LA UTILIZACIÓN DE CORPUS SOCIOLINGÜÍSTICOS EN LA ENSEÑANZA DE LA LENGUA: EL CORPUS PRESEEA

Jorge Roselló Verdeguer
Universitat de València
jorge.rosello@uv.es

Resumen

Pese a que todavía haya algunas reticencias sobre su uso, los corpus lingüísticos son una buena herramienta en la enseñanza de idiomas. Se trata de muestras reales de la lengua en contextos auténticos y presentan muchas utilidades para las clases de español como lengua extranjera, ya que permiten comprobar las teorías lingüísticas y las intuiciones y extraer reglas a partir de la observación, además de abordar una diversidad de temas con los que practicar las destrezas básicas de la lengua.

Entre la variedad de corpus existentes en español, nos centramos aquí en los corpus de texto, normalmente orales con transcripción, en los cuales se reproducen conversaciones entre varias personas sobre diversos temas. La conversación es la forma prototípica de usar la lengua y, por ello, cualquier fenómeno lingüístico debería analizarse dentro de ella. El grupo PRESEEA (Proyecto para el Estudio Sociolingüístico del Español de España y América) ha creado corpus de lengua española representativos del mundo hispánico en sus distintas variedades geográficas y sociales. A través de ellos y de las investigaciones que ha generado podemos estudiar una multiplicidad de fenómenos lingüísticos y aplicarlos al aula, sabiendo que están contrastados y nos muestran su uso en diferentes contextos y situaciones. A modo de ejemplo, ofrecemos algunas propuestas sobre el leísmo, el uso del verbo impersonal *haber*, las formas de tratamiento o la expresión del sujeto pronominal en español.

1. INTRODUCCIÓN

En los años 90 del siglo XX, Sinclair, uno de los pioneros en el estudio de la materia que nos ocupa, consideraba el corpus como "una colección de piezas de un idioma que se seleccionan y ordenan de acuerdo con criterios lingüísticos explícitos con el fin de ser utilizados como muestras de lengua" (Sinclair, 1994: 4). Aunque, actualmente, la recopilación de textos (o fragmentos de textos) se concibe en soporte electrónico, lo que facilita la organización de grandes cantidades de datos, lo cierto es que la definición de Sinclair sigue siendo válida. En ella introduce un concepto clave en la lingüística de corpus: su finalidad, ya que el corpus, como indica Cruz Piñol (2012: 23), es una fuente de datos para investigaciones lingüísticas, es decir, "no constituye una meta en sí, aunque compilarlo es una labor que requiere una profunda reflexión lingüística".

Estas muestras de habla, que pueden ser orales o escritas, deben ser representativas. La representatividad es un tema polémico debido a que hay que prestar la atención necesaria a la variación geográfica, social y estilística del español. Así pues, esta variedad puede ser muy amplia (corpus que incluyen diferencias geográficas y sociales) o más concreta (representativo solo del español coloquial o del español formal). El tamaño de la muestra será, entonces, clave para medir su representatividad. Por consiguiente, un corpus, para que sea representativo, debe estar bien diseñado (por ejemplo, dividido por edades, grupos sociales, nivel de estudios…) y tener una representación mínima por cada grupo (Elvira-García, 2021). También debe recoger todas las estructuras posibles —fonéticas, morfológicas, sintácticas o léxicas— y no solo las que el investigador quiere encontrar. Es lo que Torruella y Llisterri (1999) llaman un "corpus neutro", aunque esto es más una tendencia que una realidad, pues el investigador siempre dirige la mirada o el pensamiento hacia aquello que, de una manera consciente o inconsciente, quiere demostrar.

Algunos estudiosos se han preguntado por qué, pese a ser una herramienta útil en las clases de español como lengua extranjera, son poco atractivos, tanto para el alumnado como para el profesorado. Esto puede ser debido, como indica Abad Castelló (2019, p.150), a que muchos profesores, ya sean veteranos o nuevos, no están familiarizados con el uso pedagógico de corpus lingüísticos o también a que algunos corpus presentan una interfaz árida y poco atractiva. De igual modo, en ocasiones, los ejemplos de los corpus, si bien genuinos, puede que se hayan extraído de un contexto muy amplio y ofrezcan poco interés a los estudiantes (Widdowson, 1978). No obstante, los nuevos corpus han mejorado las condiciones de acceso y recuperación de la información textual que contienen (Vázquez Rozas y Blanco, 2022, p.344) y cada vez son más los corpus que realizan un enforque netamente didáctico. Una muestra la encontramos en Campillos Llanos, Gozalo y Moreno (2007).

Los corpus presentan muchas utilidades para las clases de ELE. Mostramos, a continuación, algunas de ellas:

a) Permiten disponer de muestras reales de la lengua en contextos auténticos. En efecto, en algunos corpus podemos ver el uso de una determinada palabra o expresión en su contexto; en otros, podemos contemplar la conversación entera para comprender los usos y las estructuras; en los corpus de aprendientes podemos consultar los errores más frecuentes cometidos por los estudiantes…

b) Permiten poner a prueba y comprobar las teorías lingüísticas y las intuiciones. A veces los profesores corregimos determinadas formas lingüísticas porque pensamos que no existen o no se dicen exactamente así, pero, como pone de relieve Gras (2018), lo que ocurre es que no están en nuestro idiolecto, es decir, para nosotros no son gramaticales, pero se usan normalmente en otras variedades. El corpus se convierte así en un argumento de autoridad en discusiones que se pueden dar en el aula.

c) Permiten extraer reglas o descripciones lingüísticas a partir de la observación de la lengua. De este modo, fomentamos el aprendizaje por descubrimiento y damos un papel más activo al estudiante, que puede deducir él mismo una regla de uso mediante la observación sistemática de estructuras de la lengua.

d) Permiten obtener información cuantitativa y, por lo tanto, verificable. Para acreditar determinados fenómenos no basta con que aparezcan una sola vez (algunos casos de leísmo, el tuteo entre personas desconocidas, la pluralización del verbo *haber* en algunas zonas hispanohablantes), sino que han de ser constantes. La cuantificación aquí sí puede ser importante para explicar un fenómeno.

e) Los temas abordados son muy variados y facilitan su aplicación y adaptación al aula. Algunos corpus se estructuran en módulos temáticos (PRESEEA, COSER) y en ellos podemos encontrarnos situaciones en las que se pueden practicar las destrezas básicas de la lengua, como describir, argumentar o conversar sobre una amplia variedad temática (la familia, los amigos, las costumbres, etc.).

f) Desarrolla también el componente pragmático, con el aprendizaje de valores y comportamientos culturales (abrir y cerrar una conversación, despedirse, pedir disculpas, reformular lo dicho, sugerir o pedir algo, etc.). Por tanto, su utilidad va más allá de las estructuras gramaticales, ya que determinados corpus, sobre todo los que incluyen conversaciones, facilitan el aprendizaje de las funciones comunicativas, la capacidad de entender los significados e inferir sus sentidos y usos (Albelda y Fernández, 2006).

2. TIPOS DE CORPUS

Los corpus se suelen clasificar atendiendo a diferentes criterios. Sin ánimo de ser exhaustivos, podemos ofrecer la siguiente clasificación (Figura 1), que está en la línea de la realizada por la mayor parte de los especialistas.

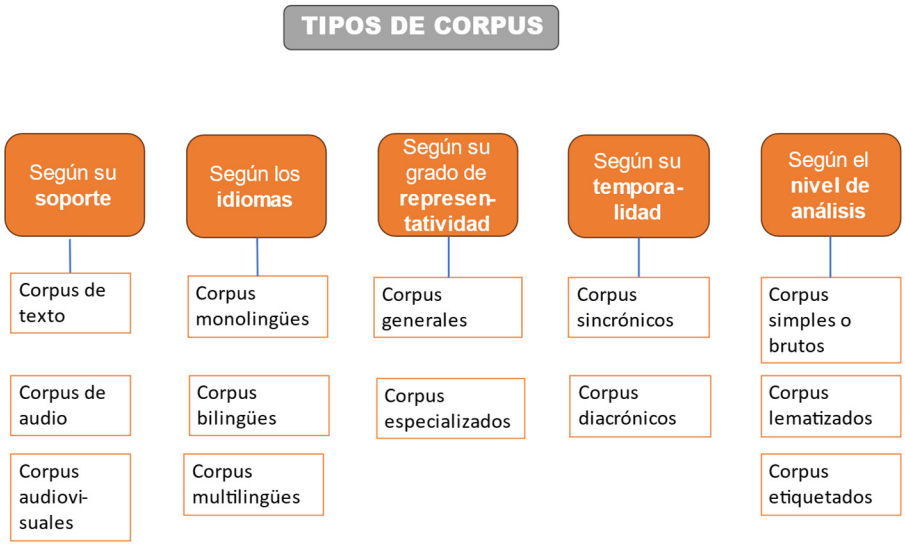

Figura 1. Tipos de corpus (elaboración propia a partir de distintos trabajos consultados)

Según el soporte o canal de producción original, los corpus pueden ser de texto (tanto orales como orales transcritos), de audio (orales con transcripción alineada) o audiovisuales (en vídeo, que sirven también para analizar los gestos).

Según su grado de representatividad, pueden ser generales, es decir, que contienen muestras de la lengua en general, y especializados. En los primeros se incluyen los corpus de referencia, como el CREA, el CORPES XXI o el Corpus del español, de Mark Davis (2002), cuyo objetivo es abarcar la mayor cantidad posible de muestras diferentes de la lengua para tratar de representarla en su totalidad. Están dotados de una interfaz de búsqueda en línea que permite a los usuarios hacer búsquedas sencillas para consultar frecuencias de palabras, frases que contengan algunas categorías gramaticales, colocaciones, etc.

Los corpus especializados se dirigen, sobre todo, a la elaboración de materiales relacionados con la lengua de especialidad (lenguaje económico, médico, jurídico, turístico…). También tendrían cabida aquí los corpus de aprendientes, que contienen datos de muestra real de los estudiantes de una determinada lengua y en los que podemos consultar los errores más frecuentes cometidos por las personas que están aprendiendo un idioma, y que nos puede ser de mucha utilidad en las clases de ELE. Uno de los más conocidos es el CAES, realizado por la Universidad de Santiago de Compostela con la ayuda del Instituto Cervantes (Rojo y Palacios, 2022).

Según los idiomas de los textos incluidos en el corpus, pueden ser monolingües, que recogen muestras en una única lengua, y bilingües o multilingües. Se incluyen aquí los llamados corpus paralelos, que contienen el mismo texto traducido a otra lengua (por ejemplo, una novela inglesa y su traducción al español). Si el corpus está alineado, el texto traducido se muestra junto con el original para que sea más fácil la comprensión. Este tipo de material es muy utilizado por los estudiantes para realizar determinadas actividades.

Según la temporalidad, pueden ser sincrónicos, si se centran en una etapa lingüística determinada, generalmente la actual, o diacrónicos, que son aquellos que contienen datos históricos, como el CORDE, que permiten ver la evolución en tiempo real de una palabra o estructura.

Por último, según su nivel de análisis, pueden ser simples o brutos, si no poseen ningún tipo de anotación; lematizados, en los que podemos encontrar, por ejemplo, cualquier forma verbal buscando el infinitivo, y etiquetados, que pueden contener distintos tipos de marcas, por ejemplo, morfológicas, que nos indica la categoría gramatical de cada palabra, o semánticas, para mostrar las diferentes relaciones semánticas —sinonimia, homonimia, antonimia…— que mantienen las palabras.

Se ha discutido si Google podría considerarse un corpus, pero, como pone de relieve Elvira-García (2021), Google no es propiamente un corpus, sino un buscador que contiene un conjunto digitalizado de datos (muy amplio, eso sí). El problema es que no está ordenado y que no podemos extraer informaciones que, tal vez, son necesarias para el usuario o para la investigación que se está realizando (autoría, edad, lengua original de quien escribió el texto, etc.).

La lingüística que hace uso de los corpus recibe la denominación de lingüística de corpus y es, propiamente, una metodología para poder llevar a cabo cualquier tipo de estudio lingüístico en cualquier nivel de la lengua (semántica, sintaxis…). Aunque no todos los estudiosos están de acuerdo en considerarla una metodología[1], lo cierto es que la lingüística de corpus proporciona datos que nos van a ser muy útiles, en nuestro caso, para enfocar y explicar determinados fenómenos lingüísticos que pueden ser problemáticos o controvertidos (por ejemplo, el leísmo, las formas de tratamiento en el mundo hispánico, la presencia o ausencia del sujeto gramatical, etc.). Su uso habitual en las conversaciones, su contextualización y su cuantificación sí que se pueden ver claramente en un corpus, y, sobre todo, en los de tipo conversacional. Por ello nos vamos a centrar en los del primer grupo de la clasificación reseñada en la Figura 1.

3. LA CONVERSACIÓN EN LOS CORPUS DE TEXTO

Los corpus de texto, orales con transcripción, se fundamentan en conversaciones o entrevistas entre varias personas. La conversación, como sabemos, es la forma prototípica de usar la lengua y, por ello, cualquier fenómeno lingüístico debería analizarse en el marco de la conversación. A través de esta interacción lingüística se adquieren la mayor parte de las estructuras sintácticas de la lengua, los elementos discursivos y los actos comunicativos lingüístico-pragmáticos.

La conversación se concibe no como una unidad lingüística superior a la oración, sino como una actividad comunicativa, un tipo de interacción social que posee una estructura y unas unidades propias e independientes. Su estructura es jerárquica, de manera que las unidades principales se van uniendo y combinando hasta formar la actividad comunicativa completa. La unidad básica principal es el turno de habla, que podemos definir como "el periodo de tiempo que comienza cuando una persona empieza a hablar y concluye cuando dicha persona deja de hacerlo" (Cestero, 2000, p.20). Otro concepto fundamental es de alternancia, esto es,

[1] Los pioneros de la lingüística de corpus la ven como una teoría. Así, Halliday (1994: 24) señala que "la lingüística de corpus pone en el mismo plano las actividades de recolección de datos y teorización". Por su parte, Tognini-Bonelli (2001:1) completa esta idea al señalar que la lingüística de corpus "ha adquirido un estatus teórico que la coloca en posición de contribuir de manera específica a otras aplicaciones". Para más información puede consultarse Zapata (2015).

cuando un hablante finaliza su mensaje y otro hablante toma la palabra para emitir el suyo, y que da lugar al cambio de turno.

La competencia conversacional se ha llevado también a la enseñanza de lenguas extranjeras, ya que es uno de los pilares básicos en la adquisición y la enseñanza de estas. Esto ha llevado a concienciar a los profesionales de la necesidad de trabajar con textos reales, tanto en el aula como en la elaboración de materiales y manuales, lo que ha propiciado que se utilicen los corpus lingüísticos, que posibilitan y facilitan la utilización de materiales discursivos auténticos (Cestero, 2005, p.72)[2].

Hemos de decir aquí que cualquier conversación nos parece adecuada en la enseñanza de ELE, incluida la de los textos literarios. Es cierto que en estos solo disponemos del texto escrito, pero, en muchas ocasiones, el autor literario, buen conocedor del lenguaje oral, reproduce con exactitud los giros de la lengua hablada, y todo ello en un texto depurado e idóneo para presentar a los estudiantes. Como muestra, ofrecemos este fragmento de la novela *Ojos de agua*, de Domingo Villar, con frases cortas y dinámicas de un intercambio lingüístico no exento de ironía.

—¿Nombre —había preguntado Estévez, dispuesto a rematar la tarea con prontitud.

—¿Mi nombre? —preguntó el chico.

—Claro, chaval, no vas a decirme el mío.

—Ya —concedió el joven traficante.

—Pues dime tu nombre.

—Francisco.

—¿Francisco algo?

—Francisco nada.

—¿No tienes apellido?

—Ah, Martín Fabeiro, Francisco Martín Fabeiro.

—¿Domicilio?

—¿Mi domicilio? —preguntó el joven.

Rafael Estévez alzó la vista

—¿Crees que quiero que me digas el mío? ¿Te parece que hemos venido a jugar a las adivinanzas?

—No, señor.

—Pues a ver si acabamos de una vez. ¿Cuál es tu domicilio?

Domingo Villar, *Ojos de agua*, Debolsillo, p.17.

El siguiente ejemplo corresponde al corpus Val.Es.Co, de tipo coloquial, que se caracteriza por la ausencia de toma de turno en los intercambios y por un nivel de formalidad mínimo de los hablantes, independientemente de su procedencia social (Pons y Ruiz Gurillo, 2005, p.244). El método de grabación es secreto, por lo que los informantes no son conscientes de ser grabados en el momento de la conversación, frente a otros corpus, como el PRESEEA, que veremos después, en que las entrevistas son semidirigidas y con micrófono a la vista. En las conversaciones de Val.Es.Co pueden participar dos o incluso más personas a la vez en la conversación, y, por ello las intervenciones habitualmente se solapan y, en ocasiones, se entienden con dificultad[3].

A: ya me lo diráas¿noo? Y yo síi síi tranquilos// qué alegría / es que nos ha- nos han hecho de rogaar↑ tíaa↑ las cinco y veinte que nos han d- dao las no[tas]

B: [y ¿qué] estábais todos allí esperandoo?§

C: §no↓ solo estábamos los que creíamos que estábamos apro(bados)

A: (( )) tía/ cuéntale lo del (( )) había↓ tenía uno↑ de teleco §

C: §eh- ese- ese que te he dicho yo que ha suspendido que se lo han dicho seguro↓et- ee- era de teleco↓ típico PAVITO/ de estos§

A: § mari como [el ingenier]

C: [de (( ))] como el INGENIER

A: [((Beniferrii/ Beniferrii))]

---

[2] Aunque hablamos de competencia "conversacional, algunos autores distinguen claramente entre conversación, sobre todo la de tipo coloquial, de otras formas dialogadas, como las entrevistas. La primera se caracteriza (Pons, 2022: 30) por "por poseer una toma de turno no predeterminada", además de desarrollarse entre hablantes que se conocen, con una relación de igualdad entre ellos, por no tratar temas especializados y tener lugar en un entorno familiar a los hablantes. Sin embargo, para nuestro propósito, lo fundamental es que su forma sea dialogada, se guarden o no los turnos. Además, como señala Terrádez (2019: 194), la entrevista semidirigida, como es el caso del corpus PRESEEA, comparte más rasgos con el lenguaje oral que con el planificado, además de recoger diferentes tipologías lingüísticas.

[3] Para las transcripciones de este y otros fenómenos se utilizan diferentes símbolos. Para saber su significado, ver Hidalgo y Sanmartín (2005).

C: [con una cara de- de pa-/ de PAVO→ y nada

A: (RISAS) (María José) ¿te acuerdas cuando fuimos a- a-?¿eras tú la que viniste(s) conmigo? [eeh (RISAS)]

B: [sí sí si] que no se enteraba

A: (( ))

C: es [igual]

A: [(( ))] lo vi pasar y nada es que me entró el ataque de [risa]

C: [pues eso] resulta que hacía cada animalada conduciendoo↑

Conversación 146 a Val.Es.Co.

4. EL CORPUS PRESEEA

En la propuesta que presentamos, vamos a utilizar el corpus PRESEEA (Proyecto para el Estudio Sociolingüístico del Español de España y América), ya que nos va a permitir reflexionar sobre determinados fenómenos gramaticales y pragmáticos del español que habitualmente se explican en las clases de ELE, fundamentados en los diversos estudios realizados por este grupo.



Figura 2. Captura de la página web de PRESEEA (https://preseea.uah.es/)

Como indica su página web (Figura 2), PRESEEA es un proyecto para la creación de un corpus de lengua española hablada representativo del mundo hispánico en su variedad geográfica y social. Esos materiales se reúnen atendiendo a la diversidad sociolingüística de las comunidades de habla hispanohablantes. Su corpus (Figura 3), de uso sencillo, está abierto y puede ser consultado por cualquier investigador.



Figura 3. Consulta del corpus del grupo PRESEEA

La metodología empleada, siguiendo a Moreno Fernández (2005 y 2021), se basa en entrevistas realizadas a hablantes residentes en núcleos urbanos monolingües o bilingües con una población hispanohablante bien asentada. En la Figura 4 se muestran las ciudades donde se asientan los equipos de investigación socio-

lingüística, tanto en España como en América. En algunos países de América, como el caso de Colombia y México, hay numerosos equipos en diferentes ciudades.

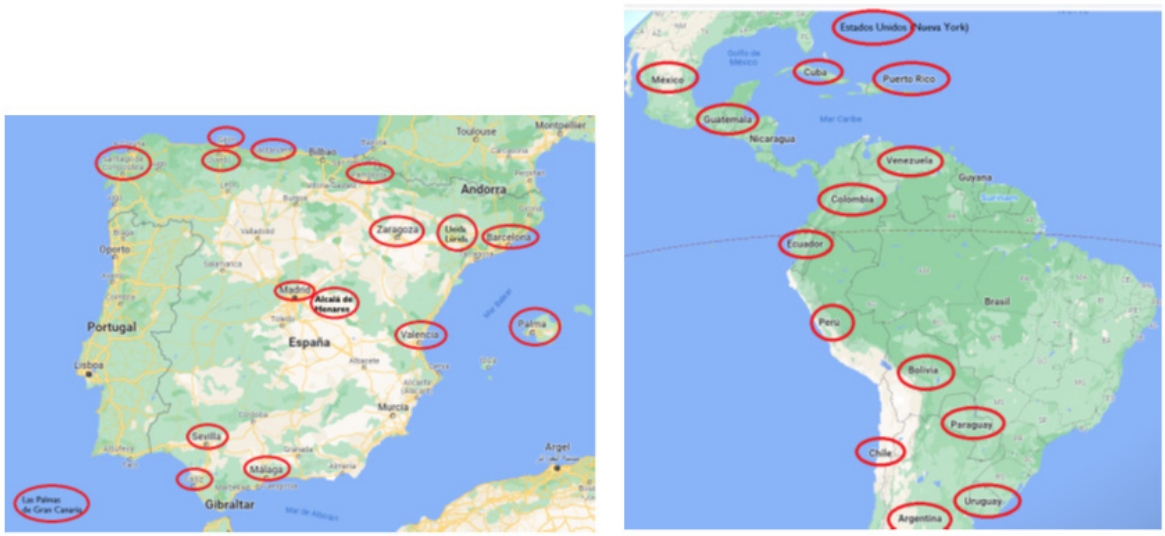

Figura 4. Lugares en los que se asientan los grupos de trabajo de PRESEEA

La selección de informantes para las entrevistas se hace siguiendo los criterios fijados en un muestreo por cuotas, creados a partir de tres variables: sexo/género, edad y nivel de instrucción o nivel sociocultural. En algunas comunidades bilingües también se incluye la lengua habitual o materna, que sirven para analizar las posibles interferencias[4]. Con respecto al tamaño de la muestra, los núcleos con un número de habitantes inferior a 500 000 suelen tener una muestra de 54 informantes, mientras que aumentaría a 72 en aquellas ciudades más pobladas.

Los materiales se recogen en conversaciones semidirigidas grabadas con algún dispositivo a la vista, en lugares de fácil acceso para los informantes y con buenas condiciones para su registro (es decir, ausencia de ruidos que dificulten después la interpretación de las palabras). Las entrevistas, con una duración aproximada de 45 minutos, se realizan proponiendo a los informantes módulos temáticos:

A) Saludos: ¿Cómo quiere que lo/la trate de tú o de usted? La verdad es que es un problema esto del tratamiento, nunca sabes cómo tratar a las personas, ¿verdad?

B) Lugar donde vive: ¿Dónde vives? ¿Cómo es tu casa? Descríbemela un poco. ¿Qué es lo que más te gusta de ella? ¿Llevas mucho tiempo viviendo allí? ¿Te gusta vivir aquí? ¿Dónde te gustaría vivir?

C) Familia y amistad: ¿Cómo conociste a tu pareja? ¿Piensas tener hijos? ¿Cómo debe ser una persona para que la consideres amiga? ¿A qué te dedicas? ¿Qué estás estudiando?

D) Costumbres: ¿Qué sueles hacer en vacaciones? ¿Qué harás en las próximas Navidades o en verano? ¿Cuál es la comida típica de Navidad en tu país? ¿Eres buen cocinero/a?

Los criterios y normas propuestos son un mínimo metodológico que busca la comparabilidad de los materiales reunidos en los distintos centros de investigación asociados al proyecto, aunque cada uno de ellos es libre de realizar otras investigaciones complementarias (actitudes lingüistas, ampliación de grupos etarios, etc.).

Antes de nombrar algunos de los estudios realizados por el grupo PRESEEA, hemos de realizar algunas precisiones terminológicas. La sociolingüística que se ha encargado de analizar estos fenómenos ha sido la llamada sociolingüística variacionista o laboviana (Labov, 1972), basada en la cuantificación y en el análisis probabilístico, lo que la aleja de los procedimientos empíricos intuitivos tradicionales. Labov considera que la lengua es variable y que la variación no es simplemente una mera manifestación de la diversidad consustancial a las actuaciones lingüísticas, sino que constituye una pieza esencial de su aprendizaje, adquisición y uso. La cuantificación es un elemento muy importante en el variacionismo, que ha tenido que desarrollar una metodología compleja y rigurosa (Moreno Fernández, 1994), diseñada específicamente para abordar el análisis de su objeto de estudio. Así pues, los principales atractivos de la cuantificación en el estudio de la regla variable son, por un lado, la posibilidad de obtención de medidas absolutas y relativas en la descripción de los fenómenos atendiendo a la influencia que ejercen una serie de factores contextuales lingüísticos y extralingüísticos sobre una determinada variable lingüística y, por otro, la elaboración de modelos estadísticos inferenciales predicti-

[4] También es posible atender a otras variables, como profesión, ingresos económicos, condiciones de alojamiento o 'modos de vida', pero se tratan como variables de post-estratificación.

vos que contribuyen a una rigurosa estimación del estado de un determinado hecho lingüístico variable y cuyos resultados pueden ser extrapolados a la población global de la que proceden las muestras de habla analizadas.

Detallamos aquí algunos de los fenómenos lingüísticos estudiados por el grupo y que pueden ser de aplicación en las clases de español, ya que muchos de ellos son conflictivos y plantean dudas a los estudiantes. Atendiendo a los trabajos realizados por los investigadores basados en los corpus de PRESEEA[5], estaremos en condiciones de explicar de una manera más precisa estas cuestiones gramaticales y pragmáticas, indicando su cuantificación, su extensión geográfica y los elementos sociales (edad, nivel de instrucción…) y lingüísticos (combinación con otras palabras, estructura en que se inserta, semántica del verbo…) que más los afectan.

- Realización de la /-s/ final de palabra o sílaba.
- Estudio de la /d/ intervocálica.
- Estudio de la /-d/ final de palabra o sílaba.
- Estudio del verbo impersonal *haber*.
- Estudio de la expresión del sujeto pronominal.
- Estudios sobre leísmo y laísmo.
- Estudios sobre las formas de tratamiento en español.
- Estudios sobre el uso del diminutivo.
- Estudios sobre la atenuación discursiva.

5. ALGUNOS EJEMPLOS DE UTILIZACIÓN DEL CORPUS

Aportamos aquí algunos ejemplos que consideramos pueden ser útiles para llevar al aula. **El leísmo**, sobre todo de persona (es decir, la utilización del pronombre *le* en lugar del etimológico *lo* para referirse al complemento directo: *A tu hermano le conocí ayer*) es, según indica la *NGLE* (2009, p.1215), un fenómeno generalizado en España, aunque infrecuente en el español de América. No obstante, según muestran diversos estudios de PRESEEA, no presentan la misma intensidad en todos los lugares de España (Roselló, 2017). Mientras en Madrid, el uso de *le* como complemento directo en el lenguaje oral alcanza el 24 %, en Málaga solo lo usa el 8,6 % y en Valencia, un 11,4 %. Además, este tipo de leísmo es más frecuente con verbos de afección, como *añorar*, *querer* (*A ese maestro le queríamos mucho*), y en determinadas estructuras, como las siguientes:

- Estructuras con verbo + complemento directo pronominal + predicativo: *Cuando le nombraron ministro, se transformó en otra persona*.
- Pronombre átono + verbo + infinitivo: *Miras por la ventana mientras le oyes hablar*.
- Se impersonal + pronombre átono: *A un amigo se le recibe en casa como a un familiar*.

Otra estructura interesante es la alternancia en singular y plural del **verbo impersonal *haber*** cuando va acompañado de un sintagma nominal en plural (*había /habían niños en el parque*), sobre todo porque en lenguas como la inglesa sí es correcta desde un punto de vista gramatical esta alternancia, según el sustantivo que acompañe al verbo sea singular o plural (*there was a chair / there were some chairs*). Esto puede crear algunas dudas al alumnado, sobre todo el de lengua inglesa, pues en algunas zonas de habla hispana es habitual el uso de este verbo en plural, incluso en el lenguaje escrito, como lo muestran estos ejemplos ofrecidos en la *NGLE* (2009, p.3063). “Como no habían medios de transporte el éxodo se intentaba a pie, en especial hacia Macaray” (García Márquez, *Cuando era feliz e indocumentado*); “Se separaron un poco para pasar sin despertar sospechas frente a la esquina cercana de la casa de Vela donde habían algunos agentes de seguridad, con radios” (Gioconda Belli, *La mujer habitada*).

La *NGLE* (2009, pp.3063-3064) reconoce que el fenómeno “está hoy en expansión, con intensidad algo mayor en América que en España”. Recomienda “en todos los casos el uso no concordado de haber”, y, por tanto, prefiere *Hubo dificultades*, en lugar de *Hubieron dificultades*; *Había suficientes pruebas para incriminarlos* y no *Habían suficientes pruebas para incriminarlos*. Tanto en el español europeo como en el americano “considera vulgar la forma *hayn*”. En los estudios realizados con los corpus PRESEEA, se observa que el uso de *haber* impersonal es alto en algunos lugares de España, como en Valencia, con el 46,2 % (Gómez Molina, 2013), y en Las Palmas de Gran Canaria, con el 36,2 % (Hernández Cabrera, 2016). En Valencia el plural es utilizado por todos los estratos sociales, incluido el alto, con estudios universitarios, y el nivel de aceptabilidad global hacia la pluralización es considerable (un 37 %). Sin embargo, en Las Palmas de Gran Canaria, el nivel alto restringe fuertemente la pluralización (que solo alcanza un 20,5 %). Esto indica que en esta comunidad este fenómeno no es prestigioso y puede estar sufriendo una regresión.

El uso de las **formas de tratamiento** en el español es otro de los temas interesantes para reflexionar en el aula. No hay que olvidar que uno de los primeros retos que se le plantea a la persona que está aprendiendo

[5] Algunos de los trabajos más significativos (los de Martín Butragueño, Cestero, Samper, Gómez Molina, Paredes, etc.) aparecen recogidos en el volumen de *Spanish in Context* coordinado por Molina, Paredes y Cestero (2020).

una lengua es la manera correcta de dirigirse a su interlocutor, ya que esta puede condicionar tanto su imagen como la conversación posterior. El correcto uso de las formas de tratamiento y su enseñanza también ha sido objeto de análisis por parte de los especialistas de español como lengua extranjera. García Aguiar (2009) llevó a cabo un análisis de una serie de gramáticas y manuales de español para extranjeros con el objeto de comprobar cómo se abordaba su estudio. Además de importantes consideraciones de carácter geográfico y dialectal, esta investigadora llegó a la conclusión de que factores como la edad, el desconocimiento y la clase social del interlocutor solían ser ignorados en los manuales. En los corpus de PRESEEA podemos encontrarnos con interesantes reflexiones sobre el uso de tú y usted, como se ve en el siguiente ejemplo.

E: bueno pues en primer lugar/ quería saber si tú ves a una persona de mi edad/ de mis características por la calle que es un desconocido/ ¿cómo lo saludas?

I: ¿cómo lo saludo?/ normalmente pues de usted

E: ¿por qué?

I: pues para mí es una pers- una persona desconocida// ee si más o menos es de mi edad// pues de usted/ si es más joven// naturalmente pues de- de tú// solamente por eso/ [hasta que coja-]

E: [o sea-]

I: hasta que puedas entrar un poco en confianza con él

E: o sea que es la edad [diríamos]

I:[sí sí/] es- es sobre todo la edad

Corpus PRESEEA-Valencia

Observando el corpus PRESEEA de Valencia, Roselló (2018), indica que el tuteo está generalizado en las conversaciones. El factor que más frena el uso de *tú* es la edad, ya que los mayores de 55 años se inclinan más por utilizar *usted* en las conversaciones, mientras que los jóvenes son los que más tutean, pese a que la edad del entrevistador es mayor que la suya. Por su parte, el entrevistador, mayor de 55 años, varón y de nivel sociocultural alto, utiliza más *tú* que *usted*. Tutea siempre a los jóvenes y a los que son de su mismo nivel sociocultural. La edad y el nivel sociocultural son, como vemos, factores fundamentales, aunque hay también otros aspectos que influyen, como la clase de actividad que se está realizando, el nivel de conocimiento de la otra persona o el tono de la interacción, que hace que se pueda cambiar de un trato más formal a otro más coloquial (por ejemplo, cuando se tratan temas más livianos, como "salir de marcha" o el ocio nocturno).

La determinación de los factores que condicionan la **expresión del sujeto pronominal** en español y cómo debe abordarse su enseñanza a estudiantes extranjeros también es un tema recurrente en los estudios de ELE, sobre todo en aprendientes anglófonos, que muestran dificultades para adquirir un parámetro ajeno a ellos como es el sujeto nulo (*Voy a clase todos los días* / *Yo voy a clase todos los días*). Los criterios que dan los manuales para su utilización son, a menudo, confusos, sobre todo porque los estudiantes no hacen un uso agramatical del pronombre sujeto, sino inapropiado, ya que lo utilizan en contextos donde un nativo, normalmente, no haría uso de él. Por ello, los estudios realizados por el grupo PRESEEA basados en sus diferentes corpus pueden darnos criterios cuantitativos y estadísticos sobre sus usos más habituales, ofreciendo a los estudiantes unas orientaciones claras para su utilización en contextos reales. Viendo los diferentes trabajos sobre el tema, lo primero que llama la atención es la variabilidad en su uso, ya que su presencia oscila entre más de un 45 % en algunos lugares de América (en Barranquilla, Colombia) a apenas un 17 % en Granada, España. Si dejamos a un lado determinados usos en que su presencia (caso de frases idiomáticas, como *Yo que sé* o usos enfáticos del tipo *Yo mismo lo hice*) o ausencia (caso de verbos impersonales como *haber* o cláusulas con *se* impersonal) es obligatoria, podemos decir, resumiendo, que la presencia del sujeto es mayor en los siguientes casos (Roselló, 2021):

- Cuando aparecen formas verbales ambiguas, sobre todo el pretérito imperfecto de indicativo (*Yo tenía dos pisos en*…; *Yo vivía en esa ciudad*…).
- Con verbos asociados a procesos mentales (*pensar, opinar, creer*) que, normalmente, aparecen en situaciones en donde los hablantes tienen que argumentar sobre alguna cuestión.
- El cambio de turno en la conversación suele llevar aparejada una mayor frecuencia en el uso del pronombre.
- Si el sujeto no es mencionado en la oración precedente, tiende a explicitarse en la siguiente intervención, lo cual no ocurre con la misma frecuencia si el sujeto es el mismo.

6. CONCLUSIONES

Así pues, hemos visto cómo los corpus pueden ser un buen material para llevar al aula de español para extranjeros, pese a que no todos se utilizan de la misma manera ni sirven para los mismos fines. Aquí nos hemos centrado en los corpus de texto, orales con transcripción, puesto que presentan materiales reales extraídos de conversaciones o diálogos entablados por hablantes nativos. En ellos, además de contextualizar la lengua,

podemos fijarnos en determinados elementos lingüísticos que pueden resultar conflictivos en el aprendizaje y que necesitan explicaciones que vayan más allá de la pura gramaticalidad o agramaticalidad del fenómeno. Por eso nos hemos centrado en PRESEEA, ya que sus materiales permiten la comparación en diversos lugares del mundo hispánico y, al utilizar la misma metodología, extraer conclusiones que nos puedan servir para explicar algunos usos lingüísticos.

Es cierto que aquí, por falta de espacio, nos hemos centrado en aspectos gramaticales, pero el grupo también ha estudiado fenómenos fonéticos interesantes para las clases de ELE, como puede ser la elisión del fonema /d/ en posición intervocálica (*olvidao, señalao*), u otros de tipo pragmático, como los mecanismos de atenuación existentes en el idioma para lograr determinados propósitos, el uso de los diferentes marcadores discursivos, etc.

7. REFERENCIAS BIBLIOGRÁFICAS

Abad Castelló, M. (2019). Uso de corpus lingüísticos por y para profesores de español como lengua extranjera. *redELE*, 31, 1-20. Recuperado de: https://www.educacionyfp.gob.es/dam/jcr:39778ef0-dddf-422f-b3c2-a8f08730225b/07-uso-de-corpus-linguisticos.pdf.

Albeda, M. y Fernández, M.J. (2006). La enseñanza de los registros lingüísticos en E/LE. Una aplicación a la conversación coloquial. *MarcoELE. Revista de Didáctica Español como Lengua Extranjera,* 3, 1-31. Recuperado de: https://marcoele.com/descargas/3/albelda-fernandez_registro-coloquial.pdf.

Campillo, L. Gozalo, P y Moreno, A. (2007). El corpus C-ORAL-ROM en la enseñanza de ELE. *Actas del XVII Congreso Internacional de ASELE: Las destrezas orales en la enseñanza del español* (pp. 1115-1128), Logroño: Servicio de Publicaciones de la Universidad de La Rioja. Recuperado de: https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/17/17_1115.pdf.

Cestero, A.M. (2000). *El intercambio de turnos de habla en la conversación. Análisis sociolingüístico*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá.

Cestero, A.M. (2005). *Conversación y enseñanza de lenguas extranjeras*. Madrid: Arco/Libros.

Cruz Piñol, M.C. (2012). *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: Arco/ Libros.

Davies, M. (2002). Corpus del Español: 100 million words, 1200s-1900s. Recuperado de: http://www.corpusdelespanol.org.

Elvira-García, W. (2021). *Uso de corpus en clase de ELE. La lengua real como modelo*. Barcelona: Difusión.

García Aguilar, L.C. (2009). Los sistemas de tratamiento en la enseñanza de E/LE. *FIAPE. III Congreso Internacional: La enseñanza del español en tiempos de crisis.* Cádiz: Universidad de Cádiz, 1-10. Recuperado de: https://www.educacionyfp.gob.es/dam/jcr:b966869d-cf53-4e3c-9749-33737c58b4d-b/2010-esp-09-07garcia-pdf.pdf.

Gómez Molina, J. R. (2013). Pluralización de *haber* impersonal en el español de Valencia (España). *Verba*, 40, pp.253-284. Recuperado de: https://revistas.usc.gal/index.php/verba/article/view/1292.

Gras, P. (2018). Más allá de las gramáticas y los diccionarios, los corpus como herramientas en la enseñanza/aprendizaje de lenguas. *MarcoELE. Revista de Didáctica Español como Lengua Extranjera*, 26, pp. 1-5. Recuperado de: https://marcoele.com/descargas/26/resena-cruz.pdf.

Halliday, M.A.K. (1994). *An Introduction to Systemic Functional Grammar* (2.ª ed.). London: Edward Arnold.

Hernández Cabrera, C. (2016). Variación de *haber* impersonal en el español de Las Palmas de Gran Canaria, *ELUA*, 30, pp. 141-162. http://dx.doi.org/10.14198/ELUA2016.30.07.

Hidalgo, A. & Sanmartín, J. (2005). Los sistemas de transcripción de la lengua hablada. *Oralia*, 8, pp. 13-36. Recuperado de: https://ojs.ual.es/ojs/index.php/ORALIA/article/view/8325.

Molina Martos, I, Paredes García, F & Cestero Mancera, A.M. (2020). Sociolinguistics Patterns and processes of convergence and divergence in Spanish, *Spanish in Context,* 17, p. 2 (Special issue).

Moreno Fernández, F. (2005). Corpus para el estudio del español en su variación geográfica y social. El corpus PRESEEA, *Oralia*, 8, pp. 123-139. Recuperado de: https://ojs.ual.es/ojs/index.php/ORALIA/article/view/8333.

Moreno Fernández, F. (2021). Metodología del "Proyecto para el estudio sociolingüístico del español de España y América" (PRESEEA). *Documentos PRESEEA de investigación, Documentos PRESEEA de trabajo, 1*». Alcalá de Henares, Editorial Universidad de Alcalá. https://doi.org/10.37536/preseea.2021.doc1.

Pons, S. & Ruiz Gurillo, L. (2005). Corpus para el estudio de la conversación coloquial. El corpus Val.Es.Co. (Valencia. Español coloquial). *Oralia*, 8, 243-263. Recuperado de: https://ojs.ual.es/ojs/index.php/ORALIA/article/view/8362.

Pons, S. (2022). *Creación y análisis de corpus orales: saberes prácticos y reflexiones teóricas*. Berlín: Peter Lang.

RAE y ASALE (2009). *Nueva gramática de la lengua española*, Madrid: Espasa.

Rojo, G. & Palacios. I.M. (2022). Los corpus de aprendientes de español como L2. En Parodi, G., Cantos, P. & Howe, C. (Coord.). *Lingüística de corpus en español* (pp. 74-88). Reino Unido: Routledge.

Roselló, J. (2017). Las formas de tratamiento en el corpus PRESEEA-Valencia. Un estudio sociolingüístico, *Círculo de Lingüística Aplicada a la Comunicación*, 76, 241- 260. http://dx.doi.org/10.5209/CLAC.62507.

Roselló, J. (2018). Los pronombres clíticos en el español hablado de Valencia. Un estudio sociolingüístico. *Diálogo de la lengua*, 10, pp. 18-55. Recuperado de: https://www.dialogodelalengua.com/articulo/pdf/10/2_Rosello_Verdeguer_DL_2018.pdf.

Roselló, J. (2021). Factores determinantes en la expresión del sujeto pronominal en español hablado de Valencia», *LinRed*, 17, pp. 1-24. Recuperado de: https://linred.web.uah.es/articulos_pdf/LR-articulo-16012021.pdf.

Terrádez, M. (2019). El léxico del español hablado: diferencias y semejanzas entre el registro planificado y no planificado. En Cabedo, A. & Hidalgo, A. (Eds.). *Pragmática del español hablado. Hacia nuevos horizontes*. Valencia: Universitat de València.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work (Studies in Corpus Linguistics: 6)*, Amsterdam/Atlanta, GA: John Benjamins.

Torruella, J. y Lliterri J. (1999). Diseño de corpus textuales y orales. *Filología e informática: nuevas tecnologías en los estudios filológicos*, pp. 45-81. Recuperado de: http://liceu.uab.es/~joaquim/publicacions/Torruella_Llisterri_99.pdf.

Vázquez Rozas, V. y Blanco, M. (2022). Corpus y enseñanza del español. En Parodi, G, Cantos-Gómez, P y Howe, C. (Coord.). *Lingüística de corpus en español* (pp. 342-356). Reino Unido: Routledge.

Widdowson, H. (1978). *Teaching Language as Communication*. Oxford: Oxford University Press.

Zapata, C. (2015). La lingüística de corpus y su incidencia en la enseñanza de lenguas extranjeras, *Educare*, 19, pp. 53-75. Recuperado de: https://revistas.investigacion-upelipb.com/index.php/educare/article/view/114.