento habitual de esta forma verbal en el aula de ELE y su actualización por parte de los hablantes nativos. En el segundo paso se consultó el PCIC (2006) y algunas gramáticas pedagógicas (Matte Bon, 2001; Martí Sánchez *et al.*, 2008; Borrego Nieto, 2013) y manuales de ELE (Equipo Prisma, 2005; Corpas *et al.*, 2013; Martínez *et al.*, 2020), con el fin de comprobar cómo abordan la descripción de este contenido gramatical. Se observó que los valores asociados al futuro se relacionan fundamentalmente con acciones futuras absolutas (*Estaré fuera de martes a jueves*), hipótesis y probabilidad (*Serán unos siete*), uso en frases interrogativas (*¿Será imbécil?*), futuro concesivo (*Estará lloviendo, pero yo voy a salir*), cortesía verbal atenuadora (*Veré qué puedo hacer*), futuro de mandato (*Obedecerás en todo lo que te digan*) y futuro impreciso (*Ya veremos*).

En el tercer paso se llevó a cabo el análisis de los datos arrojados por el corpus ESLORA. El

aprovechamiento futuro de los resultados obtenidos se materializará en un cuarto paso centrado en la elaboración de materiales a partir del uso tanto directo como indirecto del corpus. Su uso directo, como fuente de información, permitirá basar las descripciones y explicaciones didácticas sobre la expresión de futuro en datos reales y contextualizados. Su uso indirecto, como herramienta de aprendizaje, posibilitará el diseño de actividades en la línea del *Data-Driven Learning* (aprendizaje guiado por datos), que potencia la autonomía y la motivación de los aprendientes, ya que favorece una práctica lingüística significativa al trabajar con muestras auténticas de lengua. Por último, el quinto paso consistirá en el pilotaje de las actividades en un contexto de instrucción formal y requerirá la implicación activa de las docentes en el diseño de la herramienta para llevarlo a cabo, así como en la interpretación de la información obtenida y sus consiguientes implicaciones didácticas.

El examen detenido del uso oral contextualizado en contraste con las descripciones que ofrecen el PCIC y los manuales y gramáticas pedagógicas de ELE permitió no solo comprobar en el corpus los valores atribuidos al FM en la bibliografía sino también observar la productividad de los diferentes usos de la forma. De esta manera se introducen nuevos conocimientos empíricos en la práctica de la enseñanza de la lengua, que responde así de forma más adecuada a las condiciones y exigencias comunicativas reales. Además, el acceso inmediato al audio de las entrevistas y conversaciones que integran el corpus constituye una ventaja añadida para el reconocimiento de valores diversos en contextos dialógicos.

Como se muestra en la Figura 20, los usos del FM registrados de ESLORA responden a algunas de las funciones atribuidas al FM en manuales y gramáticas pedagógicas de ELE, pero el corpus no aporta ejemplos ni del futuro de mandato ni del futuro de cortesía verbal atenuadora. Esta ausencia puede considerarse indicativa de su menor extensión en el uso actual y, en consecuencia, puede determinar decisiones sobre la introducción didáctica de las funciones del FM.

Figura 20  
Expresión del futuro (FM)

PCIC, manuales y gramáticas pedagógicas de ELE	Datos de ESLORA
a) Acciones futuras absolutas	a) mañana <u>vendrá</u> por ejemplo y <pausa/> y <u>haré</u> eso <pausa/> y preparo la comida
b) Hipótesis y probabilidad	b) con los alumnos ¿qué tal? porque <u>habrá</u> alguno que <u>será</u> casi de tu edad
c) Uso en frases interrogativas	c) ¿qué le <u>pondré</u> ? <silencio/> siete <silencio/> ocho
d) Futuro concesivo	d) entonces digo bueno pues inglés no <u>sabré</u> la mandaré a clases pero creo que soy un buen ejemplo
e) Futuro impreciso (ya)	e) Ya me <u>informaré</u> cuando llegue el momento
f) Cortesía verbal atenuadora	f) Uso no documentado en el corpus
g) Futuro de mandato	g) Uso no documentado en el corpus

Por otra parte, los datos extraídos del corpus muestran la recurrencia de varias fórmulas semilexicalizadas

(entre paréntesis se indica la frecuencia absoluta): *ya verás* (14), *ya veremos* (4), *tú dirás* (7), *tú sabrás* (3), *como comprenderás* (3), *ya me contarás* (3).

El trabajo colaborativo de recuperación y análisis de los datos de ESLORA permite enriquecer la clasificación de valores del futuro que ofrecen los manuales con información empírica sobre la frecuencia y distribución de las funciones. Asimismo, el corpus facilita la comparación del uso del FM con el de la perífrasis *ir* + infinitivo, con la que comparte la expresión de valores temporales. La combinación de datos cuantitativos, extraídos de forma automática, con la clasificación semántica llevada a cabo manualmente de forma coordinada por todo el equipo, señala con claridad la asociación del FM con el sentido de conjetura o probabilidad y en menor medida con el valor temporal de posterioridad. Del análisis detallado realizado sobre una muestra parcial del corpus integrada por 338 casos de futuro y 528 de la perífrasis se concluye que el FM expresa posterioridad solo en el 28 % de los casos, frente al 69 % de usos temporales de la perífrasis. El corpus aporta, además, datos sobre la distribución de la persona en relación con la forma, morfológica o perifrástica, y con el sentido, temporal o modal. Así se comprueba que el FM se registra en 3ª persona de singular en el 45 % de los casos, frente al 18 % de la 1ª persona de singular, pero si se tienen en cuenta solo los casos de interpretación temporal, la 1ª persona es bastante más frecuente que la 3ª (46 % vs. 14 %). Por su parte, *ir* + infinitivo tiende a usarse en 1ª persona de singular (42 %) más que en 3ª (23 %), una diferencia que aumenta en el uso temporal de la perífrasis: 1ª persona de singular 50 % frente al 18 % de la 3ª.

## Consideraciones finales

A pesar de su indudable potencial para la investigación lingüística en la enseñanza y aprendizaje del español, hay que reconocer que a día de hoy el aprovechamiento de los corpus en ELE no es aún una práctica generalizada (pero v. Cruz Piñol, 2021; Vázquez Rozas y Blanco, 2022; Marcos Miguel y Bonilla, 2023, y referencias allí citadas). Para promover la difusión de la lingüística de corpus (y con corpus) en este campo deben cumplirse varias condiciones. En primer lugar, es fundamental disponer de suficientes recursos de acceso público que proporcionen muestras amplias y diversas del español, tanto escritas como orales, representativas de las variedades geográficas, sociales y de los diferentes registros de la lengua. Además, los materiales recopilados han de estar enriquecidos con la codificación y las anotaciones pertinentes que permitan un aprovechamiento completo y adaptable a diferentes finalidades. Para lograr este objetivo es preciso asimismo facilitar la recuperación de la información contenida en los corpus mediante motores de búsqueda potentes y aplicaciones de consulta claras e intuitivas. Por último, pero no menos importante, es necesario difundir e impulsar el conocimiento y el empleo de los corpus por parte de los agentes en la enseñanza de lenguas, tanto en su vertiente investigadora como pedagógica.

El diseño y construcción de ESLORA reúne las condiciones señaladas en el párrafo anterior y este trabajo responde, en particular, a la última dirección apuntada. Ciertamente, los ejemplos de utilización del corpus esbozados

en los diferentes apartados constituyen solo una muestra parcial de las m3ltiples opciones de explotaci3n que ofrece el recurso para los estudios de ELE. No obstante, esperamos que la contextualizaci3n, la presentaci3n de las caracter3sticas de composici3n y dise1o y la descripci3n de ejemplos concretos de explotaci3n del corpus sirvan para promover y facilitar su uso para la investigaci3n en espa1ol lengua extranjera.

## Agradecimientos

Este trabajo se ha realizado en el marco del proyecto de investigaci3n *El corpus ESLORA de espa1ol oral: enriquecimiento, an3lisis ling3istico y aplicaciones* (PID2020-118133GB-I00-P), financiado por MCIN/AEI/10.13039/501100011033.

## Referencias

- Barcala, Mario; Dom3nguez Eva; Fern3ndez, Alba; Rivas, Raquel; Santalla, M<sup>a</sup> Paula; V3zquez, Victoria; Villapol, Rebeca (2018). El corpus ESLORA de espa1ol oral. Dise1o, desarrollo y explotaci3n. *CHIMERA: Revista de Corpus de Lenguas Romances y Estudios Ling3ísticos*, 5(2), 217-237. <https://doi.org/10.15366/chimera2018.5.2.003>
- Blanco Dom3nguez, Marta; Barros Lorenzo, Roc3o; Paz Doce Rosa Ana (2021). Corpus y ELE: una experiencia pr3ctica colaborativa. *XXXI Congreso Internacional de ASELE. Creatividad, innovaci3n y diversidad en la ense1anza del espa1ol como LE/L2*, celebrado en la Universidad de Le3n, 1-3 de septiembre de 2021.
- Borrego Nieto, Julio (dir.) (2013). *Gram3tica de referencia para la ense1anza de espa1ol: la combinaci3n de oraciones*. Ediciones Universidad de Salamanca.
- Corpas, Jaime; Garmendia, Agust3n; Soriano, Carmen (2013). *Aula 3. Nueva Edici3n B1.1*. Difusi3n.
- Criado de Diego, Cecilia; And3n Herrero, Mar3a Antonieta (2016). Variaci3n y variedad del diminutivo en espa1ol a trav3s de dos corpus originales. Apuntes para su ense1anza como lengua extranjera. *RAEL: Revista Electr3nica de Ling3ística Aplicada*, 15(1), 87-108. <https://rael.aesla.org.es/index.php/RAEL/article/view/283>
- Cruz Pi1ol, Mar (Ed.) (2021). *E-research y Espa1ol LE/L2: investigar en la era digital*. Routledge. <https://doi.org/10.4324/9780429433528>
- Equipo Prisma. (2005). *Prisma Consolida C1*. Edinumen.
- ESLORA: *Corpus para el estudio del espa1ol oral*. <http://eslora.usc.es>
- Fern3ndez Sanmart3n, Alba (2018). La entrevista libre como m3todo para evitar la paradoja del observador. Un estudio de corpus. *CHIMERA: Revista de Corpus de Lenguas Romances y Estudios Ling3ísticos*, 5(2), 141-196. <https://doi.org/10.15366/chimera2018.5.2.001>
- Fern3ndez Sanmart3n, Alba (2023). *Teor3a y m3todos para la elaboraci3n de corpus orales: la entrevista socioling3ística*. [Tesis doctoral, Universidade de Santiago de Compostela].
- Hu, Jingyuan (2020). *La ense1anza del diminutivo en ELE a sinohablantes a trav3s de contenidos audiovisuales Espa1oles e Hispanoamericanos*. [Tesis doctoral. Universidad Complutense de Madrid]. <https://eprints.ucm.es/id/eprint/63410/>
- IGE: Instituto Galego de Estatística. <https://www.ige.gal>

- Kemmis Stephen; Robin McTaggart (1988). *The action research planner*. Deakin University Press.
- Leech, Geoffrey (1997). Teaching and language corpora: a convergence. En Anne Wichmann; Steven Fligelstone; Tony McEnery; Gerry Knowles (eds.), *Teaching and language corpora* (pp. 1-23). Longman.
- Lippi-Green, Rosina (2012). *English with an Accent: Language, Ideology and Discrimination in the United States*. (2ª ed.). Routledge.
- Malaver, Irania (2019). Diminutivos en el habla de aprendices de espaol. *MarcoELE: Revista de Didctica de Espaol Lengua Extranjera*, 29. [https://marcoele.com/descargas/29/malaver-diminutivos\\_aprendices.pdf](https://marcoele.com/descargas/29/malaver-diminutivos_aprendices.pdf)
- Malaver, Irania; Paredes, Florentino (2020). Convergences and divergences in the use of the diminutive in Medellin, Caracas and Madrid. En Isabel Molina Martos, Ana M. Cestero Mancera y Florentino Paredes Garca (eds.), *Sociolinguistic patterns and processes of convergence and divergence in Spanish*. [*Spanish in Context* 17, 2] (pp. 317-340). <https://doi.org/10.1075/sic.00061.mal>
- Manj3n-Cabeza Cruz, Antonio (2016). Los sufijos *-ico*, *-ito*, *-illo* en el espaol de Granada, Espaia. *Spanish in Context* 13(3), 421-436. <https://doi.org/10.1075/sic.13.3.05man>
- Martrnez, Daniel; Rodrguez, Marta; Ruiz, Juana; Su3rez, Elena (2020). *Agencia ELE 5. Nueva Edici3n B2.1*. SGEL.
- Matte Bon, Francisco (2001). *Gramtica comunicativa del espaol. Tomo 1 De la lengua a la idea*. (Nueva edici3n revisada). Edelsa.
- Marcos Miguel, Nausica; Bonilla, Carrie L. (2023). Utilizaci3n de corpus de espaol L1 en el aula de LE/L2/LH: ¿por qu3 y para qu3? *Journal of Spanish Language Teaching*. <https://doi.org/10.1080/23247797.2022.2157083>
- Mart3 S3nchez, Manuel; Penad3s Mart3nez, Inmaculada; Ruiz Mart3nez, Ana Mar3a (2008). *Gramtica espaola por niveles*. (Volumen 2). Edinumen.
- McNiff, Jean (2013 [1988]). *Action research. Principles and practice*. (3ª ed.). Routledge.
- McNiff, Jean; Jack Whitehead (2006). *All you need to know about action research*. Sage Publications.
- NGLE: RAE y ASALE (2009). *Nueva gramtica de la lengua espaola*. Espasa.
- PCIC: Instituto Cervantes (2006). *Plan curricular del Instituto Cervantes. Niveles de referencia para el espaol* (3 vols.). Biblioteca Nueva. [https://cvc.cervantes.es/ensenanza/biblioteca\\_ele/plan\\_curricular/default.htm](https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/default.htm)
- PRESEEA: Proyecto para el Estudio Sociolingüístico del Espaol de Espaia y de Am3rica. <https://preseea.uah.es/>
- V3zquez Rozas, Victoria; Barcala, Mario; Dom3nguez Noya, Eva; Fern3ndez Sanmart3n, Alba; Rojo, Guillermo y Santalla, Mar3a Paula (2020). Codificaci3n y anotaci3n del habla en un contexto bilingüe: el corpus ESLORA de espaol de Galicia. En 3ngel J. Gallego y Francesc Roca Urgell (eds.), *Dialectologia digital del espaol*, [Anexo nº 80 de *Verba*] (pp. 191-226). Servicio de Publicaci3n e Intercambio Científico da USC. <https://dx.doi.org/10.15304/9788418445316>
- V3zquez Rozas, Victoria y Blanco, Marta (2022). Corpus y enseanza del espaol. En Giovanni Parodi, Lewis C. Howe y Pascual Cantos (eds.), *Lingüística de corpus en espaol/ The Routledge Handbook of Spanish Corpus Linguistics* (pp. 342-356). Routledge. <https://doi.org/10.4324/9780429329296-26>