

Lenguaje inclusivo en usuarios de Twitter en Argentina: un estudio de corpus

Gender-inclusive language in Twitter users from Argentina: a Corpus study

Santiago Kalinowski¹

AAL-DILyF²

Resumen

El llamado lenguaje inclusivo es objeto de fuertes especulaciones desde que adquirió visibilidad pública en Argentina a principios del año 2018, pero poco se conocía sobre datos básicos como las fluctuaciones en el uso de sus distintas variantes a través del tiempo y las regiones, la dimensión del fenómeno en relación a las emisiones totales, o sus porcentajes de usuarios activos. En este artículo, se presentan y se analizan los primeros resultados de un trabajo de corpus, tomado de usuarios de Twitter en Argentina, en el que se relevó este tema durante la última década en todo el país.

Palabras clave: lenguaje inclusivo; lingüística de corpus.

Abstract

The so-called inclusive language has been the object of strong speculation since it gained public visibility in Argentina in early 2018, but little was known about basic data such as the fluctuations in the use of its different variants over time and regions, the dimension of the phenomenon in relation to total emissions, or its percentages of active users. In this article, the first results of a corpus work, taken from Twitter users in Argentina, are presented and analyzed, in which this topic was surveyed during the last decade throughout the country.

Key words: inclusive language, corpus linguistics.

¹ Licenciado y profesor en Letras, Facultad de Filosofía y Letras, Universidad de Buenos Aires. Magíster en Lexicografía Hispánica, Escuela de Lexicografía Hispánica, Real Academia Española. Doctor en estudios hispánicos, Universidad de Western Ontario, Canadá. Director del Departamento de Investigaciones Lingüísticas y Filológicas de la Academia Argentina de Letras. Sus áreas de trabajo son la lexicografía, la lingüística de corpus y la divulgación lingüística.

² El autor no expresa la postura institucional de la Academia Argentina de Letras.

La intervención sobre la morfología de género que recibe el nombre de “lenguaje inclusivo” es objeto de diversas disputas: acerca de su legitimidad, su eficacia, su naturaleza, su extensión, su localización, sus posibilidades de persistir en el tiempo, etc. Las valoraciones sobre el fenómeno, mucho más que en el examen de sus rasgos constitutivos o lo que puede verificarse que sucede en el uso, suelen ordenarse en torno de antagonismos en el campo sociopolítico. En ese contexto, los saberes lingüísticos que se ponen en juego siguen, frecuentemente, una lógica prescriptiva que persigue el fin, o bien de apalancar su uso, difundirlo o ampliar su base de usuarios; o bien de obstaculizarlo, censurarlo o arrinconarlo en sitios cada vez más reducidos del espacio público.

El trabajo de relevamiento del fenómeno en usuarios³ de Twitter en Argentina, tesis de grado a cargo de Matías Guerson codirigida por el Dr. Agustín Gravano y por mí, en el marco del Departamento de Computación de la UBA y del Laboratorio de Inteligencia Artificial Aplicada, aspira a poner a disposición de especialistas y del público en general una fuente de datos fehacientes de uso del fenómeno en la última década en el país, con un abordaje centrado en la descripción.

Breve descripción del corpus

El corpus contiene datos desde marzo de 2007 a febrero de 2020 y dispone de 128.649.761 tuits generados por 56.308 usuarios de las 23 provincias, para un total de

³ Es una premisa del presente artículo que el masculino no marcado, o masculino genérico, como también se lo conoce, no implica la expresión de una opción ideológica sexista en los hablantes. Al asignarle ese contenido a algo que usan de modo intuitivo más de quinientos millones de hablantes, no solo no se describe adecuadamente el objeto sino que se promueve una estigmatización cuyo decurso probable es el surgimiento de actitudes de discriminación lingüística. El masculino no marcado, más que en sí mismo un problema, es el eco gramatical de un problema. Como tal, en muchas situaciones enunciativas, constituye una buena práctica evitarlo por medio de los diferentes recursos disponibles para ello, pero ese gesto no debe inducir a extender un juicio negativo general sobre él en tanto rasgo gramatical.

1.152.639.682 de palabras y 3.328.981 de palabras diferentes.⁴ En promedio, se eligieron al azar unos 2450 usuarios por provincia, cada uno de los cuales produjo unos 2285 tuits. Eso arroja un promedio de 5,6 millones de tuits y alrededor de 50 millones de palabras por provincia.

Armado

El corpus es sometido a una serie de operaciones para poder ser usado como fuente de datos lingüísticos. La separación del texto en un conjunto de *tokens* a través de la supresión de hipervínculos, emojis, menciones, *hashtags*, etc., se denomina **tokenización**. La **normalización** consiste en convertir todo a minúscula, remover las tildes, regularizar las palabras con letras repetidas, etc. Una vez obtenidos los *tokens*, a través de una operación de **filtrado**, se determina cuáles de ellos serán considerados palabras.

A partir de un diccionario de partida que permite elaborar un repertorio de las palabras que pueden ser intervenidas, se procedió a convertirlas a sus diferentes variantes. Por ejemplo, las formas intervenidas contempladas para la palabra *amigo* son:

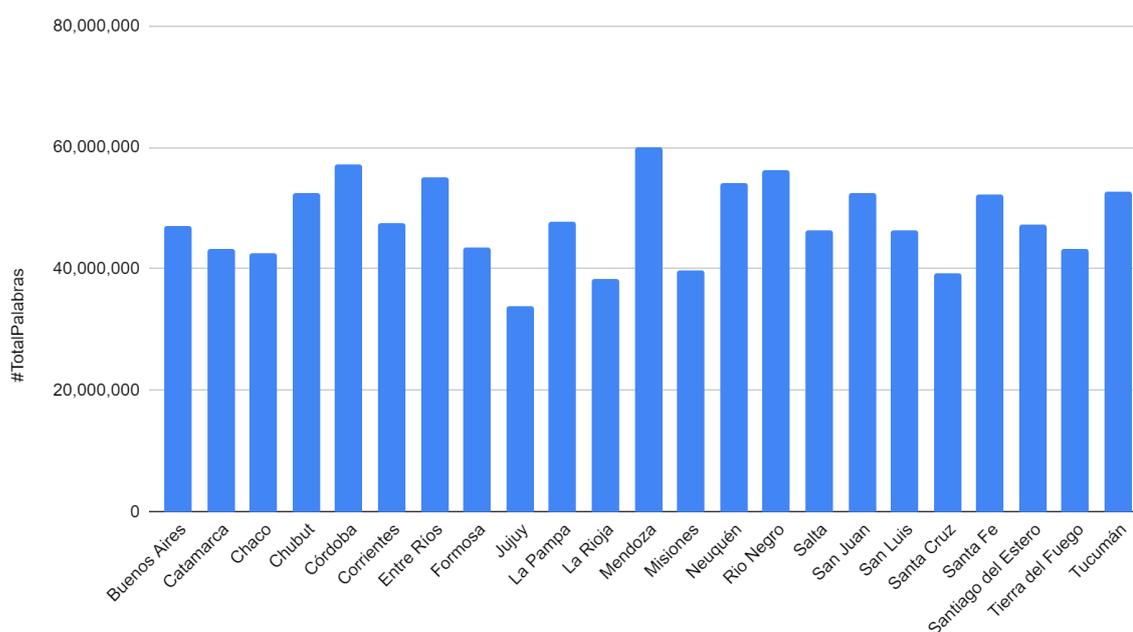
- Variante con desdoblamiento: *amigos y amigas, amigas y amigos, amigo o amiga, amiga o amigo, amigo/a(s), amiga/o(s)*
- Variante con arroba: *amig@(s)*
- Variante con equis: *amigx(s)*
- Variante con e: *amigue(s)*

⁴ Este número tan alto de palabras diferentes se debe a que se trata de un corpus con alto nivel de ruido por la gran cantidad de variantes que, por diferentes motivos, introducen los usuarios a nivel de la codificación ortográfica.

Por último, simplemente se buscaron palabras intervenidas de esas maneras específicas entre los tuits del corpus para determinar frecuencias de uso por región, por año, y establecer un repertorio.

Siempre que se reúnen grandes cantidades de datos, existen diferencias entre las variables que son inevitables. Por ejemplo, los números absolutos de cantidad de palabras entre provincias muestran algunas diferencias que oscilan entre un mínimo de 33,786,434 de palabras para la provincia de Jujuy y un máximo de 60,046,714 de palabras para la provincia de Mendoza.

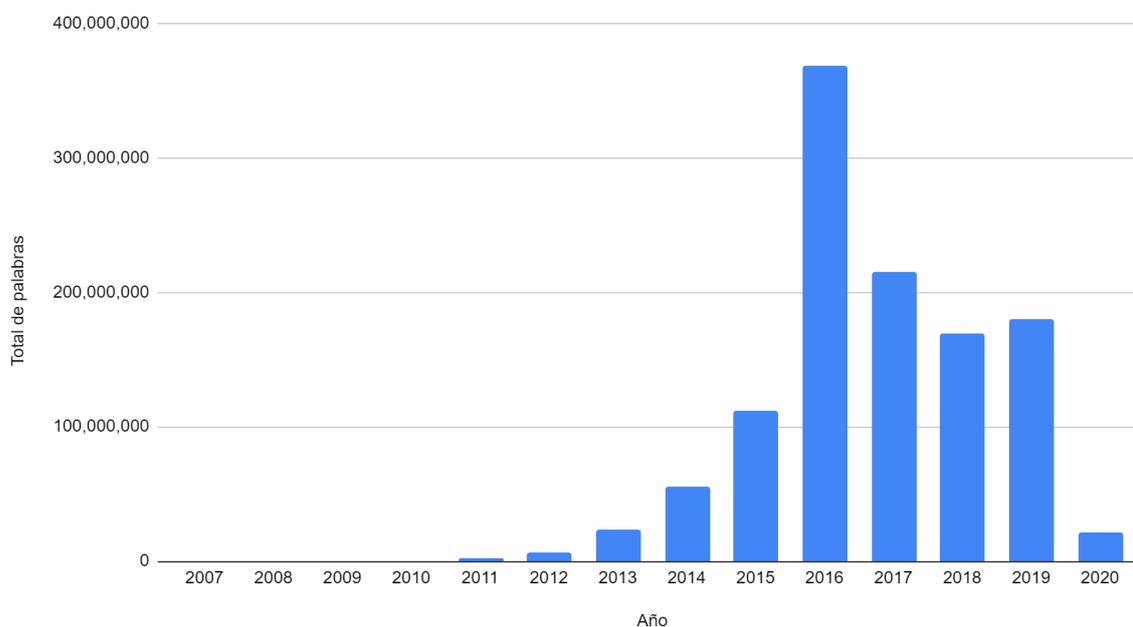
#TotalPalabras vs Provincia



Con un mínimo de 2.843, en 2007, y un máximo de 368.297.095, en 2016, mucho más abruptas son las diferencias entre años. Esto se debe a razones metodológicas que tienen que ver con la permanencia de los datos en el tiempo y con cambios en las políticas de acceso de parte de la plataforma. El primer año para el cual el corpus cuenta con datos de todas las provincias es 2009, aunque, con algo menos de cincuenta mil, se

trata de un número de palabras por debajo del necesario para derivar conclusiones válidas.

#Total de palabras vs Año



El primer año con una cantidad aceptable de palabras para todas las provincias es 2010 con 423.609. A partir del año siguiente, los datos no solo son suficientes sino que sobreabundan generosamente de acuerdo con cualquier estándar en lingüística de corpus.

La disparidad entre las distintas variables es una realidad inherente a la recolección de datos lingüísticos y obliga a trabajar siempre con medidas normalizadas, que pueden ser porcentajes, para ciertas cosas como cantidad de hablantes que reúnen cierta condición en relación al total o frecuencias normalizadas, la medida básica de la lingüística de corpus, que consiste en contar la cantidad de casos de un determinado uso por cada millón de palabras del corpus.

Por ejemplo, si se busca la palabra *pantalón* en el *Corpus del español del siglo XXI* (CORPES XXI) de la Real Academia Española, el primer resultado que se obtiene es que el país donde más se usa la palabra es España, con 4.362 casos, en segundo lugar México, con 1.799, tercero Argentina, con 1.190, cuarto Colombia, con 1.112 y así sucesivamente. Sin embargo, al observar la cantidad de veces que aparece la palabra por cada millón de palabras, este orden cambia. En primer lugar, aparece Guatemala con 71,22, segundo Perú, con 63,17, tercero Guinea Ecuatorial, con 60,14, cuarto Puerto Rico, con 55,56 y así sucesivamente. España más que duplica los casos del segundo país de la lista cuando se consideran los números absolutos, pero retrocede hasta el puesto 13, con 42,29, cuando se consideran las frecuencias normalizadas. Haciendo una división simple, la palabra aparece una vez cada 14.401 palabras en Guatemala, mientras que, en España, lo hace cada 23.646 palabras. Cuando se pasa a medidas normalizadas, emerge una imagen completamente diferente y, por lo tanto, las conclusiones que permiten caracterizar el uso de la palabra cambian de manera drástica.

La razón detrás de este hecho es muy simple. Dentro de CORPES XXI, al mes de octubre de 2020, momento en que se hizo esta consulta, España cuenta con 118.448.834 de palabras mientras que Guatemala, con 5.154.412. Es decir, el corpus tiene 2,53 palabras por habitante para España,⁵ mientras que para Guatemala ese número es de 0,28, lo cual significa que hay nueve veces más datos para los hablantes españoles que para los guatemaltecos. Con esos niveles de disparidad entre países, la única medida comparable es la frecuencia normalizada ya que comparar valores absolutos no es más que un modo de reformular los sesgos que tiene el corpus, debidos a distintos factores, entre los que con toda probabilidad debe contarse la decisión explícita de privilegiar el país de donde el corpus es nativo y con cuyos fondos se financia.

⁵ Según datos del Fondo de Población de las Naciones Unidas: www.unfpa.org/es

Desde el punto de vista metodológico, eso significa que esa frecuencia normalizada de 42,29 en España está basada en más casos que la de 71,22 de Guatemala, por lo que es de mejor calidad. No obstante, eso no debe inducir a desestimar una frecuencia que surge de un subcorpus nacional de más de cinco millones de palabras. Algo similar pasa en nuestro caso con el año 2010 o 2011 en relación a los números que tiene 2016 o 2017: de ningún modo el hecho de que las medidas normalizadas surjan de subcorpus de menor tamaño debe inducir a desestimarlas, especialmente en el caso del año 2011, que cuenta ya con 2.136.182 de palabras. Esto permite concluir que la composición de los datos del corpus autoriza a considerar que la muestra recogida es representativa del español de la Argentina usado en Twitter durante la última década.

Otro de los puntos desde los cuales puede resultar ilustrativa la comparación con CORPES XXI tiene que ver con el tipo de lengua que cada corpus captura. CORPES XXI tiende a reunir materiales que fueron publicados. La categoría “Mensaje en redes sociales”, por ejemplo, reúne 27.139 palabras, o el 0,0077 % del total. El medio “oral” es algo más abundante con 4.543.110, o el 1,29 % del total, pero cuando se observa que se subdivide, a su vez, en las categorías “debate”, “discurso”, “entrevista”, “magacines y variedades”, “otros”, “retransmisiones deportivas”, “sorteos y concursos” y “tertulia” es forzoso concluir que el porcentaje del corpus que no proviene de algún tipo de escritura profesionalizada es muy marginal. Es, por esta razón, un mal lugar donde ir a buscar coloquialismos. No quiere decir que no los contenga, o que eso lo convierta en una mala herramienta de trabajo. Simplemente, se trata de un lugar donde está ausente un numeroso repertorio de formas coloquiales, usadas cotidianamente por millones de hablantes de la lengua pero que no llegan a publicarse. Es, en cambio, una excelente herramienta para observar usos más establecidos, aun si son populares o coloquiales,

que superan las diferentes instancias de selección y corrección propias de los espacios en donde la lengua es material de desenvolvimiento profesional, como revistas, documentos oficiales, novelas, diarios, etc.

El corpus que se puede reunir de Twitter tiene un sesgo igualmente pronunciado, pero en el sentido opuesto. Es escritura que no debe superar más instancia que aquella a la que la somete el usuario individual, en busca de sus propios rendimientos. Es una fuente prácticamente insuperable para coloquialismos y neologismos actuales, y, dado que permite reunir una cantidad de datos inconcebible para cualquier corpus tradicional, las frecuencias normalizadas que arroja son de una calidad excelente. Como ya queda consignado al principio de este artículo, el corpus del cual surge el presente estudio supera los 1150 millones de palabras, mientras que CORPES XXI está actualmente en algo más de 350. En un trabajo hecho esencialmente por dos personas (Matías Guerson ampliando lo hecho por Damian Aleman, en una etapa anterior del proyecto) se reunieron más de tres veces la cantidad de palabras que CORPES XXI en los años que lleva de desarrollo, con varios equipos de trabajo en España y uno en Argentina. Por supuesto, las diferencias son muy grandes más allá de la cantidad de palabras. CORPES XXI es un corpus anotado, lo cual significa que prácticamente cada palabra tiene información adicional que indica si se trata de una preposición, un sustantivo (en singular o en plural, masculino o femenino), un verbo (con datos sobre la persona, el modo, el tiempo y el número), etc. Nuestro corpus fue recolectado por métodos automáticos y no está anotado. El número tanto mayor de palabras es necesario para compensar, aunque sea en parte, esa información faltante.

En cuanto a la edad de los usuarios cuyos tuits conforman el corpus, es poco lo que puede decirse de manera fehaciente. En general, no se trata de una red social que esté hoy especializada en un grupo etario específico o definido, como tal vez fue el caso

en sus comienzos. Parece claro que quienes primero adoptan una nueva red social son los más jóvenes y son ellos quienes, por un tiempo, definen en gran medida la identidad del nuevo espacio. Ciertamente, es eso lo que puede decirse actualmente de TikTok, donde es común todavía que exista la idea de que hay una suerte de límite de edad. BBC en español, por ejemplo, agrega ese rasgo en su caracterización: “Con unos 500 millones de usuarios activos, TikTok es una de las aplicaciones de móviles más populares del mundo, principalmente entre los adolescentes y jóvenes” (BBC, 8/11/2019).⁶ También es significativo que todavía sea posible hacer humor con la idea de que alguien es “demasiado viejo” para estar en TikTok. Por ejemplo, el video “Too Old for TikTok”⁷ publicado en Youtube por Jon Lajoie, un usuario con 1,37 millones de suscriptores, tiene más de trescientas mil vistas y termina con la frase “tengo 40 años, esto no luce bien. Mi edad es más apropiada para Facebook”. No obstante, conviene recordar que ese era el caso también de Facebook, donde era un requisito, en sus inicios, tener una dirección de email con dominio universitario para poder abrir una cuenta, lo cual hacía forzoso que la gran mayoría de los usuarios tuviera veinticinco años o menos. Aunque resulte hoy difícil de creer, Facebook estuvo también, en su momento, asociada primariamente a jóvenes y adolescentes. A medida que la red social se mantiene en el tiempo y aumenta la edad de sus usuarios, esta asociación empieza a debilitarse. Ese parece ser el camino recorrido por Twitter desde su lanzamiento en julio de 2006, hasta un presente en el que es un medio de comunicación elegido por políticos y figuras públicas de todas las edades para comunicarse de manera directa con un gran público. Otro indicio importante, derivado de este uso, es el hecho de que es cada vez más

⁶ <https://www.bbc.com/mundo/noticias-50341717>

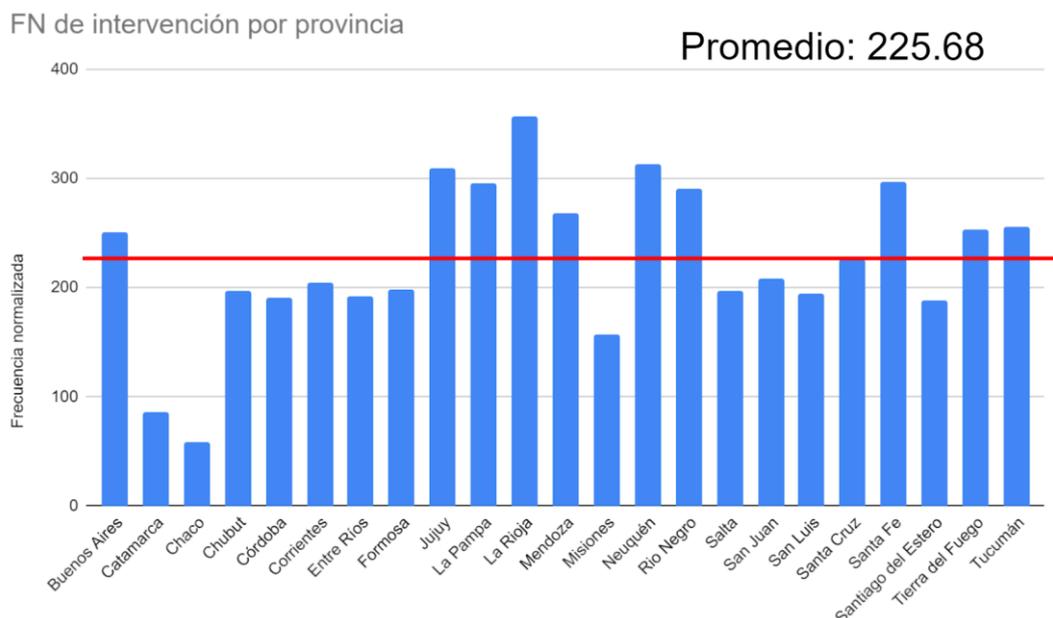
⁷ <https://www.youtube.com/watch?v=DZ2ZB-zcy9A>

común que Twitter sea fuente de noticias publicadas en medios como la radio, la televisión o los diarios.

Todos estos elementos parecen indicar que, fuera, acaso, de los adultos mayores de setenta años, no existen grupos etarios ausentes de Twitter. No sabemos si esto es algo que se va a mantener en el futuro, pero, por el momento, no solo es una excelente fuente de datos lingüísticos porque, a diferencia de Whatsapp o el propio Facebook, todo lo publicado ahí es de acceso libre salvo que se indique lo contrario, sino porque cuenta entre sus usuarios a población que pertenece a casi todas las variables demográficas en términos de edad, clase social, nivel educativo y localización.

Resultados

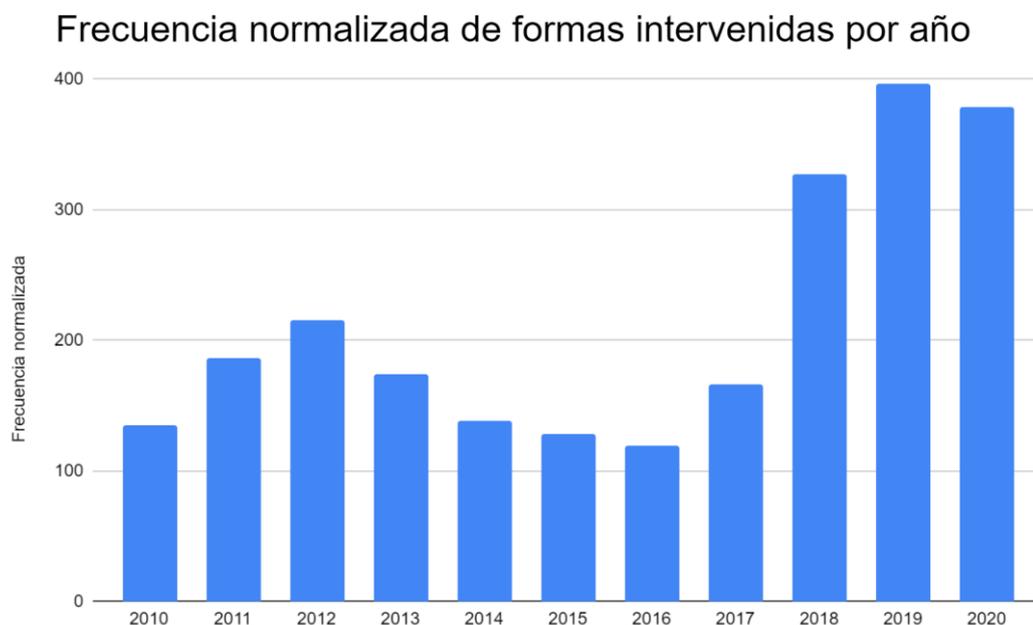
Cuando se detallan las frecuencias de las formas intervenidas en general, es decir, para todas las variantes contempladas, desde el punto de vista territorial, surgen algunos contrastes notables. Por encima del promedio, se destacan Jujuy, La Rioja y Neuquén, mientras que las provincias con frecuencias muy por debajo del promedio son Catamarca y Chaco. Los valores que muestran estas últimas tienen diferencias tan importantes, que logran bajar el promedio nacional en alrededor de 15 puntos.



Ese valor promedio quiere decir, entonces, que por cada millón de palabras del corpus, aparecen 225,68 formas intervenidas de alguna de las maneras contempladas por el corpus: un desdoblamiento masculino-femenino o femenino-masculino, desarrollado o con barra, arroba, letra *x* o vocal desinencial *e*, y cada una de estas variantes en plural o singular. Dicho de otro modo, es preciso leer 4431 para encontrar una palabra intervenida de alguna de estas maneras en el corpus. Esta frecuencia es comparable a las que tienen, en CORPES XXI, estructuras como el potencial compuesto (*habría ido, habríamos querido*, etc.), con 117,2, o del infinitivo compuesto (*haber ido, haber querido*, etc.), con 345,9. Esta característica podría inducir a considerar que estas intervenciones son comparables con estructuras del sistema lingüístico, pero es necesario considerar como término de comparación, no estructuras relativamente poco frecuentes como las mencionadas, sino el propio masculino no marcado. Así, en relación con aquel rasgo que es objeto de la intervención, se puede dimensionar preliminarmente su alcance.

No existen maneras precisas de cuantificar la frecuencia del masculino genérico ya que es un rasgo que depende de interpretaciones que están por encima de lo que se hace (¿o puede hacerse?) con las técnicas y tecnologías disponibles en la actualidad para el procesamiento de lenguaje natural. Sin embargo, puede asumirse que se trata de frecuencias muy altas. El infinitivo simple (*andar, saber, caminar*, etc.), por caso, tiene en CORPES XXI una frecuencia de 28.568, esto significa que aparece un infinitivo cada 35 palabras. Es plausible, sin contar con el dato concreto, que la del masculino no marcado se parezca a la frecuencia normalizada del infinitivo simple. En este punto, cabe lanzar un desafío a la comunidad de especialistas en programación: tiene un enorme interés para la investigación de este fenómeno desarrollar métodos confiables de cuantificar el masculino no marcado en español. Es altamente probable, además, dado el rumbo que ha tomado la consideración pública de este tema, que ese potencial desarrollo tecnológico encuentre aplicaciones muy concretas, más allá de las puramente académicas o científicas. Por ahora, no obstante, es preciso apoyarse en otros indicadores, como los que se verán a continuación.

Particularmente reveladoras de la naturaleza del fenómeno son las frecuencias que muestra la intervención a través de los años.



Parece claro, más allá de un uso estable con frecuencias normalizadas de un mínimo de 100 que pueden atender demandas y necesidades diversas, que los aumentos y las disminuciones en el uso por encima de ese número guardan una relación estrecha con eventos de la vida política del país. La curva ascendente que culmina en el pico de 2012 coincide con el tratamiento y aprobación de la Ley de Identidad de Género. Pasado ese hito, lo que se ve es un descenso muy considerable que no se revierte hasta el año 2017, cuando se movilizan los sectores afines a la presentación del Proyecto de Ley de Interrupción Voluntaria del Embarazo (IVE), algo que sucedió al año siguiente. A partir de ese momento, la intervención parece asociarse más directamente con ese evento legislativo, aunque esa asociación más directa nunca implica que abandone los demás temas de la agenda de género que son de interés para sus usuarios. Lo que se ve que alterna es, entonces, la preponderancia de la asociación de la intervención con cada uno de esos distintos temas. Es decir, periodos más centrados, por ejemplo, en

transexualidad/no binarismo, y periodos más centrados en derechos de la mujer/aborto, dependiendo de los objetivos políticos o legislativos perseguidos en cada uno de ellos.

La diferencia que puede verse entre lo que sucede después de 2012 y lo que muestran los datos después de 2018 es un indicio clave para caracterizar ese rasgo central del uso. En el año 2012 la Ley de Identidad de Género resultó aprobada, en cambio, en 2018, el Proyecto de Ley IVE no logró superar la instancia en el Senado. Parece muy probable que ese traspie legislativo haya animado a muchos usuarios de lenguaje inclusivo a redoblar los esfuerzos durante el año siguiente, en 2019. Esto refuerza la idea de que la intervención tiene un marcado componente político, en tanto rasgo saliente de la configuración discursiva de diversos grupos activos en temas de igualdad de género.

La evolución en cuanto a la cantidad de usuarios detrás de estas fluctuaciones describe una trayectoria que parece desacoplada de la que describe el uso para el mismo periodo. El siguiente cuadro muestra el porcentaje de usuarios de Twitter que usaron al menos una forma intervenida durante cada año.



Este puede deberse a un fenómeno paralelo pero diferenciado. A medida que la visibilidad de la intervención fue en aumento, también empezó a ser objeto de conversaciones en Twitter, por un público mayor, no necesariamente comprometido con las causas sociales y políticas con las que está vinculada la intervención desde sus inicios. En estas conversaciones, empezó a ser común que alguna forma intervenida sea usada con ánimo de descalificar su uso, estigmatizarlo o, incluso, insultar a sus usuarios. Inevitablemente, estos casos pasan también a engrosar los diferentes valores del corpus, tanto en términos de frecuencias normalizadas de las formas intervenidas, como en porcentajes de usuarios que las usan al menos una vez. No constituye esto un problema, puesto que, más allá del ánimo o la intención de quien incluye una forma intervenida en un tuit, la idea del corpus es cuantificar la extensión del fenómeno y un uso adverso también es una de las maneras en las que el fenómeno alcanza a más hablantes. De todas formas, a pesar de que no hay métodos sencillos de automatizar la detección de estos usos, todo parece indicar que son minoritarios. La curva de cantidad de usuarios es una medida que acompaña el nivel de conocimiento público de la intervención y no necesariamente su uso. De hecho, llamativamente, un primer pico de cantidad de usuarios (cerca del 30%) se dio en el piso histórico de uso para la década. Esto sugiere que se estaban empezando a dar las condiciones para el debate público sobre la intervención y sus temas relacionados que tendría lugar con su máxima intensidad a partir del año 2018. Es decir, estaba aumentando el nivel de conocimiento público y se estaban empezando a definir las posturas que su ubicarían luego en extremos opuestos de la polémica.

El otro desacople interesante es el de 2020. La frecuencia de uso mantiene los niveles altos de 2018 y 2019, pero el porcentaje de usuarios que usan formas intervenidas al menos una vez desciende de un modo abrupto. Si se tiene en cuenta que

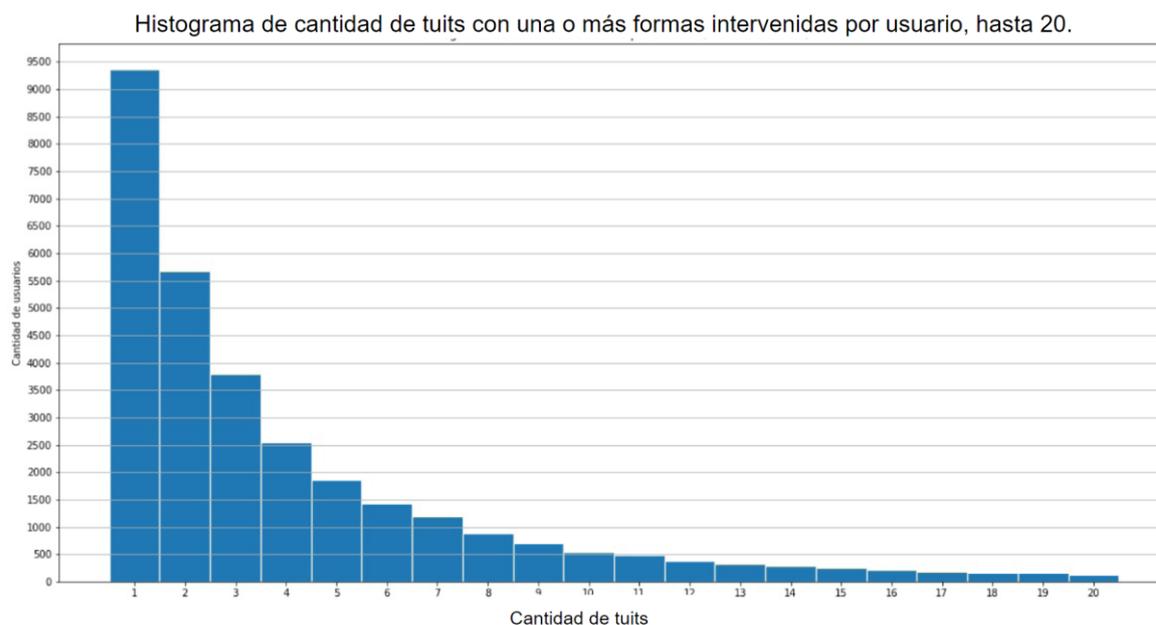
el corpus solo cuenta con datos de enero y febrero para ese año, una posible explicación tiene que ver con el contexto de esos dos meses: la pandemia, que tomaría por asalto la agenda pública a partir de marzo, era todavía un evento que ocurría en países lejanos y el nuevo gobierno, inaugurado un mes antes, había generado varias noticias importantes alrededor de la intervención. Notablemente, que PAMI utilizaría el llamado lenguaje inclusivo en la redacción de disposiciones⁸ y la propuesta de parte de la ministra de género de la provincia de Buenos Aires para que se use en la administración estatal.⁹ Como es esperable, estos anuncios capturaron inmediatamente la atención de los medios de prensa y generaron una intensificación del debate en redes sociales que impactó de manera directa en nuestro corpus. Esta mayor intensidad, combinada con el receso de verano, puede proveer una explicación satisfactoria para la contradicción aparente de que se hayan mantenidos los altos niveles de frecuencia de los años anteriores ante una disminución porcentual tan notable de sus usuarios.

Usuarios por cantidad de formas intervenidas

Un indicio adicional de que muchos de los indicadores están siendo alimentados no por usuarios activos de la intervención sino por “comentadores” lo da el hecho de que, con casi 9500 usuarios, la columna más populosa del siguiente cuadro sea la de quienes tienen una sola aparición de alguna forma intervenida en toda su historia con la red social incluida en el corpus.

⁸ “La directora del PAMI estableció que todas las disposiciones sean escritas en lenguaje inclusivo”, [Infobae, 10/1/2020](#).

⁹ “La ministra de Género bonaerense propuso utilizar el lenguaje inclusivo en toda la administración pública de la provincia y ahora debe definirlo Axel Kicillof”, [Infobae, 21/1/2020](#).

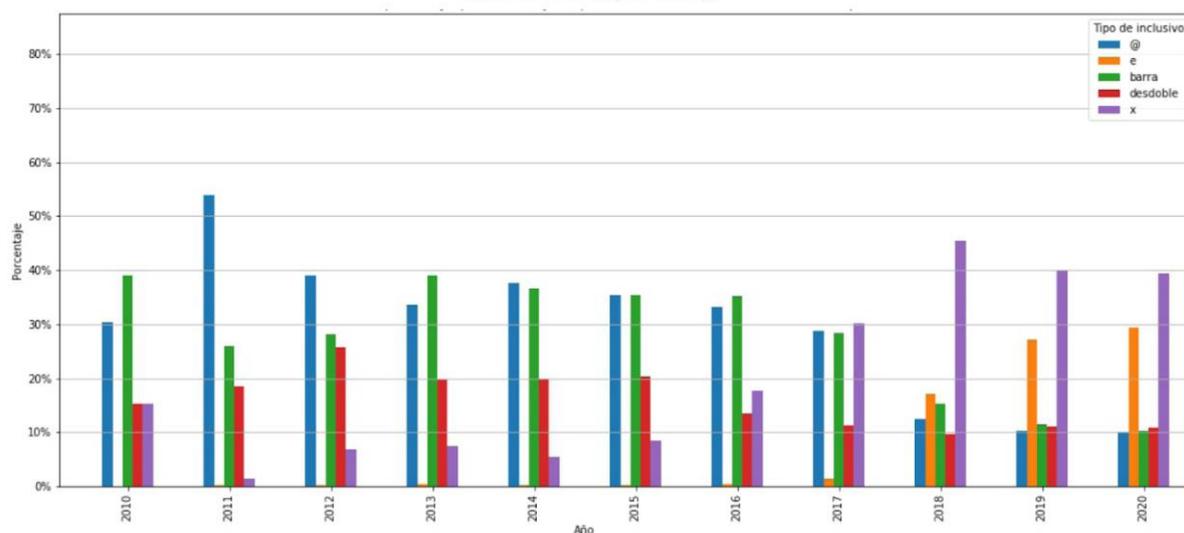


Este es un hecho interesante en sí mismo para caracterizar al menos unos de los usos de la intervención, pero sobre todo presenta un obstáculo difícil de franquear para quienes quieran sostener que estamos ante un fenómeno gramaticalizado, aun de manera incipiente. La curva de un fenómeno gramatical tendría la forma de un sombrero de ala como el del Principito (en realidad, para más precisión, de boa que acaba de ingerir un elefante), puesto que lo más inusual serían los extremos, esto es, que una palabra o estructura del sistema lingüístico se usara una sola vez o más de quince en las publicaciones de una persona en Twitter.

Evolución de variantes

Al discriminar por tipo de intervención, aparece una trayectoria desde los desdoblamientos (o bien desarrollados *-todos y todas-* o abreviados por medio de la barra *-todos/as-*) y la arroba (*tod@s*), que predominaron durante casi dos tercios de la década, hacia el uso de *x* (*todxs*) y *e* (*todes*), que son preponderantes a partir de 2017.

Formas intervenidas por año y tipo (porcentaje respecto del total de tuits con presencia de intervención, por año).



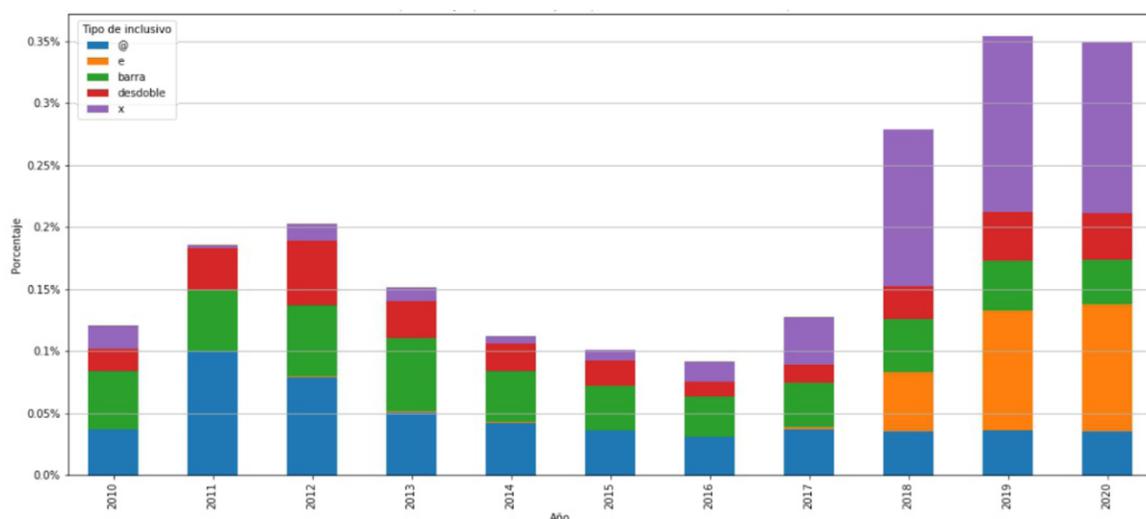
Esto puede deberse a cuestiones de distinto orden. En primer término, más allá de que tiene el beneficio de ser pronunciable, el desdoblamiento desarrollado es algo costoso estilísticamente, puesto que puede generar secuencias largas. Esto animó a que se buscara una opción abreviada por medio del signo arroba. Una solución, que, durante los años 2011 y 2012, muestra niveles de adhesión altos, para luego empezar un lento declive que parece coincidir con el aumento de visibilidad pública de las demandas de la comunidad transexual, sobre todo en beneficio de la letra *x* en reemplazo de la vocal desinencial, como superación del binomio tradicional que se consideraba contenido, gráficamente, dentro de ese signo. La necesidad de encontrar una forma que pudiera trascender los límites de la escritura resultó en la adopción de la vocal *e*, dado su rol histórico en muchos epicenos,¹⁰ como los adjetivos *grande* o *libre*, los sustantivos *cantante* o *amante* e, incluso, en pronombres como el dativo *le* o *les*, por contraste con otros que marcan género como *él*, *ella*, *lo*, *la*, *nosotros*, *nosotras*, etc. Esta característica

¹⁰ Yarubi Diaz Colmenares & David Heap. *Variation et changement dans les accords du français inclusif*. Congrès de l'Association canadienne de linguistique, 30 mai au 1er juin, 2020.

forma parte de un sistema, en español, en el que predominan las palabras masculinas terminadas con la vocal *o*, las femeninas terminadas con la vocal *a* y las epicenas terminadas con la vocal *e*. Este patrón general es el que da cuenta de que los hablantes hayan derivado asociaciones prototípicas entre las tres vocales desinenciales de la lengua y el género, aun cuando existen numerosas excepciones en cada caso (por ejemplo, las palabras femeninas *mano*, *libido*, *modelo*, etc., las masculinas *mapa*, *problema*, *poeta*, etc., y los plurales de sustantivos deverbales terminados en *-or*, que tienen vocal desinencial *e* pero son todos masculinos, como *administradores*, *lectores*, *profesores*, etc.). Este patrón puede también citarse como la razón de que no se haya preferido alguna de las dos vocales restantes (*i* o *u*) para la intervención, cuando no había, en principio, ningún impedimento para hacerlo.

Cuando se proyecta esta evolución sobre la totalidad de los tuits disponibles de cada año y se apilan todas las variantes en una sola columna, se obtiene una imagen clara de los alcances de la intervención en cada momento, que, en tanto ambas son medidas normalizadas, prácticamente, reproduce la trayectoria que describen las frecuencias normalizadas en el periodo.

Formas intervenidas por año y tipo (porcentaje respecto del total de tuits por año).



El año con mayor uso es 2019 con un porcentaje apenas por encima del 0,35 % del total. Esto significa que el 99,65 % de los tuits de ese año no presentan ninguna forma intervenida, en ninguna de sus variantes. El contraste entre estos valores y aquellos de usuarios con alguna forma intervenida, que rondan el 30 % de 2016 en adelante, hace evidente la empinada diferencia que existe entre la visibilidad pública de la intervención, el alto nivel de conocimiento que de ella tienen los hablantes, y sus muy modestos alcances como práctica concreta con la lengua. Se trata de un dato que, desde un ángulo diferente, también desautoriza cualquier hipótesis de gramaticalización y soporta la noción de que se trata de una acción explícita, conciente y calculada cuyo funcionamiento se captura más adecuadamente con herramientas provenientes del análisis del discurso que con aquellas propias de la descripción gramatical o lingüística. Es decir, aunque no avale la hipótesis de que forma parte del sistema lingüístico de los hablantes, la evidencia que arroja el corpus vuelve innegable que la intervención forma parte ya de su sistema discursivo/retórico.

Repertorio

Otra pregunta central para el análisis es si la intervención se da más o menos homogéneamente sobre todas las palabras pertinentes (aquellas referidas a grupos o individuos humanos) o si, por el contrario, se enfoca en ciertos elementos léxicos concretos dentro de una lista limitada. El siguiente cuadro muestra un detalle de los veinte masculinos no marcados más intervenidos dentro del corpus, separados por tipo de variante.

	Desdoblamiento barra	Desdoblamiento y/o	@	x	e
1	amigo (9856)	todo (7970)	todo (11608)	el (15460)	todo (9405)
2	el (7396)	el (3807)	amigo (9074)	todo (11431)	chico (7426)
3	todo (3337)	amigo (2998)	el (5229)	amigo (10147)	amigo (6108)
4	novio (3041)	chico (1788)	chico (1676)	chico (4462)	hijo (1583)
5	hijo (1558)	vecino (1781)	novio (1321)	hijo (2291)	nosotros (995)
6	chico (1205)	trabajador (668)	bienvenido (1266)	uno (2120)	muchacho (617)
7	boludo (858)	nene (662)	hijo (1010)	otro (1351)	otro (377)
8	uno (699)	pibe (524)	boludo (701)	novio (1315)	diputado (186)
9	hermano (683)	hermano (286)	solo (535)	nosotros (1259)	novio (175)
10	otro (632)	hijo (276)	alguno (528)	alguno (1245)	alumno (161)
11	alguno (593)	alumno (238)	otro (524)	boludo (1115)	uno (153)
12	pibe (550)	novio (236)	puto (500)	pibe (1060)	querido (129)
13	solo (543)	argentino (199)	agradecer (462)	lo (969)	bienvenido (123)
14	lindo (528)	neuquino (164)	afortunado (461)	trabajador (837)	sobrino (81)
15	lo (473)	solo (146)	hermano (439)	hermano (814)	egresar (64)
16	vecino (385)	abuelo (144)	vecino (410)	solo (794)	mogolico (39)
17	ese (348)	tio (144)	nosotros (379)	mucho (675)	jubilado (35)
18	profesor (341)	mucho (140)	loco (372)	lindo (625)	mismo (35)
19	sobrino (331)	maestro (126)	mucho (348)	vecino (602)	interesado (30)
20	nene (324)	rojano (125)	lindo (347)	loco (540)	indignar (29)

En el cuadro se incluyen 180.156 instancias de palabras intervenidas de 241.819, es decir, el 74,5 % del total. De estos 100 lugares que tiene el cuadro surge una lista de 45 palabras diferentes. Las palabras se incluyen dentro del cuadro en masculino singular. Esto significa que, si elegimos una palabra cualquiera, por ejemplo *novio*, las variantes incluidas son, por columna y de izquierda a derecha: *novios/as*, *novios* y *novias* (también, *novio o novia*), *novi@s*, *novixs*, *novies*. En el caso del artículo masculino singular *el*, las variantes, en el mismo orden, serían: *los/as*, *los* y *las* (también, *el o la*), *l@s*, *lxs*, *les*.

Si se reúnen estas 45 palabras y se suman todos los casos de cada una, lo que surge es el siguiente cuadro.



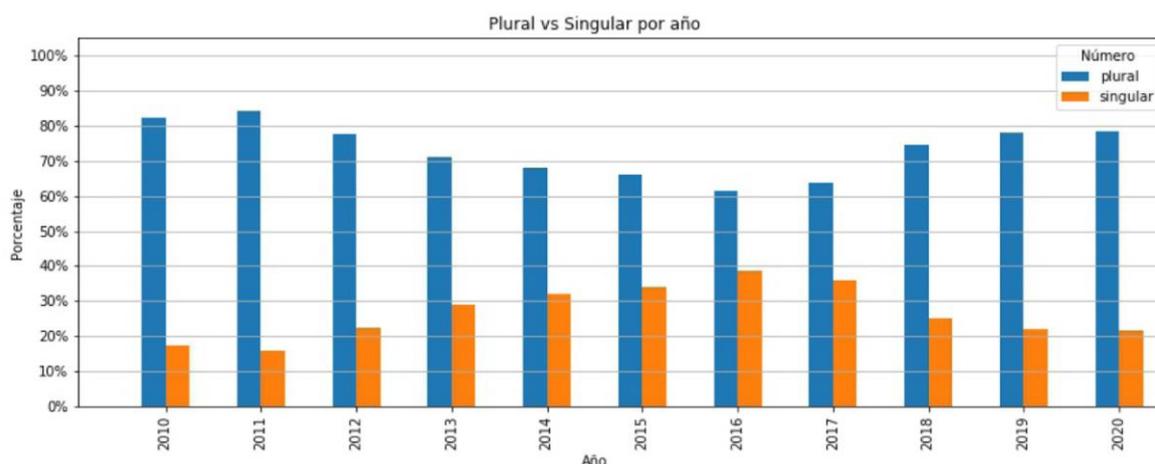
La imagen que emerge así es contundente: el 72,37 % de todas las instancias de intervención reunidas en el cuadro se concentran en solo cuatro palabras (*todo*, *amigo*, *el* y *chico*), mientras que el 27,63 % se reparte entre las 41 palabras restantes. Esta distribución es compatible con la hipótesis de que se trata de un fenómeno de orden léxico: parecen haberse identificado algunas palabras como emblemáticas receptoras de la intervención y son esas las que se usan mayoritariamente. Esto permite proponer que estos hablantes, antes que ajustando o modificando el modo en que se estructura el género gramatical en español, están incorporando una serie relativamente restringida de marcadores discursivos neológicos, pasibles de ser insertados voluntariamente a lapsos variables, sin incurrir en el gasto de energía derivado del intento de hacer todas las modificaciones exhaustivamente. De este modo, se va actualizando el efecto retórico, discursivo y performativo de la intervención, pero merced a un esfuerzo moderado que no distrae recursos de procesamiento valiosos al momento de dar forma a un enunciado.

No puede anticiparse si este repertorio va a sufrir modificaciones en el futuro. El proceso por medio del cual ciertas palabras se vuelven receptoras emblemáticas de la intervención es relativamente prolongado en el tiempo y requiere de una cantidad muy considerable de usos para que se consoliden en ese rol. Es imposible, en este punto, saber si otras palabras están en camino de transitar ese proceso, pero los datos muestran que las cuatro palabras más intervenidas de nuestro corpus ya lo han completado, lo cual las convierte en objeto de interés para quienes hacen lexicografía en lengua española. En especial si es integral, puesto que, según lo que se puede observar de modo general, no parece ser un fenómeno restringido a un país o a una región en particular. Claro que, para sustanciar esta última observación, sería preciso darle alcance panhispánico al presente estudio.

Plural vs. singular

Por último, existe la pregunta acerca de la cantidad de usos que se da en plural, la cantidad que se da en singular y acerca de las conclusiones que pueden derivarse de la proporción entre ambos.

El siguiente cuadro reúne todas las variantes discriminadas de esta manera.

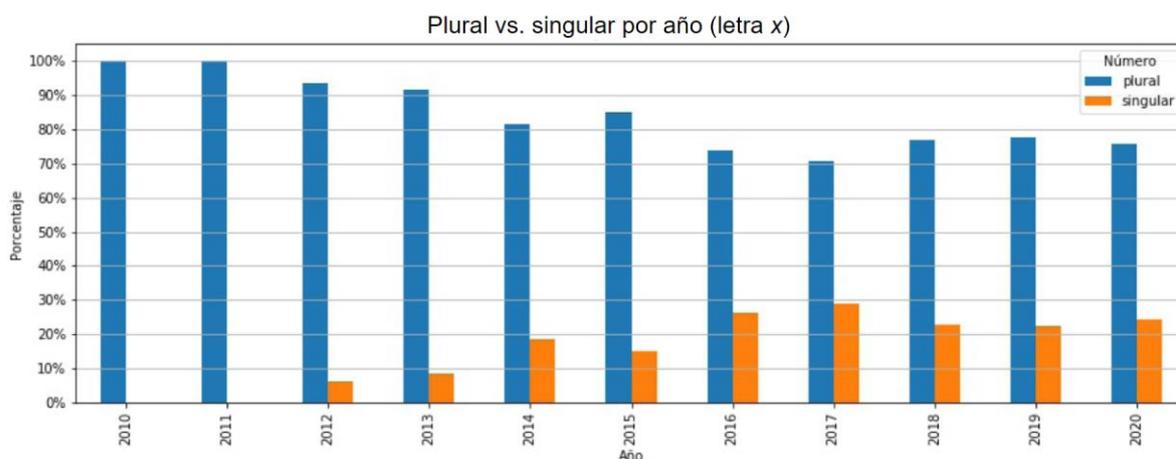


El primer hecho notable es que la brecha entre ambos usos fluctúa considerablemente. Otro hecho, también muy notable, es que el momento de mayor paridad coincide con el piso histórico de uso de la intervención en general. Tienen mucho interés estas fluctuaciones, en una intervención que se asocia primariamente con los feminismos, porque muestran su permeabilidad en cuanto a los diferentes temas de la agenda de género que puede acoger. El plural está más típicamente asociado con evitar el uso de masculino no marcado para designar grupos humanos compuestos por hombres y mujeres, aunque también incluyan otras identidades. Su uso en singular puede asociarse, y lo hace frecuentemente, con la cuestión de la sexualidad no binaria.

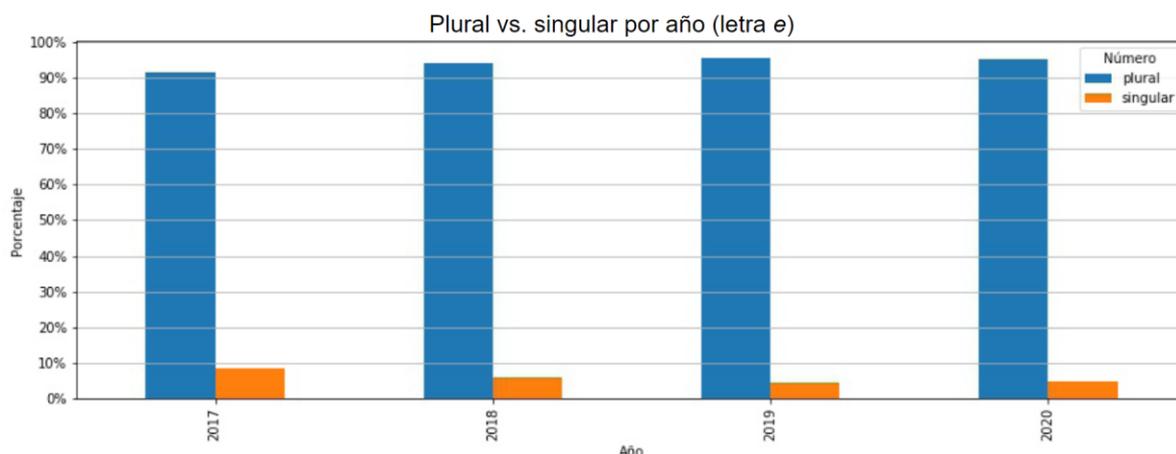
Las proporciones entre ambos usos que se observan en el corpus podrían sugerir que se abrió un periodo, tras la aprobación de la Ley de Identidad de Género en el año 2012, en el que la atención pública fue virando paulatinamente, desde la preocupación tradicional de la intervención más dominada por temas de la mujer, hacia las problemáticas de personas transexuales y no binarias. A medida que ese hito legislativo se alejaba en el tiempo y sin otras instancias comparables que enfoquen la agenda pública en temáticas de género, la intervención fue mermando su uso en general, pero manteniendo el foco en los temas que se habían instalado en 2012 a raíz de la ley. Así,

podrían explicarse dos hechos concomitantes: la disminución del uso de todas las variantes en el periodo 2012-2016 que muestran las frecuencias normalizadas y el aumento, paulatino y simultáneo, de la proporción de usos en singular respecto del plural. Esta situación daría un giro abrupto, a partir de 2017, cuando se preparó el proyecto de Ley de Interrupción Voluntaria del Embarazo que se presentaría al año siguiente, y los temas de la mujer volvieron a dominar la intervención. Si esta sucesión temporal es correcta, daría cuenta del proceso por medio del cual tomó forma el escenario que llega hasta el presente en el país.

Estos elementos permiten interpretar también la evolución del singular con *x*, que, como se ve en el cuadro siguiente, hace su aparición en el corpus durante el año 2012, cuando los desdoblamientos y su forma abreviada, la arroba, llevaban tiempo bajo presión porque se consideraba que no atendían a las demandas de quienes no se identificaban con alguna de las opciones de género dominantes. Esa aparición coincide con la mencionada sanción de la Ley de Identidad de Género, su uso llega a un máximo durante el año 2017 y mantiene niveles que rondan el 25 % durante los años siguientes, lo cual parece indicar que es la variante más consistentemente asociada con temas de sexualidad no binaria.



Un contraste notable respecto de esta situación es que la aparición más tardía de la variante con *e* puede haber obstruido, en parte, su asociación con temas de sexualidad no binaria en un contexto ya dominado por la cuestión de la mujer y la interrupción voluntaria del embarazo.



Este hecho es el que puede dar cuenta de que es la variante que más claramente se presenta como asociada con el uso en plural. Sin abandonar por completo la temática no binaria, que, aun cuando no es incompatible con el plural, puede estar representada por estos porcentajes que rondan el 6 % de promedio de uso en singular, la variante parece asociarse claramente con el objetivo buscado en el año 2018 (ley de IVE), que, adicionalmente, quedó pendiente para el futuro.

Este último punto es el que hace esperable que el uso de la intervención en general, y no ya solo el de la variante con *e*, tenga un comportamiento diferente al declive sostenido que mostró entre 2012 y 2016. Habrá que esperar a una nueva recolección de datos para saber si estas frecuencias más altas conforman una tendencia que se mantiene, merced, o no, a un nuevo potencial tratamiento legislativo del proyecto

de ley de IVE, si disminuye debido a algún factor, como puede ser la profunda crisis que precipitó la pandemia del COVID-19, o si se abre un nuevo escenario diferente.