



Diseño de corpus específicos para el estudio histórico gramatical: el caso de las construcciones con clítico femenino

The creation of specific corpora for the historical study of grammar: the case of constructions with the feminine clitic

Nicolás Arellano

Universidad de Buenos Aires (UBA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires / Argentina

nicolas.a.arellano@gmail.com

<http://orcid.org/0000-0002-5197-5428>

Resumen: Este artículo busca analizar las ventajas de una aproximación a los datos lingüísticos históricos a partir de la confección de corpus específicamente diseñados. Para ello, en primer lugar, se presentan las principales limitaciones de los corpus generales de referencia, particularmente del CORDE (RAE) y *Corpus del Español* (BYU), no solamente en cuanto al acceso de sus motores de búsqueda, sino también a la disponibilidad de los textos que los componen. En segundo lugar, se hace uso de un caso en específico, el del origen y desarrollo de las construcciones con clítico femenino, para ilustrar la propuesta. A continuación, se contrasta esta propuesta de abordaje de datos con otras investigaciones que utilizan corpus generales. Se evidencia, así, que este modo de acceso a las emisiones lingüísticas históricas favorece el estudio de procesos gramaticales novedosos de interfaz que se circunscriben a ámbitos informales, populares, orales y diatópicamente marginales.

Palabras clave: locuciones idiomáticas; clítico femenino; corpus generales; corpus históricos; español rioplatense.

Abstract: The aim of this article is to analyze the advantages of an approach to historical linguistic data based on the creation of specifically designed corpora. For this, the main limitations of general reference corpora are presented in the first place; particularly of CORDE (RAE) and *Corpus del Español* (BYU). The limitations are not only presented with regard to the access on their search engines, but also to the availability of texts that are part of said corpora. Secondly, a particular case is utilized, which shows the origin and development of the constructions with the feminine clitic, so as to exemplify the proposal. Next, the mentioned proposal on the approach of

data is contrasted to other research that use general corpora. Thus, it is demonstrated that this way of accessing historical linguistic utterances benefits the study of novel grammatical interface processes that deal exclusively with informal, popular, oral, and dialectically peripheral fields.

Keywords: idioms; feminine clitic; general corpora; historical corpora; Río de la Plata Spanish.

Recebido em 7 de setembro de 2020

Aceito em 26 de outubro de 2020

1 Introducción

Con la afluencia del giro computacional de las ciencias sociales y humanas (DE MATTEIS, 2015), el acceso y desarrollo de bases de datos en línea, corpus electrónicos generales y de referencia, así como la multiplicidad de nuevos géneros y prácticas discursivas en Internet –correos electrónicos, mensajería instantánea, entradas en redes sociales, comentarios en periódicos, intervenciones en foros, por citar algunos–, han cobrado un lugar preponderante dentro de la investigación del lenguaje (VELA DELFA; CANTAMUTTO, 2015). Este escenario no es nuevo y se enmarca al mismo tiempo en una realidad histórica: las investigaciones filológicas y de las primeras décadas del siglo XX fueron precursoras del análisis lingüístico a partir de la evidencia empírica directa (BERBER SARDINHA, 2000). A partir de la década de 1960, con la incorporación de ciertas herramientas computacionales y la progresiva pérdida de la influencia de los métodos formales en los estudios gramaticales, el trabajo con datos lingüísticos auténticos comenzó a ganar cada vez más terreno (BIBER, 1990; ROJO, 2008). Si bien actualmente existe consenso en los beneficios de acercarse a las hipótesis lingüísticas de una manera que no excluya mutuamente a las dos corrientes (FILLMORE, 1992; LÜDELING; KYTÖ, 2008), esto es que tenga presente tanto fuentes directas como armado de oraciones sin necesariamente una constatación empírica, la obligación del acceso a los datos, a los “hechos del lenguaje” (WEISSER, 2016), se hace más evidente ante algunos tipos de fenómenos gramaticales.

Los avances, si bien importantes, con relación a los grandes corpus de referencia no se limitan en los alcances de las herramientas computacionales y de la lingüística de datos reales. De manera sostenida,

a partir de la década de 1990 (ROJO, 2008), la disponibilidad de la web como corpus lingüístico potencialmente infinito y el diseño y puesta en práctica de corpus específicos de discursos especializados han hecho avanzar la disciplina y la variedad de las investigaciones de forma notable. La lingüística de corpus resulta así una metodología en cierta manera joven, que permite al lingüista desarrollar un entendimiento más acabado de cómo es y cómo funciona el lenguaje a partir de la consulta cualitativamente significativa de muestras de lengua reales, provenientes de diversas fuentes, en grandes cantidades. Los corpus, que deben ser colecciones electrónicas organizadas de texto, seguir criterios de representatividad y ser guiados por parámetros puntuales, constituyen las herramientas específicas desde las cuales se accede a la verificación de una hipótesis en particular que motiva la investigación o desde las cuales se indaga sobre una problemática de la que aún no se tienen muchos datos o información (PARODI, 2008; SILVÉRIA OLIVEIRA, 2015).

La siguiente investigación se centra en el análisis de las herramientas de la lingüística de corpus que pueden aplicarse a fenómenos gramaticales novedosos e históricos que ilustran algún tipo de liminalidad entre los niveles de la lengua, sobre todo entre el morfosintáctico y el léxico. En particular, se analizan los inconvenientes que se desprenden de la utilización de los métodos tradicionales en la investigación del desarrollo histórico de las construcciones pronominales con clítico femenino –*arreglárselas, pegarla, susanearla*– (ARELLANO, 2020; MARE; CASARES, 2017; MASULLO; BÉRTORA, 2014), las cuales cuentan con una expansión marcada en la variedad rioplatense; de ellos también se ha propuesto su presencia en las variedades ibéricas (CIFUENTES HONRUBIA, 2018). En la sección 2, en primer lugar, resumimos la historia y abordamos las características de los corpus generales de referencia. En segundo lugar, focalizamos sobre las propiedades características de los corpus históricos en general y de los dos corpus históricos del español más importantes –*Corpus del Español* y CORDE– en particular. En la sección siguiente, presentamos el fenómeno de las construcciones con clítico femenino y analizamos críticamente una propuesta de investigación a partir de corpus generales. En la sección 4, mostramos las bases de la creación de una base de datos adaptada para la investigación de este fenómeno. En la sección 5, se contrastan los resultados obtenidos a partir de la utilización de ambos tipos de corpus. En la última sección, presentamos las conclusiones de la investigación.

2 Corpus generales: características y limitaciones

En cuanto a los corpus generales de referencia, es decir, las bases de datos lingüísticas en línea constituidas por millones de palabras, estos posibilitan el acceso a muestras de lengua de una manera sencilla y sin la obligación de obtener elicitaciones o recuperar contenido desde fuentes primarias o desconocidas. Al mismo tiempo, la sistematización de las producciones de los hablantes permite que otras variables metalingüísticas, a menudo ignoradas, puedan ser tenidas en cuenta a la hora de buscar el dato lingüístico necesario para la investigación. Los grandes corpus generales disponibles en línea no constituyen solamente una colección de textos accesibles digitalmente,¹ sino que representan fundamentalmente muestras de lengua anotadas según una variedad de categorías útiles para una multiplicidad de investigaciones y enfoques lingüísticos (DAVIES, 2009). En la mayoría de ellos, puede encontrarse información relacionada con la situación de comunicación, el género discursivo, la variedad del español y el lugar y momento de emisión,² por citar algunas características. Debido a la facilidad de acceso que representan los grandes corpus, muchas investigaciones del lenguaje han relegado la inclusión del “discurso real” a la búsqueda de material lingüístico en estas bases de datos. De manera inductiva, deductiva y cuantitativa, el lingüista reflexiona e hipotetiza a través de los datos presentes en los corpus y asume la existencia, productividad y frecuencia de los fenómenos a través de los resultados obtenidos.

¹ Conviene diferenciar, así, los archivos y las bibliotecas electrónicas (los primeros, depósitos de textos sin organización; los segundos, colecciones con algún tipo de criterio, sobre todo de género textual o no estrictamente lingüístico) de los corpus o los subcorpus, que constituyen una parte específica de un archivo o una biblioteca organizadas con un diseño y unos objetivos explícitos (BERBER SARDINHA, 2000).

² Nuevos proyectos lingüísticos orientados al análisis de discurso interaccional, las humanidades digitales y la sociolingüística han comenzado a incluir otras variables en la notación de sus bases de datos, al mismo tiempo que implementan medidas y decisiones para obtener el consentimiento de sujetos y/o anonimizar la identidad de los enunciadorees. En este tipo de empresas se suele agregar la información tipológica (según el canal de comunicación, escrito u oral), metodológica (disponible o elicitado), genérica, de la situación de comunicación y de las características identitarias básicas de los interlocutores (DE MATTEIS, 2015; HUNSTON, 2008; VELA DELFA; CANTAMUTTO, 2015). Los datos que no se recuperan de corpus lingüísticos generales en la investigación fueron sistematizados siguiendo estos lineamientos lingüísticos y éticos.

Sin embargo, aunque la aparición de los grandes corpus lingüísticos representa sin dudas mejores posibilidades para el desarrollo de determinadas investigaciones, estos no dejan de traer aparejados ciertos inconvenientes y limitaciones, sobre todo en relación con la elección de los textos que los componen. Como dentro de sus objetivos normalmente se asume la compilación de la mayoría de los aspectos de una lengua o por lo menos una variedad de ella, generalmente se han señalado las incompatibilidades del estudio de la variación, debido a la neutralización que se deriva al intentar dar cuenta de muchos tipos de discursos simultáneamente (RISSANEN, 2008). Más allá de esta crítica, que se corresponde con las primeras instancias del desarrollo de la disciplina, en las que la cantidad de palabras y textos estaban sujetos a la posibilidad de alojar determinada cantidad de textos, se han presentado también otros problemas.

Por una parte, se estima que los corpus están influenciados en gran medida por contextos escritos (ROMAINE, 2008; WEISSER, 2016). Como consecuencia, los discursos orales se encuentran, con suerte, subrepresentados o directamente no son tenidos en cuenta (CLARIDGE, 2008). En el caso de aparecer, suelen estar relacionados con ámbitos cultos y formales. Muchos corpus cuentan con ejemplos de discursos públicos, disertaciones y clases orales, por citar algunos casos, que generalmente pueden ser actos orales de un discurso previamente escrito (HUNDT, 2008). Por otra parte, como consecuencia de la ausencia de textos del lenguaje hablado, también escasean muchos tipos de textos que, además de ser orales, suelen suceder en interacción y/o son de carácter popular o informal. Este tipo de discursos, al estar raramente disponibles por escrito, ser difíciles de transcribir e involucrar participantes que generalmente no son tenidos en cuenta por los investigadores, suelen ser abandonados. Esta falta afecta el trabajo en general de cualquier investigador de la lengua, pero sobre todo de un conjunto de áreas, entre ellas la gramática, la dialectología, la sociolingüística, la lexicografía y los estudios de cambio lingüístico.

En relación con esto último, efectivamente el panorama desfavorable se acrecienta cuando se lidia con corpus históricos, y no emisiones lingüísticas actuales, en los que el acceso al material de forma representativa se vuelve aún más complejo por razones evidentes. De por sí, los corpus históricos suelen ser más pequeños que los sincrónicos (CLARIDGE, 2008). Asimismo, la disponibilidad de registros orales

grabados incluso durante la primera mitad del siglo XX es realmente exigua. En caso de existir ejemplos, es posible no contar con toda la información lingüística y metalingüística necesaria para llevar a cabo algunas investigaciones, debido a la calidad del material audiovisual o a las dificultades en el rastreo de los participantes de las emisiones registradas, entre otras cuestiones. Los ejemplos anteriores al siglo XX se concentran especialmente en transcripciones de discursos orales, en los que también suelen abundar los registros más formales o los usuarios más cultos. Así, son las fuentes literarias como un conjunto las que ocupan gran parte del total de textos que pertenecen a las bases de datos históricas.

En la actualidad, se cuenta con dos grandes corpus históricos disponibles en español:³ el Corpus Diacrónico del Español (CORDE),⁴ administrado por la Real Academia Española; y el *Corpus del Español Genre/Historical*, desarrollado por Mark Davies y la Universidad de Brigham Young (BYU).⁵ En ambos, los textos analizados van desde los primeros registros escritos con los que cuenta el español (en el siglo XI) hasta mediados del siglo XX. En términos de cantidad de palabras, resultan ambos proyectos de importante capacidad, incluso con un total de *tokens* superior al *Helsinki Corpus*, una de las bases de datos más extensas para el análisis histórico del inglés (DAVIES, 2009).

Para el caso del *Corpus del Español*, hasta el año 1900, está compuesto de 1670 textos, manuscritos o facsímiles, particularmente tomados de textos literarios disponibles de la Biblioteca Virtual *Miguel*

³ Estos no constituyen los únicos dos corpus disponibles en línea, sino solo los más importantes en cuanto a cantidad de palabras, influencia y extensión en las investigaciones, debido a su gratuidad y las universidades e instituciones que los desarrollan. Para trabajar con corpus que fundamentalmente se ocupan de variedades ibéricas desde un punto de vista histórico, pero que se centren en variedades no formales, se recomienda consultar, por ejemplo, *PostScriptum* (Centro de Lingüística de la Universidad de Lisboa), que se destaca por ser un corpus histórico que incluye epístolas, cartas y géneros escritos informales. El Corpus Diacrónico y Diatópico del Español de América (CORDIAM), coordinado por la Academia Mexicana de Letras y dirigido por Concepción Company Company, es quizá la única excepción a la tendencia peninsular. Otros corpus históricos son relevados por Contreras Seitz (2009) y Enrique-Arias (2012).

⁴ *Corpus diacrónico del español* (CORDE). Recuperado de: <http://corpus.rae.es/cordenet.html>. Acceso en: 6 set. 2020.

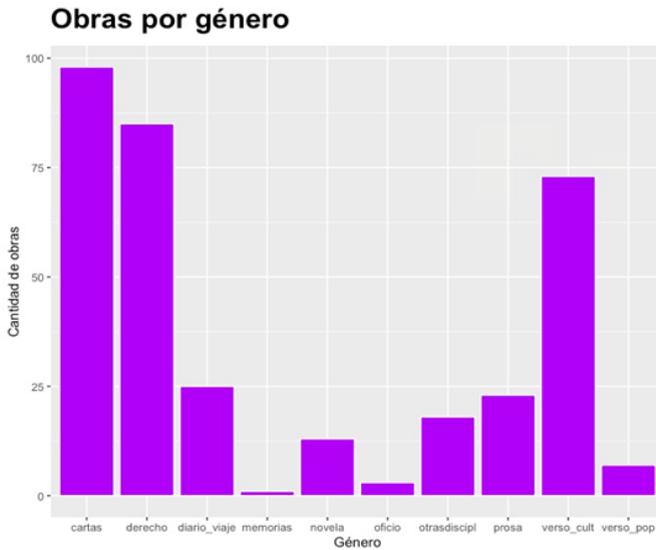
⁵ *Corpus del Español Genre/Historical*. Recuperado de: <https://www.corpusdelespanol.org/hist-gen/>. Acceso en: 6 set. 2020.

de Cervantes y otros proyectos recopilatorios literarios electrónicos. Se destaca la presencia de 392 novelas para el período que abarca el siglo XIX (DAVIES, 2002). La presencia de otros registros, ya sea otro tipo de literatura, textos periodísticos u orales, comienza a realizarse a partir del siglo XX, por lo que el equilibrio que se sostiene mantener se corresponde únicamente con las décadas finales de la franja temporal elegida, esto es los últimos años del siglo XX, sobre todo la década de 1990. No obstante, entre los puntos débiles, se encuentra la utilización de fuentes orales, las cuales se basan principalmente en el ámbito del dominio público de personalidades importantes de Iberoamérica, por lo que se destacan entrevistas a representantes de partidos políticos, intervenciones en sesiones legislativas y discursos y notas periodísticas de diarios españoles de tirada nacional.

Al mismo tiempo, no se especifican de manera extensiva las variedades del español que se toman en cuenta en los textos del siglo XIX ni en otras épocas, por lo que no puede determinarse el porcentaje de textos latinoamericanos y si estos tienen relación con el criterio poblacional que sigue adelante el equipo de la Universidad de Brigham Young en la selección de textos para sus corpus sincrónicos. En estas bases, al contrario del corpus histórico, se constata una clara mayoría de textos y cantidad total de palabras de variedades latinoamericanas, en concordancia con el menor peso demográfico que aporta la península ibérica.

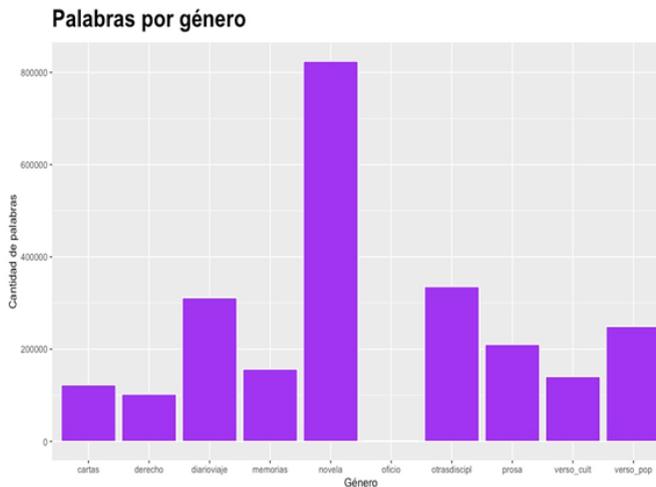
En relación con el Corpus Diacrónico del Español (CORDE), este se compone de 5500 “grandes obras” de la lengua española, con un 74% de datos provenientes de España, y un 25% correspondiente a América Latina (1% corresponde al judeoespañol), priorizando la delimitación de 6 grandes áreas diatópicas (ROJO, 2008; SÁNCHEZ SÁNCHEZ; DOMÍNGUEZ CINTAS, 2007). Cuenta con un total de 180 millones de palabras (PARODI, 2008). En principio, el peso de las obras literarias parece ser pequeño, si se tiene en cuenta que solamente el 44% de los textos se corresponde con literatura. De hecho, cuando se observa la segmentación según el tipo textual, género discursivo o área del conocimiento que compone el corpus, se observa que las cartas, los tratados de derecho, los versos cultos y el diario de viaje se corresponden con las cuatro entradas más frecuentes en el corpus (FIGURA 1). Sin embargo, cuando se analiza el peso por cantidad de palabras, la situación cambia drásticamente: el género novela aporta la mayoría de ellas (FIGURA 2).

FIGURA 1 – Gráfico de confección propia a partir del número de obras por género en Corpus Diacrónico del Español (CORDE) sección Argentina.



Fuente: Datos disponibles en: <http://corpus.rae.es/cordenet.html>

FIGURA 2 – Gráfico de confección propia a partir del número de palabras por género en Corpus Diacrónico del Español (CORDE) sección Argentina.



Fuente: Datos disponibles en: <http://corpus.rae.es/cordenet.html>

Al contrario del *Corpus del Español*, las posibilidades de búsqueda en el CORDE para investigaciones gramaticales más allá del léxico, sobre todo relacionadas con fenómenos morfosintácticos y semánticos, se ven reducidas por la ausencia de lematización y de anotación específica de clases de palabras más allá de las principales (BUENAFUENTES DE LA MATA; SÁNCHEZ LANCIS, 2012; DAVIES, 2009). A pesar de estas observaciones, aun así los corpus cuentan con más dificultades para estudios de tipo morfológico o sobre unidades perifrásticas (DAVIES, 2002). En este escenario, los estudios sobre afijos poco regulares, no monosémicos o con restricciones semánticas particulares (por ejemplo, en los que la adición de determinado afijo quede determinada al aspecto o la valencia del verbo) representan un problema adicional que la sintaxis de expresiones regulares no siempre puede saldar.

Este tipo de fenómenos no son poco frecuentes en las lenguas del mundo (BYBEE, 1985) ni en español, por lo que muchos procesos morfosintácticos, incluso de alta frecuencia, no pueden ser buscados con facilidad a través de algunos corpus. Hasta en las búsquedas más avanzadas siempre se presume la existencia de una equiparación única uno a uno entre una determinada palabra y su clase. Esta presuposición supone un problema para palabras homomórficas o con distinto criterio ortográfico, préstamos y, sobre todo, para palabras funcionales, que pueden ser categorizadas como artículos, partículas, pronombres y clíticos, todas nomenclaturas correctas dependiendo de los contextos gramaticales en los que aparecen o el sistema de notación que se siga. Esta particularidad resulta principalmente evidente para los clíticos, cuya proclisis o enclisis en combinación con determinados verboides o bien cuenta con más de una posibilidad de realización o bien no presenta restricciones gramaticales tan claras en algunos contextos.

De este modo, se evidencia la necesidad de un método de aproximación a los datos que pueda ser aprovechado desde el punto de los investigadores y la implementación de soluciones creativas más allá de los grandes corpus de referencia y de los textos electrónicos más fácilmente disponibles (BUENAFUENTES DE LA MATA; SÁNCHEZ LANCIS, 2012; ENRIQUE-ARIAS, 2012). A continuación, se presenta un fenómeno particular de la variedad rioplatense, de amplio desarrollo actual e histórico, que ilustra las problemáticas, limitaciones y características desarrolladas hasta ahora.

3 Construcciones verbales con clítico femenino: aproximación desde los corpus generales

Las locuciones verbales con clítico femenino *la* no anafórico, inherente o marginal constituyen un proceso novedoso de formación de palabras en el que el pronombre acusativo femenino *la* puede unirse a bases patrimoniales y neológicas del español. De esta manera, se pueden formar o bien construcciones más bien idiomáticas (1), en las que el significado de las bases difiere en cierta medida de la locución formada, o bien un tipo de unidad que sirve para intensificar (ALBELDA MARCO, 2004) el valor de las acciones descritas (2).

- (1) Sepan que si **la pego**, les traigo a todos conmigo al éxito (Facebook, 2015).
- (2a) ¿Hace falta **cancherearla**? (Facebook, 2019).
- (2b) Sé que el disco **la flopeó** mal, pero es un buen disco (Twitter, 2014).
- (2c) Quiero contarte que empecé a **mariekondearla** y saqué la mitad de la ropa (Twitter, 2017).

En el caso de (1), el pronombre *la* en la construcción *pegarla* ('tener éxito, generalmente con suerte') contribuye a la formación de una locución con un significado diferente al valor general asociado al lexema *pegar*. En los ejemplos de (2), el clítico aporta un matiz elativo que sirve para maximizar la forma de realización del evento descrito. Todos los casos de este grupo pueden funcionar –y, de hecho, se registra así– sin el pronombre *la*: *cancherear* ('actuar como un canchero', es decir, fanfarronear), *flopear* (del inglés *to flop* 'fracasar comercialmente') y *mariekondear* ('actuar como Marie Kondo', es decir, ordenar).

Este tipo de construcciones generalmente son proferidas por hablantes jóvenes y pueden encontrarse fácilmente en el discurso de las redes sociales o en las interacciones propias de la mensajería instantánea. Asimismo, también están presentes en discursos plenamente orales (ARELLANO, 2020).

Si bien los últimos estudios (ARELLANO, 2020; ARIAS, 2018; SILVA GARCÉS, 2017, por citar algunos) se han centrado sobre el valor sincrónico y el estado actual de las construcciones en la lengua, otras investigaciones de carácter más o menos reciente han señalado tangencialmente la pertenencia histórica del fenómeno. Para la variedad

argentina, Gobello (1991), Conde (2011, 2013) y Ghio y Albano (2013) han recuperado algunas construcciones de la primera mitad del siglo XX al establecer una relación entre este tipo de locuciones y el léxico lunfardo rioplatense. Se pueden observar también algunos ejemplos que se forman a partir de verbos patrimoniales en combinación con otros pronombres femeninos y reflexivos, que no son exclusivamente el clítico *la*.

- (3) Milonguera, bullanguera, que **la va de** alma de loca (FONT, 1927 *apud* GHIO; ALBANO, 2013, p. 111)
- (4) **enchufarla, saberla lunga** (CONDE, 2013, p. 100)
- (5) El tipo **se las trae** (GHIO; ALBANO, 2013, p. 114)
- (6) El indio no le da ni de comer a la mujer, que **se las tiene que rebuscar** sola (GOBELLO, 1991, §Rebuscar)

Pese a estas menciones, encontradas sobre todo en letras de tango y refranes, el carácter histórico del fenómeno de creación de construcciones verbales con clítico femenino no ha sido abordado desde una perspectiva filológica y diacrónica. Generalmente se asume así la presencia de este tipo de locuciones y no se toman en cuenta de manera sistemática otras descripciones de unidades y las distintas entradas en diccionarios generales que recuperan algunas unidades lexicalizadas (*DLE*, *DAMER*, entre otros).

En una primera aproximación, y como ya fue mencionado, surgen algunas complicaciones para la búsqueda y ampliación de los datos lingüísticos de este fenómeno en corpus generales. Los clíticos se homologan a los artículos definidos y los pronombres anafóricos de objeto directo. En efecto, pueden aparecer unidos ortográficamente a un verboide o separados cuando se relacionan con verbos conjugados o en construcciones con infinitivo o gerundio. Asimismo, los ejemplos se encuentran en gran medida en discursos del tipo fenoménico, en los que la incidencia del factor coloquial y oral está más presente. Los registros, géneros textuales y variedades dialectales en las que esperamos encontrar este tipo de ejemplos también escasean en representación.

De todos modos, existen investigaciones que hacen uso de los corpus lingüísticos clásicos para analizar la naturaleza de las construcciones con clítico femenino. Cifuentes Honrubia (2018) constituye uno de esos casos. En este trabajo, el investigador hace un relevamiento histórico de un considerable conjunto de construcciones que

combinan los pronombres *la(s)* y *se*, o en los que simplemente aparece alguna de las formas del clítico pronominal femenino, que abarca desde los primeros registros del español hasta la actualidad. Esto se lleva a cabo con el uso del CORDE⁶ de manera general y el uso esporádico del CREA (*Corpus de la Real Academia Española*, el corpus sincrónico de la RAE), limitándose al estudio del español europeo (CIFUENTES HONRUBIA, 2018, p. 13). Aparecen, de todas maneras, un buen número de ejemplos de variedades latinoamericanas en su argumentación, especialmente a medida que los ejemplos encontrados se acercan al presente.

Más allá de las conclusiones que establece la investigación acerca del origen y desarrollo de la construcción, metodológicamente no está libre de obstáculos. Cifuentes Honrubia (2018, p. 13) reconoce las dificultades del estudio histórico “por la propia complicación del manejo de corpora”. Así, el lingüista no limitó su trabajo al uso de los corpus mencionados, sino que, en una primera instancia, dio cuenta de una “labor de rastreo y acreditación de las distintas construcciones consideradas” por medio de la confección de una lista de construcciones que obtuvo “a través de trabajos lexicográficos y fraseológicos”. Más allá de estas líneas, no se precisan otras reflexiones o indicaciones acerca de la obtención de los ejemplos históricos y literarios que registra en su investigación.

Las características, tanto del fenómeno como del motor de búsqueda del CORDE, contribuyen a la decisión de comenzar las investigaciones a partir de otras fuentes. Enrique-Arias (2012) particularmente señala dos limitaciones que se desprenden de la utilización directa de este tipo de corpus generales. La primera es la noción de perspectiva, es decir, relacionada con la forma de acceso a los datos y es de la que da cuenta Cifuentes Honrubia. Efectivamente, para conocer las formas de antemano que el investigador intenta buscar, se necesita contar con un acceso anticipado a través de gramáticas, diccionarios, lexemarios o algún tipo de documento o lista organizados con el criterio deseado. El otro problema es el de la comparabilidad. Este se relaciona con la idea de que los cambios lingüísticos se suceden en distintas etapas: “un estadio original anterior al cambio, una fase en la que triunfa la nueva estructura

⁶ Por las similitudes en la conformación de ambos corpus, nos concentramos únicamente en el análisis derivado del Corpus Diacrónico del Español (CORDE). Al igual que lo que ocurre con este corpus, los resultados obtenidos a partir de la utilización de una base propia presentan pocas coincidencias con las unidades que se pueden obtener a través del motor de búsqueda del *Corpus del Español* (BYU).

y una etapa intermedia en la que coexisten el sistema innovador y el original” (ENRIQUE-ARIAS, 2012, p. 89).

En investigaciones que intentan dar cuenta del desarrollo de un fenómeno a lo largo de un considerable período de tiempo, resulta fundamental seguir un horizonte de equivalencia entre los distintos textos que conforman el total de la base de datos a desarrollar. De esta forma, se puede dar cuenta de una manera más acabada acerca del proceso de expansión y retracción de determinadas formas lingüísticas.

Con todo, la investigación con corpus debe cumplir en el mejor de los casos una doble misión: una primera, heurística, dedicada a conocer más acerca del fenómeno a través de las unidades formales que la representan; y una segunda, que permita ir desde la comprensión del fenómeno a la búsqueda de instancias que la verifiquen.

4 Base de datos propia: lineamientos generales

Puede pensarse –y con razón– que los corpus generales ofrecen una gran cantidad de ventajas que pueden ser difíciles de compensar. En principio, son herramientas que llevan años de desarrollo y exitosamente han contribuido a un sinnúmero de investigaciones. Además, cuentan con una selección criteriosa de géneros y contienen una gran cantidad de textos en su conformación. Sobre todo en relación con la representatividad, se cree que los corpus generales resuelven la cuestión del equilibrio de textos y formas a partir de la concentración de un gran volumen de palabras. No obstante, entendemos que no hay criterios objetivos para la determinación de la representatividad (PARODI, 2008). Es posible hablar de ella en términos relativos, alrededor de otros factores que no son la extensión. Según Berber Sardinha (2000), por ejemplo, los corpus compilados en escala pequeña por investigadores individuales terminan siendo más representativos, debido a que los corpus generales “no son necesariamente adecuados para la investigación de cualquier característica lingüística” (BERBER SARDINHA, 2000, p. 348; mi traducción).

Así, resulta claro que cualquier investigación histórica, y en especial aquellas que pretenden dar cuenta de la evolución de estructuras, construcciones y determinados comportamientos morfológicos o sintácticos, requiere de un tipo de aproximación heterogénea. Para ello, se debe priorizar la confección de bases de datos propias, a partir de distintas fuentes, incluso corpus generales de referencia, aunque estos deben ser

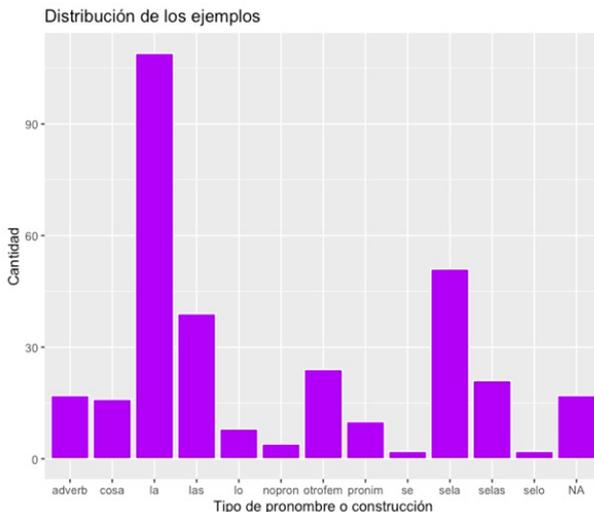
tenidos en cuenta de modo secundario. Las fuentes deben encontrarse y anotarse específicamente según la pertinencia del estudio que se quiere llevar a cabo. Si los estudios están destinados a una palabra o fenómeno en particular, es habitual hacer selecciones de texto que contengan el ítem; esto es particularmente cierto si el ítem es infrecuente (HUNSTON, 2008). Sin embargo, para la emergencia de nuevos fenómenos en el habla coloquial, las prioridades tienen que estar fijadas en la obtención de ejemplos de español hablado, informal y deben ser diatópicamente pertinentes.

Si bien es evidente la contradicción inherente que subyace a la idea de registros orales en investigaciones históricas anteriores al siglo XX, entendemos que esta empresa puede llevarse a cabo a través de registros y géneros textuales particulares que habiliten la presencia de interlocutores o personajes populares y que introduzcan lenguaje hablado (CLARIDGE, 2008). El género literario que mejor se adapta a estas exigencias es el teatro (CONDE, 2010). En general, se trata de obras de carácter realista en las que se procura reflejar la lengua coloquial (FONTANELLA DE WEINBERG, 1970). Sobre todo para la etapa que va desde finales del siglo XIX hasta mediados del siglo XX, el teatro argentino experimenta un auge que no es ajeno a la vasta producción de sainetes, el género teatral popular por excelencia. Esta expansión coincide con la época del aluvión migratorio y la consecuente modificación del entramado social y cultural, la transformación de las estructuras económicas, la urbanización acelerada y el comienzo de un proceso de industrialización en las metrópolis argentinas (BRAVO HERRERA, 2015; FLØGSTAD, 2014; FONTANELLA DE WEINBERG, 1970). En cuanto a lo lingüístico, se señala que es una etapa en la que existe “una tendencia general a una evolución de los tratamientos asimétricos hacia tratamientos simétricos” (FONTANELLA DE WEINBERG, 1970, p. 17); así, sostiene Fontanella de Weinberg, “se ha pasado a un trato igualmente recíproco pero cercano” (1970, p. 22).

Además, este período representa el momento de creación y expansión del lunfardo, argot porteño, cuyo ámbito predilecto de difusión es el lenguaje literario y periodístico y los tangos. El lunfardo no constituye una lengua en sí, sino un tipo de léxico, un sociolecto, un modo de expresión del *populus minutus* porteño (CONDE, 2013, p. 80). Representa un tipo de habla marginal, opuesta a la estandarizada, y por lo tanto con ella “se procura crear una situación comunicativa desjerarquizada” (CONDE, 2013, p. 78).

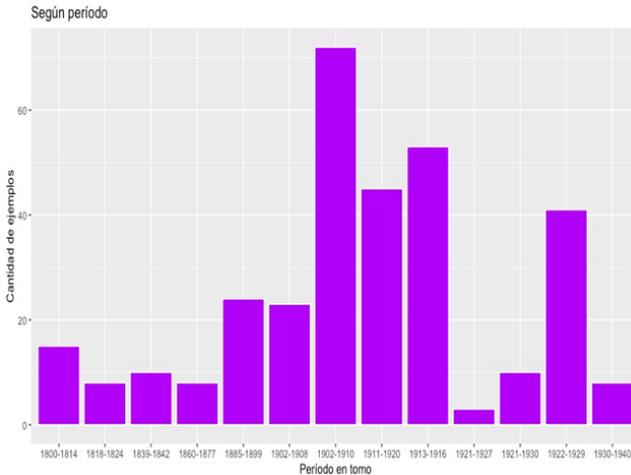
La base de datos histórica confeccionada y analizada en este estudio se compone de 303 oraciones, extraídas de los 15 primeros tomos de la *Antología de obras de teatro argentino*, compilada por la investigadora especialista en teatro Beatriz Seibel, y publicada por el Instituto de Teatro Argentino. La colección abarca un período de 140 años ininterrumpidamente, desde principios del siglo XIX hasta la década de 1940, y presenta obras pertenecientes a distintos subgéneros teatrales, retratando las tendencias estéticas de cada momento. En el 55% de las obras se encontraron construcciones pronominales. La edición se realiza manteniendo los manuscritos originales y está disponible de manera totalmente digital y gratuita. Estos últimos elementos resultan de fundamental importancia para un correcto análisis de texto a través de herramientas computacionales (CLARIDGE, 2008). Eventualmente, el análisis puede complementarse con otra porción de datos tomada de diccionarios y proyectos lexicográficos del lunfardo y letras de tango (CONDE, 2011; GOBELLO, 1991), que suman un total de otros 63 ejemplos. A continuación, se especifican los resultados por período según la especificación de cada tomo (FIGURA 3) y la distribución de los tipos de pronombres asociados a las construcciones (FIGURA 4).

FIGURA 3 – Distribución de ejemplos en base de datos histórica según período.



Fuente: Confección propia a partir de Seibel (2008).

FIGURA 4 – Distribución de ejemplos en base de datos histórica según tipo de construcción.



Fuente: Confección propia a partir de Seibel (2008)

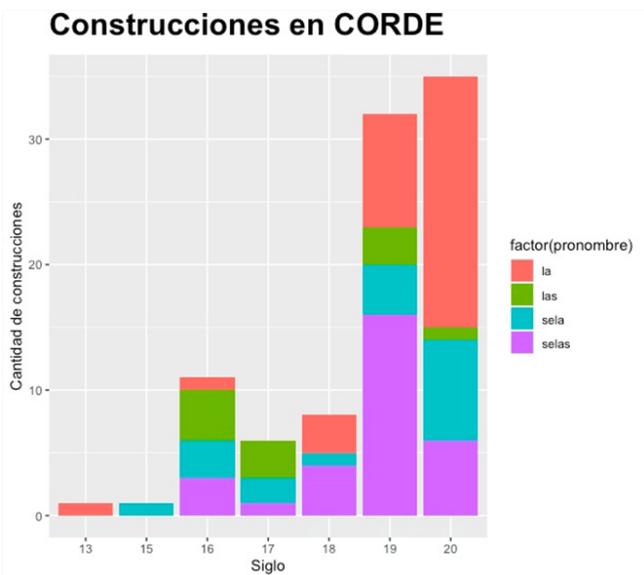
5 Contrastes entre corpus

En relación con el origen de las construcciones, Cifuentes Honrubia (2018) rastrea las primeras oportunidades en que la construcción aparece sin una clara referencia anafórica y, por lo tanto, no puede ser confundida por la combinación de un verbo y un objeto directo pronominalizado convencional. Además de los casos de *hacerla* ('faltar a lo que se debía') y *guardársela* ('aplazar la venganza o castigo de una ofensa'), cuyos primeros ejemplos se ubican en los siglos XIII y XV, a partir de siglo XVI comienzan a detectarse nuevos lexemas y formaciones: 11 para este siglo, 6 para el XVII y otros 8 para el XVIII. A partir del 1800, el número va en aumento y se detectan 32 para el siglo XIX y 35 para el siglo XX (13 de ellos antes de 1950). Solamente 11 de las construcciones relevadas en todo el período analizado pierden frecuencia y no se localizan en textos recientes (FIGURA 6).

En relación con los tipos y la cantidad de pronombres utilizados, si bien el aumento del número de construcciones puede estar relacionado con la disponibilidad propia de las fuentes, y no específicamente con un aumento de la cantidad de construcciones habilitadas en la lengua en el

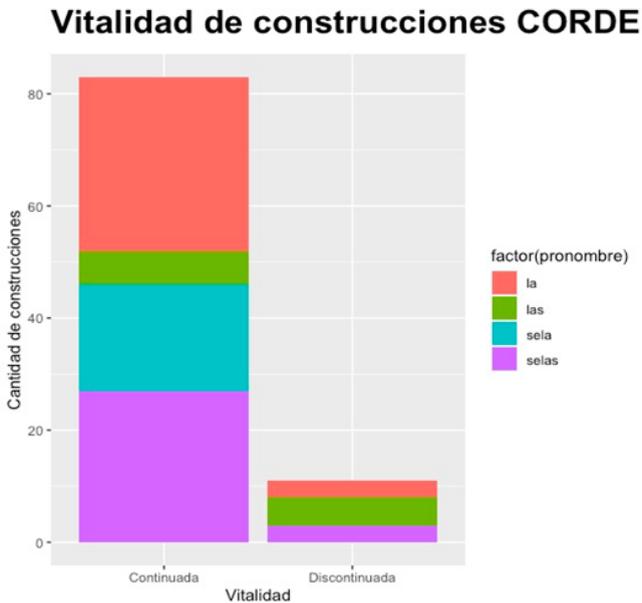
momento analizado, se evidencia una tendencia en el mantenimiento de la disponibilidad de los cuatro tipos de formantes (*se+la*, *se+las*, *las* o *la*), aunque con algunos cambios. En particular, se observa una leve propensión hacia una pérdida del clítico acusativo femenino plural *las* como formante, un aumento en proporción de la cantidad de lexemas construidos con el clítico femenino *la* y la consolidación de la unión de *se+la* como formante más frecuente (FIGURA 5), sobre todo a partir del siglo XIX (los datos del siglo XX corresponden hasta 1940, para poder comparar con la muestra del español argentino en un mismo período).

FIGURA 5 – Distribución de ejemplos según cantidad de construcciones novedosas por siglo y los pronombres que se utilizan como formantes.



Fuente: Confección propia a partir de la base de datos de Cifuentes Honrubia (2018).

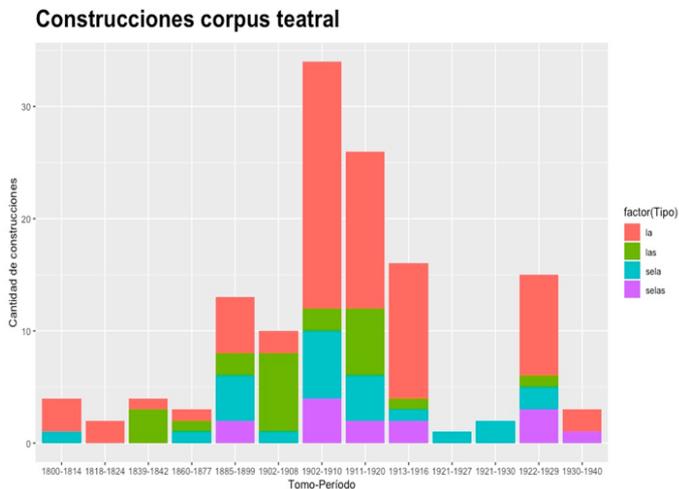
FIGURA 6 – Pérdida de vitalidad de las construcciones según los pronombres que se utilizan como formantes.



Fuente: Confección propia a partir de la base de datos de Cifuentes Honrubia (2018)

La situación difiere cuando se analiza la variedad argentina, a partir de las obras de teatro editadas por la antología de Beatriz Seibel. Mientras que para el siglo XIX se observan pocos ejemplos, de igual manera puede observarse la disponibilidad de múltiples formantes. La prevalencia de las construcciones con clítico pronominal *la*, y en menor medida con *se+la*, comienza a ser evidente desde principios de siglo XX y se extiende, aunque con una menor cantidad de ejemplos, hacia mediados del mismo siglo (FIGURA 7).

FIGURA 7 – Distribución de ejemplos para la base de datos de confección propia según cantidad de construcciones novedosas por período de tomo y los pronombres que se utilizan como formantes.



Fuente: Confección propia a partir de Seibel (2008).

La comparación lexemática también arroja diferencias interesantes. Solo algunas pocas construcciones aparecen en ambas bases y a menudo lo hacen con cambios, particularmente en la elección de pronombres o la necesidad de un predicativo secundario. Para el caso argentino, *haberlas* (‘disputar’), *tirlarla* y *echarla(s)* (ambos ‘tener ínfulas de’) aparecen sin el pronombre *se*, al contrario de lo que sucede en el CORDE. En el resto de los casos, como en *dársela* o *dárselas* (‘tener ínfulas de’), *armarla* (‘generar un revuelo’) y *cantarlas* (‘decir lo que se piensa’) se observan exactamente las mismas construcciones, aunque en la base de datos teatral también aparecen algunos predicativos, como *dársela seca* y *dárselas todas* (‘pegar violentamente’) y *cantarlas claro*. Como consecuencia, existe una cantidad relativamente alta de construcciones que no se encuentran en la base opuesta. Si bien en el caso de la base de datos del español europeo no tenemos un acceso directo al total de ejemplos por cada construcción, todas las construcciones relevadas cuentan con una multiplicidad de instancias, que son mostradas por Cifuentes Honrubia (2018) en su argumentación. En este sentido, podemos asumir por lo menos una relativa frecuencia asociada a cada una de estas construcciones. De la misma manera, si nos atenemos a los

casos de la base de datos construida a partir de los ejemplos de obras de teatro argentinas, algunas construcciones con alta frecuencia no aparecen en pocas o ninguna oportunidad en la base de datos ibérica. En particular, se destacan *tenerla con* ('guardar rencor') y las diferentes variantes de *irla de —irlas de/con—* ('creerse algo que no se es'); todos estos lexemas han aparecido por lo menos 5 veces en el conjunto de obras argentinas.

Particularmente puede observarse una importante vitalidad del fenómeno si se tiene en cuenta el análisis de la base de datos de las obras de teatro argentinas en dos aspectos fundamentales. El primero se corresponde con el aumento de las construcciones relevadas en general y en especial con el aumento en proporción de las variantes con exclusivamente pronombre *la* en comparación con los ejemplos del mismo período y con años anteriores. En una segunda instancia, se destaca el tipo de bases verbales con las que el pronombre se combina. Así, se encuentran combinaciones con verbos intransitivos, incluso inacusativos o con un régimen pronominal distinto (*irla, palpitarla, protestarla, equivocarla, trabajarla*), occasionalismos restringidos a la situación comunicativa en especial retratada en la obra o relacionados con el personaje en particular que lo enuncia (*epilogarla, chapariarla*), verbos con *-ear* (*gorjearla*) y otras bases verbales novedosas, formadas a partir de sustantivos (*balconearla*) o préstamos léxicos de otras lenguas o relacionadas al ámbito del lunfardo (*escabiarla, morfarla*). En todos los casos, la adición del pronombre *la* no contribuye idiomáticamente a la construcción, por lo que los nuevos significados se pueden derivar de los lexemas originales.

- (7) Yo nació para **protestarla**. Lo tengo adentro, en la entraña. (*Los disfrazados, apud SEIBEL, 2008, Tomo 8, p. 232*)
- (8) Pero conmigo... ¡la chapariolan! (*El conventillo de la paloma, apud SEIBEL, 2008, Tomo 12, p. 185*)
- (9) ¿Y cómo **la gorjeás** cuando estás en la batea? ¿O cuando querés que cante el pajarito? (*El debut de la piba, apud SEIBEL, 2008, Tomo 9, p. 140*)
- (10) Eso es pa otra gente, que toca el piano y se levanta a las once sin gana 'e **morfarla**... (*Los disfrazados, apud SEIBEL, 2008, Tomo 8, p. 233*)
- (11) Vamos a **balconearla**. (*El rey del cabaret, apud SEIBEL, 2008, Tomo 12, p. 107*)

Gran parte de estos ejemplos se ven asimismo confirmados por otras bases de datos secundarias de la misma época, que contienen datos provenientes de letras de tango, lexemarios dedicados al lunfardo y otras obras literarias populares no teatrales. Así, los fenómenos en cada lado del Atlántico tienen desarrollos distintos, tanto en sus particularidades históricas como actuales. La explosión de principios del siglo XX en Argentina llega, aunque con intermitencias, hasta principios del siglo XXI con una nueva expansión del fenómeno hacia bases nominales más complejas y préstamos léxicos, como en los ejemplos de (2), que reponemos aquí con nuevas variantes.

- (12) Si tengo pago, no **la fantasmeo**. (*Gas Montana Remix*, 2018)
- (13) Pienso **buquearla** un rato con el día de la mujer, ahí voy. (Twitter, 2014)
- (14) Te guardaste un añoito, te teñiste y saliste a **susanearla**. (Twitter, 2014)

Con todo, en el caso de la variedad argentina, y a partir del nuevo corpus, se puede dar cuenta de una serie de características que están ausentes de otros análisis. En primer lugar, se detecta una mayor frecuencia y productividad del fenómeno, especialmente a partir de las primeras décadas del siglo XX. En segundo lugar, este aumento se ve acompañado por una tendencia a la regularización y morfologización del pronombre *la*, al mismo tiempo que se evidencia una pérdida progresiva de la carga idiomática que aporta el clítico. Asimismo, puede postularse una línea temporal de desarrollo del fenómeno, que brinda una posible explicación a la relación del proceso actual con uno histórico. Los ejemplos registrados en Argentina en diccionarios, obras del lunfardo y algunas investigaciones lingüísticas particulares no constituyen casos esporádicos, sino más bien instancias de realización de un fenómeno de amplia difusión en la lengua.

6 Conclusiones

En la introducción identificamos la importancia de las herramientas de la lingüística de corpus en las investigaciones actuales y la relación de esta metodología con otras subdisciplinas dentro de los estudios del lenguaje.

A partir de la indagación sobre la posibilidad de los corpus de referencia, en la sección 2 establecimos sus características más importantes. Nos concentramos en las limitaciones que han sido señaladas por la bibliografía y otras analizadas aquí. Particularmente, recuperamos las observaciones sobre las restricciones para el estudio de la dialectología, la sociolingüística, los estudios gramaticales y la historia de la lengua. En relación con esta última cuestión, se describieron las críticas que se realizaron específicamente sobre los tipos de datos lingüísticos que se recogen en los corpus históricos: falta de registros orales, informales e interactivos y alta dependencia de discursos cultos y literarios. Estas características se encuentran presentes en los dos corpus de mayor alcance del español: CORDE y *Corpus del Español*.

En la sección 3, presentamos las propiedades y antecedentes de estudio de las construcciones con clítico femenino, un proceso de interfaz entre morfología, sintaxis y léxico de amplia expansión en la actualidad en discursos orales e informales. En particular, se indicaron las pocas referencias de estudio histórico que presenta el fenómeno, pese a las referencias hechas a la presencia de estas locuciones en géneros textuales de carácter más popular.

A partir de estos datos y las limitaciones de los corpus generales introducidas en la sección 2, se presentan los lineamientos a favor de la confección de corpus de discursos específicos en la sección 4. Específicamente para el estudio de las construcciones con clítico femenino, se describe el armado de un corpus basado en obras teatrales populares rioplatenses a través de una antología de obras que ocupa la totalidad del siglo XIX y la primera mitad del siglo XX.

Finalmente, en la sección 5, se contrastan los resultados obtenidos del estudio histórico de las construcciones con clítico femenino a partir de la utilización del CORDE para el caso del español ibérico con el análisis de los datos del español de Argentina a partir del corpus histórico teatral. Si bien ambos análisis muestran resultados satisfactorios en tanto encuentran instancias en sus respectivas bases de datos del fenómeno analizado, se evidencian las ventajas de una investigación histórica específica de la variedad rioplatense. El proceso de morfologización, regularización y productividad del clítico femenino *la* en esta zona toma características particulares que no se muestran en el análisis del fenómeno europeo. En efecto, una gran porción de los lexemas, incluso de alta frecuencia, encontrados en el corpus teatral no cuentan con un correlato en CORDE

incluyendo o no las variedades americanas. También el origen y la escisión en el desarrollo del fenómeno en Argentina y en España pueden identificarse con cierta certeza a partir del corpus específico presentado.

Asimismo, una de las mayores ventajas de este tipo de aproximación más heterogénea es la de poder apreciar procesos de surgimiento, afianzamiento, expansión y retracción de determinado fenómeno lingüístico a través de la delimitación y diseño de un corpus específico controlado. De esta manera, la confección de determinadas bases de datos anotadas puede mejorar los estudios de cambio histórico y describir las características propias de fenómenos de gramaticalización y lexicalización del español. Este tipo de registros informales, además, pueden resultar útiles para la investigación de otro tipo de procesos gramaticales, que pueden estar circunscritos solamente al habla coloquial o tener poca incidencia en ámbitos de prensa y literatura. Por último, la confección de corpus específicos puede contribuir a la aportación de otras perspectivas a fenómenos ya investigados a partir de corpus generales, no solo a estudios históricos, sino también sincrónicos.

Referencias

ALBELDA MARCO, M. *La intensificación en el español actual*. 2004. 444f. Tesis (Doctorado en Filología) – Departamento de Filología Española, Facultat de Filologia, Universitat de Valencia, 2004.

ARELLANO, N. Entre la morfología y la sintaxis: una aproximación a la creación de verbos con pronombre acusativo «la». *Forma y Función*, Bogotá, v. 33, n. 2, p. 81-108, 2020. DOI: <https://doi.org/10.15446/fyf.v33n2.80194>

ARIAS, J. J. Clítico inherente/marginal *la* en el español rioplatense: ¿de qué la va esta construcción? *Quintú Quimün*, Rio Negro, Argentina, n. 2, p. 74-103, 2018.

BERBER SARDINHA, T. Lingüística de corpus: histórico e problemática. *Delta: Documentação de Estudos em Lingüística Teórica e Aplicada*, São Paulo, v. 16, n. 2, p. 323-367, 2000. DOI: <https://doi.org/10.1590/S0102-44502000000200005>

BIBER, D. Methodological Issues Regarding Corpus-Bases Analyses of Linguistic Variation. *Literary and Linguistic Computing*, Oxford, v. 5, n. 4, p. 257-269, 1990. DOI: <https://doi.org/10.1093/llc/5.4.257>

BRAVO HERRERA, F. E. *Huellas y recorridos de una utopía: la emigración italiana en la Argentina*. 1. ed. Buenos Aires: Teseo, 2015.

BUENAFUENTES DE LA MATA, C.; SÁNCHEZ LANCIS, C. E. Procesos de gramaticalización y lexicalización a la luz de los corpus académicos. In: JIMÉNEZ JULIÁ, T. E.; LÓPEZ MEIRAMA, B.; VÁZQUEZ ROZAS, V.; VEIGA RODRÍGUEZ, A. (coord.). *Cum corde et in nova grammatica: estudios ofrecidos a Guillermo Rojo*. Santiago de Compostela: Servicio de Publicaciones e Intercambio Científico/ Universidad de Santiago de Compostela, 2012. p. 153-165.

BYBEE, J. *Morphology. A Study of the Relation between Meaning and Form*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1985. DOI: <https://doi.org/10.1075/tsl.9>

CIFUENTES-HONRUBIA, J. L. *Construcciones con clítico femenino lexicalizado*. Madrid: Verbum, 2018.

CLARIDGE, C. Historical corpora. In: LÜDELING, A.; KYTÖ, M. (ed.). *Corpus Linguistics*. Berlin; New York: Walter de Gruyter, 2008. v. 1, p. 242-259.

CONDE, O. El lunfardo en la literatura argentina. *Gramma*, Buenos Aires, v. 21, n. 47, p. 224-246, 2010.

CONDE, O. *Diccionario etimológico del lunfardo*. Buenos Aires: Taurus, 2011.

CONDE, O. Lunfardo rioplatense: delimitación, descripción y evolución. In: VILA RUBIO, N. (ed.). *De parces y troncos. Nuevos enfoques sobre los argots hispánicos*. Lleida: Edicions de la Universitat de Lleida, 2013. p. 77-105.

CONTRERAS SEITZ, M. Hacia la constitución de un corpus diacrónico del español de Chile. *RLA. Revista de Lingüística Teórica y Aplicada*, Concepción, Chile, v. 47, n. 2, p. 11-134, 2009. DOI: <https://doi.org/10.4067/S0718-48832009000200007>

DAVIES, M. Un corpus anotado de 100.000.000 palabras del español histórico y moderno. *Procesamiento del Lenguaje Natural*, Jaén, Espanha, n. 29, p. 21-27, 2002.

DAVIES, M. Creating Useful Historical Corpora. A Comparison of CORDE, the Corpus del Español, and the Corpus do Português. In: ENRIQUE-ARIAS, A. (coord.). *Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde la lingüística de corpus*. Madrid: Iberoamericana/Vervuert, 2009. p. 137-166. DOI: <https://doi.org/10.31819/9783865278685-009>

DE MATTEIS, L. Ejes para un debate sobre el uso ético de datos interaccionales escritos y obtenidos en línea. In: CANTAMUTTO, L. (ed.). *Actas de las I Jornadas de Humanidades Digitales*. Buenos Aires: Facultad de Filosofía y Letras/Universidad de Buenos Aires, 2015. p. 235-247.

ENRIQUE-ARIAS, A. Dos problemas en el uso de corpus diacrónicos del español: perspectiva y comparabilidad. *Scriptum Digital*, Barcelona, n. 1, p. 85-106, 2012.

FILLMORE, C. J. “Corpus Linguistics” or “Computer-Aided Armchair Linguistics”. In: DIRECTIONS IN CORPUS LINGUISTICS. NOBEL SYMPOSIUM, 1991, Stockholm. *Proceedings* [...]. Berlin; New York: Mouton de Gruyter, 1992. p. 35-60.

FLØGSTAD, G. *Forking Paths: Category Change, Subfunction Variation, and Preterits in Porteño Spanish and beyond*. 2014. 247f. Tesis (Doctorado en Filosofía) – University of Oslo, Oslo, 2014.

FONT, J. *Alma de loca* (letra de tango). Buenos Aires, 1927.

FONTANELLA DE WEINBERG, M. B. La evolución de los pronombres de tratamiento en el español bonaerense. *Thesaurus: Boletín del Instituto Caro y Cuervo*, Bogotá, v. 25, n. 1, p. 12-23, 1970.

GHIO, A.; ALBANO, H. ‘Locuciones verbales’ con pronombre personal átono ‘la’/‘las’ en el español coloquial de Buenos Aires. *Gramma*, Buenos Aires, v. 24, n. 51, p. 102-116, 2013.

GOBELLO, J. *Nuevo diccionario de lunfardo*. Buenos Aires: Corregidor, 1991.

HUNDT, M. Text corpora. In: LÜDELING, A.; KYTÖ, M. (ed.). *Corpus Linguistics*. Berlin; New York: Walter de Gruyter, 2008. v. 1, p. 168-187.

HUNSTON, S. Collection Strategies and Design Decisions. In: LÜDELING, A.; KYTÖ, M. (ed.). *Corpus Linguistics*. Berlin, New York: Walter de Gruyter, 2008. v. 1, p. 154-168.

LÜDELING, A.; KYTÖ, M. Introduction. In: LÜDELING, A.; KYTÖ, M. (ed.). *Corpus Linguistics*. Berlin, New York: Walter de Gruyter, 2008. v. 1, p. iv-xi. DOI: <https://doi.org/10.1515/9783110211429>

MARE, M.; CASARES, M. F. *¡A lingüístiquearla!* Neuquén: Educo, 2017.

MASULLO, P.; BÉRTORA, H. Objetos acusativos expletivos en el español rioplatense. In: CONGRESO INTERNACIONAL DE LETRAS, VI., 2014, Buenos Aires. *Anales [...]*. Buenos Aires: Universidad de Buenos Aires, 2014. p. 195-205.

PARODI, G. Lingüística de corpus: una introducción al ámbito. *RLA. Revista de Lingüística Teórica y Aplicada*, Concepción, Chile, v. 46, n. 1, p. 93-199, 2008. DOI: <https://doi.org/10.4067/S0718-48832008000100006>

RISSANEN, M. Corpus linguistics and historical linguistics. In: LÜDELING, A.; KYTÖ, M. (Ed.). *Corpus Linguistics*. Berlin, New York: Walter de Gruyter, 2008. v. 1, p. 53-68.

ROJO, G. Lingüística de corpus y lingüística del español. In: CONGRESO DE LA ALFAL, XV., 2008, Montevideo. *Actas [...]*. Montevideo: ALFAL, 2008. Ponencia plenaria. p. 1-31.

ROMAINE, S. Corpus Linguistics and Sociolinguistics. In: LÜDELING, A.; KYTÖ, M. (ed.). *Corpus Linguistics*. Berlin, New York: Walter de Gruyter, 2008. v. 1, p. 96-112.

SÁNCHEZ SÁNCHEZ, M. S.; DOMÍNGUEZ CINTAS, C. El banco de datos de la RAE: CREA y CORDE. *Per Abbat: Boletín Filológico de Actualización Académica y Didáctica*, La Rioja, Espanha, n. 2, p. 137-148, 2007.

SEIBEL, B. (ed.). *Antología de obras de teatro argentino*. Desde sus orígenes a la actualidad. Buenos Aires: Instituto Nacional del Teatro, 2008. Tomos 1-15.

SILVA-GARCÉS, J. Clíticos marginales en verbos denominales en -ear. *Quintú Quimün*, Rio Negro, Argentina, n. 1, p. 34-60, 2017.

SILVÉRIA OLIVEIRA, F. Comparação linguística e perfilação gramatical sistêmica em um corpus combinado. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 23, n. 3, p. 727-768, 2015. DOI: <http://dx.doi.org/10.17851/2237-2083.23.3.727-768>

VELA DELFA, C.; CANTAMUTTO, L. Problemas de recogida y fijación de muestras del discurso digital. *Chimera: Romance Corpora and Linguistic Studies*, Madrid, v. 2, p. 131-155, 2015.

WEISSER, M. *Practical Corpus Linguistics: An Introduction to Corpus-Based Language Analysis*. 1. ed. Oxford: John Wiley & Sons, 2016. DOI: <https://doi.org/10.1002/9781119180180>