

LA CONCORDANCIA *AD SENSUM*

Carlota de Benito Moreno

1. Introducción

A pesar de la vocación de internacionalidad con que Chambers (2000, 2004) acuñó el concepto de universal vernáculo (*vernacular universal*), este no parece haber alcanzado mucho éxito en el estudio del español, incluso con la importancia que la sociolingüística de corte anglosajón tiene en la lingüística hispánica¹. Los editores de este volumen avanzan algunos de los posibles motivos en su propuesta, como la poca importancia dada a la lengua de los semicultos. En mi opinión esta falta de éxito puede también achacarse a una insuficiente caracterización teórica del fenómeno, que dificulta su encaje en la tradición hispánica.

Chambers (2000, 2004) parte de una observación empírica, a saber, la existencia de variantes subestándar comunes a una mayoría o la generalidad de las variedades vernáculos del inglés. Esto es lo que Thomas (2008 [1994]) llama vernáculo general (*general vernacular*) en su descripción del inglés galés, término con el que identifica variantes no estándar presentes en esta variedad, que no son exclusivas de ella. Para el español, estas variantes se suelen clasificar como vulgarismos generales (García de Diego 1978; Parodi y Santa Ana 1997; Benito Moreno 2020). Chambers (2000, 2004) busca ofrecer amparo teórico a este tipo de fenómenos dentro de la Sociolingüística entendida como teoría general del lenguaje y para ello se basa en tres oposiciones: 1) la oposición entre estándar (como variedad no regional y propia de las clases urbanas educadas) y variedades vernáculos (marcadas diatópica, diastrática o diafásicamente); 2) la oposición entre rasgos primitivos, que surgen de forma natural

¹ Agradezco profundamente los comentarios realizados a versiones o partes del manuscrito original a Araceli López Serena, Johannes Kabatek, Irene Areses y Yoselin Henriques, que me han permitido mejorarlo. Por supuesto, cualquier error que subsista es solo mío.

y son omnipresentes en la lengua, y rasgos aprendidos, de distribución limitada; y 3) la diferencia entre difusión de un rasgo lingüístico y poligénesis. A partir de estas oposiciones, propone que existe una serie de rasgos primitivos que aparecen independientemente en todas las hablas vernáculas y que, por tanto, son universales. Por su carácter primitivo, deberían encontrarse también en otras lenguas. Estos son los universales vernáculos.

En mi opinión, sin embargo, esta presentación teórica adolece de varias confusiones, la más importante de las cuales tiene que ver con el uso de *universal* en dos sentidos distintos. Por un lado, Chambers colige la universalidad de estos rasgos de su ubicuidad en las diferentes variedades del inglés, equiparando *universal* con ‘general’ o ‘ampliamente difundido’. Por otro lado, Chambers aduce que la generalidad de estos rasgos se debe a que los mecanismos subyacentes que justifican su origen son universales en un sentido chomskiano, es decir, a que pertenecen a la facultad general del lenguaje, equiparando *universal* a ‘natural’. Al hacer confluir estos dos sentidos de *universal* se olvida la distinción crucial entre innovación y difusión, que, si bien fue plenamente elaborada por Coseriu (1958, 1973), fue popularizada por Weinreich, Labov y Herzog (1968) dentro del ámbito de la sociolingüística anglosajona. La innovación es un acto individual, mientras que la difusión es un proceso social y, por lo tanto, las motivaciones tras ambos procesos son distintas: equipararlas supone una visión teleológica del cambio lingüístico. En otras palabras, no todas las formas difundidas en todas las variedades subestándar de una lengua se deben a una poligénesis causada por mecanismos universales². Por ejemplo, la doble negación del inglés, ampliamente extendida en las variedades vernáculas pero censurada en el estándar, presenta esta distribución por haberse perdido en esta variedad (como indica Trudgill 2009 en su crítica a Chambers).

A mi juicio, en la descripción de estas formas vernáculas generales (Thomas 2008 [1994]), necesitamos una mayor claridad teórica, para lo que puede auxiliarnos la lingüística de las variedades alemana de impronta coseriana. Por ello, creo fundamental que mantengamos las siguientes distinciones:

1. Los planos del hablar: universal (de la actividad del hablar) e histórico (de las lenguas). Esta distinción se halla *mutatis mutandis* en la propuesta de

² Como tampoco todas las innovaciones que responden a motivaciones universales han de difundirse en todas las variedades de una lengua: la poligénesis es esperable, pero el triunfo de la innovación no es inevitable.

Chambers, como se observa en la siguiente descripción de lo que él denomina “default singulars” (extensión de las formas del singular al plural en el paradigma verbal)³:

My main point in all this is to show that there is a fundamental difference between attempts at explaining the primitiveness of default singulars and attempts at explaining the subject hierarchy that promotes them. The latter is accidental, couched as it is in language-specific idiosyncrasies such as numberless *there* and syncretic *you*. [...] Default singulars and the other root vernacular features, by contrast, cannot be language-specific. They must be universal. As primitive features of the language faculty, they cannot be English-specific (Chambers 2004: 141-142).

2. La ubicación del fenómeno en el espacio variacional de la lengua. La distinción entre estándar y vernáculo usada por Chambers (2000, 2004) y Trudgill (2009) nos parece insuficiente. Estos autores utilizan *vernáculo* como ‘subestándar’ y lo ubican en el nivel diastrático, aunque Trudgill indica que otro uso posible del término es el de ‘informal o no monitorizado’, ubicándolo en el nivel diafásico. De hecho, Trudgill objeta a los planteamientos de Chambers que algunos de los fenómenos que este considera universales vernáculos se encuentran también en variedades estándar⁴. Así, se hace patente la necesidad de introducir el nivel concepcional, que, en el modelo de la cadena de variedades de Koch y Oesterreicher (2007 [1990]) es el único que tiene tanto una dimensión universal como una dimensión histórica.

³ Cuestión distinta es que su observación sea certera: Trudgill rechaza la calificación de universal de estos “singulares por defecto” y nota que el fenómeno universal subyacente es la “regularización”, pues en algunas variedades lo que se observa es la extensión de las formas plurales. Siendo así, estos singulares por defecto del inglés no obligan a interpretar los fenómenos de externalización del plural (entre los que se encuentra la concordancia *ad sensum* que nos ocupa en este trabajo) que Barrio de la Rosa (2020) describe como universal vernáculo del español (que el inglés no compartiría), puesto que los primeros serían reflejo de una tendencia a la regularización morfológica de los paradigmas, mientras que los segundos lo serían de la necesidad de marcar formalmente algunas distinciones semánticas, como el número. Aunque ambas tendencias puedan producir innovaciones opuestas, las dos son universales.

⁴ Coliden también la concepción del estándar como una variedad, propia de un grupo determinado, que ha sufrido un proceso de estandarización y que, por tanto, tiene hablantes nativos (Trudgill 1999), más frecuente en la tradición anglosajona, con la más frecuente en la tradición hispánica, de un modelo lingüístico adecuado en determinadas circunstancias (véase Rey Quesada 2021 para una discusión del concepto).

De acuerdo con estas distinciones, los universales vernáculos sobre los que teoriza Chambers son rasgos que se sitúan en el nivel universal de la inmediatez comunicativa y no deben confundirse con los rasgos marcados diatráticamente como bajos pero sin marcación diatópica, que corresponden a lo que se suele considerar vulgarismos generales en la tradición hispana —y que creo equivalentes a lo que Thomas (2008 [1994]) refiere con *general vernacular use*—⁵. Hace falta, por lo tanto, una categorización cuidadosa de los fenómenos que lista este autor: la simplificación de grupos consonánticos o la falta de concordancia en existenciales (*there's cars*), por ejemplo, encajan perfectamente con los ejemplos que Koch y Oesterreicher (2007 [1990]) mencionan dentro del plano universal de la inmediatez. Como señala López Serena (2002: 265), el carácter universal de este tipo de fenómenos reside en ser “el resultado de condiciones comunicativas universales, aunque su realización material solo sea posible a través de lenguas históricas particulares, donde pueden adoptar una estructuración diferente”. Dicha realización material, que se sitúa en el plano idiomático, tendrá su propia historia y podremos encontrar, dentro de una lengua particular concreta, que un fenómeno resultado de estas condiciones comunicativas universales adquiera una marcación diatrática o diatópica determinada, manteniéndose en otras variedades como un rasgo propio de la inmediatez comunicativa.

En este trabajo nos centramos en una construcción *ad sensum* —fenómeno que Koch y Oesterreicher (2007 [1990]) también adscriben al nivel universal—, concretado en la concordancia entre verbo y sujeto cuando este último es un sustantivo colectivo singular o un pronombre indefinido singular. Este fenómeno, que no ha sido muy investigado para el español, sí lo ha sido para el inglés, donde los estudios muestran que la concordancia plural (de algunos sustantivos) puede convertirse en un rasgo propio de algunas variedades o épocas (Levin 2001, 2006). Este trabajo busca contribuir al estudio de la concordancia de sentido en español con nuevas aportaciones empíricas, especialmente incluyendo la perspectiva geográfica, preguntándose si este fenómeno propio de la oralidad concepcional ha llegado a enraizarse en algunas variedades diatópicas concretas.

⁵ Otra cuestión, de gran interés, se refiere a la selección de rasgos en la estandarización y las motivaciones subyacentes, en las que podrían también encontrarse patrones generales o muy extendidos debidos precisamente a su marcación concepcional en el plano universal (Kabatek 2004: 385).

En el apartado 2 ofrecemos un estado de la cuestión sobre este fenómeno en español; en el 3 se establece la extensión geográfica de la concordancia semántica en el mundo hispanohablante (con datos de Twitter); y en el 4 se analizan los condicionantes lingüísticos del fenómeno en el habla rural de España. Se ofrecen algunas conclusiones en el apartado 5.

2. La concordancia *ad sensum* entre sujeto y verbo

Es bien sabido que los sujetos morfológicamente singulares con rasgos semánticos de pluralidad (como los sustantivos colectivos y pronombres indefinidos en singular) pueden desencadenar la concordancia en plural (es decir, de sentido) con el verbo. Puesto que esta es más frecuente cuando dichos sujetos encabezan construcciones partitivas o pseudopartitivas con una coda plural (como *la mayoría de las personas* o *una decena de manifestantes*), es este tipo de construcciones el que más interés ha despertado en la literatura. La concordancia variable puede entonces explicarse por los elementos morfológicamente singulares y plurales que contienen los sujetos (aunque dicha explicación puede basarse en varios análisis estructurales; RAE/ASALE 2009: § 33.8b y ss.).

Si no hay complemento preposicional plural, la concordancia *ad sensum* parece mucho menos frecuente. De hecho, cuando el sujeto y el verbo se encuentran en la misma oración y, además, son contiguos, se dice que es prácticamente inexistente o incluso agramatical o anacolútica (Bello 1888 [1847-1860]; Gili Gaya 1961; Ortega y Morera 1981-1982; Martínez García 1999; San Julián Solana 2018)⁶. Se ha notado, sin embargo, que existen factores que favorecen la concordancia en plural del verbo. Uno es la distancia entre este y el sujeto, ya sea lineal —es decir, concebida como número de palabras o elementos interpuestos entre el verbo y el sujeto (Gili Gaya 1961; Martínez García 1999; RAE/ASALE 2009: § 33.6f)— o estructural —concebida como pertenencia o no a la misma oración (Gili Gaya 1961; Bello 1888 [1847-1860]; Martínez García 1999; Soler Arechalde 2012 [*apud* Oggiani 2013]; RAE/ASALE 2009: § 33.6g)—. Se ha propuesto también que la posición del verbo respecto al sujeto tiene un efecto en la concordancia *ad sensum*, que

⁶ Hay algunas excepciones, como los sustantivos *parte*, *resto*, *mitad*, *tercio*, que parecen encontrar menos dificultades para concordar en plural con verbos y adjetivos (Bello 1888 [1847-1860]; Gili Gaya 1961).

sería imposible “[s]i el verbo va antepuesto y realiza su predicación antes de formularse el sujeto léxico” (Martínez García 1999: 2767-8). Atendiendo a la semántica, se ha apuntado que la concordancia en plural es más frecuente cuando los individuos referidos por el sustantivo singular presentan un bajo grado de determinación u homogeneidad (Gili Gaya 1961; Bello 1988 [1847-1860]; Fält 1972; Soler Arechalde 2012 [*apud* Oggiani 2013]), es decir, sustantivos como *enjambre* tienden a mantener la concordancia singular de forma más persistente que *gente*. Por último, Barrio de la Rosa (2020) ha sugerido que la concordancia plural sirve como marca formal de la interpretación distributiva del sujeto, lo que parece ir en la misma dirección que el grado de determinación.

Los trabajos basados en corpus confirman algunas y matizan otras de las afirmaciones anteriores. En primer lugar, si bien la concordancia verbal en plural con sujetos colectivos singulares es poco frecuente⁷, no puede dudarse de que existe y tiene vitalidad, además de no ser exclusivo del habla o la lengua escrita de la inmediatez. Sánchez Avendaño (2007) la documenta en un corpus de lengua hablada en Costa Rica; Al-Amar y Ababneh (2017) la encuentran también en el habla popular de Sevilla; Areses (2019) localiza numerosos ejemplos tanto en búsquedas en la red como en el CORPES; y Barrio de la Rosa (2020) la registra en el español rural europeo. El trabajo más exhaustivo hasta la fecha es el de Fält (1972), que expurga obras literarias, periódicos y revistas (todos posteriores a 1960 y procedentes de España), analizando un gran número de sustantivos en singular y en una amplitud de contextos considerable. Este autor observa que aproximadamente un 1 % de los casos localizados en los que el sujeto y el verbo se hallan en la misma oración la concordancia se produce en plural, aunque esta casi nunca ocurre si el sujeto y el verbo son adyacentes.

Estos trabajos también confirman que, cuando estos sujetos se retoman fuera de la cláusula en que se mencionan explícitamente, la concordancia en plural es más probable (Fält 1972) e incluso mayoritaria en algunos contextos y con algunos sustantivos —los de referencia indeterminada— (Fält 1972;

⁷ Resulta interesante notar que el estar inserto en una construcción partitiva parece favorecer la concordancia *ad sensum* incluso con estos colectivos en singular (frente a su uso en solitario), a pesar de que en esta circunstancia el sujeto tampoco cuente con ningún rasgo morfológico de pluralidad. Así parecen atestiguarlo los resultados de Areses (2019), actualmente basados en frecuencias absolutas: estos podrán confirmarse con el trabajo que continúa llevando a cabo esta investigadora para su tesis doctoral.

Sánchez Avendaño 2007). Asimismo corroboran que a mayor distancia (estructural), mayor probabilidad de concordancia plural, las observaciones de Fält (1972) de que las oraciones de relativo la desfavorecen más cuando el relativo es el sujeto, y que las oraciones coordinadas con *y* también la desfavorecen en contraste con otras coordinadas y yuxtapuestas⁸.

De estos estudios parece colegirse también cierto estatuto especial del sustantivo *gente*, que presenta frecuencias relativamente altas de la concordancia *ad sensum* en comparación con otros sustantivos en todos los estudios consultados (Fält 1972; Sánchez Avendaño 2007; Al-Amar y Ababneh 2017; Areses 2019). El origen del presente trabajo, de hecho, se halla en la observación informal realizada durante las campañas de encuesta para el COSER en Andalucía de que *gente* parecía concordar en plural con el verbo con inusitada frecuencia. Sin embargo, estos resultados exigen cautela, porque *gente* es también el sustantivo que mayor frecuencia absoluta presenta en estos corpus, por lo que contamos con menos datos sobre los otros sustantivos⁹.

Con todo, creo que merece la pena detenerse en la aguda observación de Sánchez Avendaño (2007) sobre la proximidad funcional de este sustantivo con los mecanismos de indefinición referencial del sujeto (es decir, las construcciones impersonales), puesto que “muchas veces el sustantivo ‘gente’ presenta la particularidad de servir como mecanismo de indiferenciación referencial, como un recurso para generalizar algo con respecto a un colectivo humano” (Sánchez Avendaño 2007: 217). Este autor nota la cercanía entre estas construcciones con *gente* y las construcciones impersonales expresadas con la morfología de 3PL en el verbo y sin sujeto expreso. Una observación similar se halla en Fält, para quien *gente* “puede expresar un matiz impersonal: *dice la gente...* no da más información sobre el sujeto que frases del tipo *dicen que...*” (1972: 119). De hecho, al analizar datos de uso real de la lengua, llega un momento en que es imposible determinar si nos hallamos

⁸ En el ámbito pronominal esto es más acusado. Como nota Areses (2019), fuera de la cláusula estos sujetos deben ser retomados necesariamente por pronombres plurales. Levin (2006), en su estudio del inglés, propone que mientras que con sujetos colectivos el número no marcado en la concordancia con el verbo es el singular, en la concordancia pronominal el número no marcado y, por tanto, el esperable, es el plural. Este comportamiento no es sorprendente si entendemos la concordancia como un mecanismo de cohesión textual, como hace Sánchez Avendaño (2007).

⁹ Levin (2006) no encuentra correlación entre concordancia *ad sensum* y *token frequency* en su estudio sobre la concordancia con sujetos colectivos del inglés, aunque sí encuentra importantes diferencias por lexema.

ante una construcción impersonal o ante un caso de persistencia referencial de un sujeto con *gente* mencionado anteriormente —cf. ejemplo (1), donde la relación con el sujeto explícito va difuminándose paulatinamente—. Como indica Sánchez Avendaño (2007), lo mismo puede decirse del sustantivo contable y singular *persona* y del indefinido *alguien*. Se puede añadir no solo el indefinido *alguno/a*, sino también el SN *todo el mundo*, que también indican la indefinición del sujeto¹⁰.

- (1) En fin, que había *mucha gente que tenía olivos*, mucha. Pues eso cuando llegaba el día de la Purísima, cuando llegaban los Santos, *lo acotaban*. *Echaban* un pregón que no se podía entrar en el local a por aceitunas ni el que las tenía podía ir, porque *decían* que no fuera ser que las cogiera de otro (Orellana de la Sierra, Badajoz, COSER-723).

Si bien la mayoría de los trabajos se centran en la concordancia de 3PL, esta no es la única posibilidad. Martínez García observa que la concordancia en 1PL/2PL es posible no solo con cualquier frase nominal en plural (porque estas en realidad no pertenecen a ninguna de las personas gramaticales, véase RAE/ASALE 2009: § 33.6j), sino también con algunas en singular, en “construcciones frecuentes aunque censuradas normativamente” (Martínez García 1999: 2758) como *En estas situaciones nadie nos atrevemos a protestar* o *La gente no sabéis comportaros*. Sobre las frases nominales en plural indica que “la interpretación semántica de las oraciones *fuerza* la correferencia entre la frase nominal y la persona verbal” (Martínez García 1999: 2764, el subrayado es mío), es decir, que si el hablante o el interlocutor forman parte de la referencia del sujeto, la concordancia en 1PL/2PL es obligatoria. Martínez García no se pronuncia acerca de si también lo es en los sustantivos singulares con interpretación plural, pero creo que ejemplos como (2) dan motivos para dudar de que así sea, ya que es evidente que el hablante se incluye en *la familia*, que concuerda en 3SG, pues sigue un uso de la 1PL. En otras palabras, asumo que la concordancia de plural, independientemente de la persona, está siempre en alternancia con la 3PL con estos sujetos morfológicamente singulares.

¹⁰ Nótese que en portugués *a gente* se ha gramaticalizado como pronombre de 1PL, ayudado por la posibilidad de que comparten sustantivo y pronombre de expresar indefinición: es interesante notar que, sin embargo, el pronombre ha mantenido mayoritariamente la concordancia verbal en 3SG (Lopes 2003), aunque es posible que haya diferencias regionales (véase Martins 2009).

- (2) el hacer el, el cerdo y luego, pues... eh se mataba, *se juntaba toda la familia, nos pasábamos* dos días prácticamente de fiesta (Anaya, Segovia, COSER-3701).

Sobre la concordancia *ad sensum* con indefinidos en singular la bibliografía es escasa. *Alguno/a* y *ninguno/a* participan sin restricciones en la construcción partitiva, donde pueden tomar un complemento plural. En estos casos, la concordancia mayoritaria es en singular (*Alguna de las estudiantes aprobó*), aunque si el complemento plural es un pronombre personal plural referido al hablante o al interlocutor (*nosotros/as, vosotros/as, ustedes*) la concordancia semántica es posible, pero “se considera más adecuada” la concordancia formal (RAE/ASALE 2009: § 33.9g). Si, por el contrario, ese mismo complemento —con el pronombre— está sobreentendido, la concordancia en 1PL/2PL es necesaria (RAE/ASALE 2009: § 33.9g).

Sobre *alguien* y *nadie*, que “tienden a rechazar los complementos partitivos en los registros más cuidados” (RAE/ASALE 2009: § 33.9j), solo se dice que la concordancia con un verbo en 1PL/2PL es poco frecuente, aunque es usada “a veces en el registro coloquial o en el habla espontánea” (RAE/ASALE 2009: § 33.9j), por lo que se recomienda evitarlas. Es decir, más allá de la mención de Sánchez Avendaño (2007) sobre *alguien*, no tenemos muchas indicaciones de que estos indefinidos desencadenen concordancia de 3PL en el verbo.

En este trabajo nos centramos en tres sustantivos colectivos (*familia, gente* y *mundo*, en la locución *todo el mundo*) y dos pronombres indefinidos en singular (*alguien, nadie*). Mientras que *familia* es un colectivo contable, *gente* es un nombre de masa (Bosque 1999; Areses 2019; Barrio de la Rosa 2020). Ambos se han considerado en la mayoría de los estudios de corpus mencionados arriba (Fält 1972; Sánchez Avendaño 2007; Al-Amar y Ababneh 2017; Areses 2019), aunque Barrio de la Rosa (2020) se centra fundamentalmente en *gente*. Fält muestra que *familia* es “uno de los colectivos que más fácilmente admiten la concordancia de sentido” (Fält 1972: 118)¹¹. *Mundo*, dentro de la locución *todo el mundo*, se comporta como un sustantivo colectivo (Bosque 1999: § 1.4.5.1). Solo Al-Amar y Ababneh (2017) consideran esta locución en su estudio.

¹¹ Dicha concordancia se produce especialmente en oraciones coordinadas (pero no cuando la conjunción es *y*), yuxtapuestas y subordinadas (aunque no en las relativas de sujeto) y en oraciones que se encuentran en distinto periodo sintáctico.

En lo concerniente a los indefinidos, la diferencia más importante entre los pares *alguno/a*, *ninguno/a* y *alguien*, *nadie* radica en la facilidad con que aparecen en la construcción partitiva: hemos descartado estos casos, así como aquellos en los que nuestros sustantivos aparecían en el interior de construcciones partitivas (*la mayoría de la gente*) y pseudopartitivas (*un montón de gente*) o en los que los propios sustantivos explicitaban sus miembros con un complemento con *de*¹².

3. Extensión geográfica de la concordancia *ad sensum* en el mundo hispanohablante

Para averiguar la extensión geográfica de la concordancia *ad sensum* aquí tratada, empleamos un corpus de más de diecisiete millones de tuits geolocalizados recogidos durante mayo y junio de 2016 para el proyecto ProVaIn. Twitter presenta abundantes muestras de habla de la inmediatez concepcional (Estrada Arráez y Benito Moreno 2016), por lo que debería ofrecer más ejemplos del fenómeno que nos ocupa que los corpus de referencia. Además, la geolocalización de los tuits es más detallada que la que obtenemos en dichos corpus. En comparación con un corpus oral como el PRESEEA, el corpus de tuits ofrece una extensión geográfica mayor¹³. Sin embargo, debe notarse que los datos de Twitter, a pesar de estar extendidos por todo el mundo hispanohablante, no lo están de forma uniforme. Su frecuencia está parcialmente determinada por la población de cada país, así como por la facilidad de acceso a internet y la penetración de la propia plataforma en cada país. En nuestro corpus, aproximadamente un tercio de todos los tuits proviene de Argentina, un quinto, de España, y un décimo, de México, mientras que el tercio restante se reparte (también de forma poco equitativa) entre los demás países hispanohablantes (véase la figura 1). Por último, si bien no podemos conocer las características sociales individuales de los usuarios, sí sabemos que la franja de la población que utiliza Twitter tampoco es una muestra representativa de la sociedad, sino que la población joven y urbana está sobrerrepresentada —especialmente en

¹² No se han excluido los complementos con *de* con otra función semántica: es decir, *una familia de veraneantes* no se ha considerado, mientras que *la familia de mis primos*, sí.

¹³ Además, los tuits suelen ofrecer un contexto lingüístico completo, lo que permite evaluar mejor el contexto que los breves resultados del PRESEEA, frecuentemente interrumpidos.

el caso de los datos geolocalizados (Hecht y Stephens 2014; Pavalanathan y Eisenstein 2015)—.

Conviene explicar qué ejemplos se buscaron y cómo se trataron. En el texto del tuit, transformado enteramente a minúsculas, se buscaron, por medio de expresiones regulares, las secuencias de palabras *la gente*, *la familia*, *todo el mundo*, *alguien* y *nadie*¹⁴. Decidimos excluir los indefinidos *alguno/a* y *ninguno/a* porque en su función de pronombres son poco frecuentes, por lo que introducirían mucho ruido en un corpus de tamaño considerable. Nótese que solo hemos buscado los casos en que los sustantivos están precedidos por el artículo definido (con el único objetivo de simplificar el proceso). Posteriormente se excluyeron, también con expresiones regulares, todos los tuits en que dichas expresiones estaban precedidas por una preposición. No se hicieron más esfuerzos para encontrar posibles faltas de ortografía o erratas que considerar variantes con o sin tilde, pues el trabajo creativo necesario para imaginar todas las posibles variantes es probablemente mayor que el beneficio que reportaría en términos de nuevos casos.

Debido a la gran cantidad de ejemplos encontrados (tabla 1), y teniendo en cuenta que la concordancia semántica tiene una frecuencia muy baja, optamos por lematizar y etiquetar los tuits de forma automática para simplificar el filtrado de los ejemplos relevantes. Para lematizar y etiquetar el corpus, se empleó el modelo *ancora* de la librería *udpipe* de R (Wijffels 2020). Al comprobar que el modelo producía frecuentes errores al considerar todo el contexto (probablemente por las características de la lengua de los tuits, con frecuentes erratas o lenguaje innovador), decidimos aplicarlo a cada palabra separada de su contexto. Para ello, primero delimitamos (con expresiones regulares) un contexto alrededor de la primera ocurrencia de cada una de las expresiones buscadas: las tres palabras precedentes y las cuatro posteriores (en caso de que las hubiera). Ese contexto garantiza poder localizar todos los casos de verbos inmediatamente antepuestos al sujeto (sea una forma simple, un tiempo compuesto o una perífrasis con o sin nexos entre ambos verbos),

¹⁴ Estas secuencias estaban precedidas y seguidas por el ancla `\b`, para evitar la recuperación de ejemplos como *la familiaridad* o *mala gente*. Los datos empleados en este artículo, así como el código empleado tanto para la recuperación de los ejemplos como para su análisis, pueden encontrarse en la siguiente dirección: https://github.com/Carlotadbm/Concordancia_ad_sensum_2020.

y los inmediatamente pospuestos al verbo (incluso si están precedidos por clíticos o la negación)¹⁵.

Cada palabra de este contexto precedente y posterior fue lematizada y etiquetada de forma automática y, después, se seleccionaron aquellos ejemplos en que el modelo había detectado una forma verbal de plural. Estos se revisaron después de forma manual, para comprobar que se trataba verdaderamente del contexto buscado. En la tabla 1 se muestra la cantidad de ejemplos de concordancia de sentido obtenidos siguiendo este procedimiento¹⁶.

Sujeto	Ocurrencias	Plurales detectados automáticamente	Concordancia detectada manualmente
<i>alguien</i>	96 840	507	40
<i>nadie</i>	92 853	661	92
<i>la gente</i>	56 860	477	211
<i>la familia</i>	4517	129	24
<i>todo el mundo</i>	9790	110	51
Totales	260 860	1884	418

Tabla 1. Ocurrencias encontradas en cada paso del proceso

En la tabla 2 pueden observarse las ocurrencias por persona, posición y sujeto. Como ocurría en todos los trabajos de corpus mencionados, el más frecuente es *la gente*. Salvo en el caso de *la familia*, la concordancia en 3PL es más frecuente que en cualquier otra persona. Por otra parte, las diferencias entre *nadie* y *alguien* a la hora de incluir al hablante (o al interlocutor, aunque estos casos son siempre menos frecuentes, pues solo pueden documentarse inequívocamente en los tuits producidos por españoles) son evidentes: *nadie* lo hace con mucha más facilidad que *alguien*. Nótese que, aunque hay una diferencia evidente en la pluralidad de estos indefinidos (*nadie* siempre se interpreta como una pluralidad, pero *alguien* no necesariamente), ambos admiten

¹⁵ Un contexto posterior de cuatro palabras excluye la posibilidad de encontrar verbos precedidos por la negación y una combinación de tres clíticos, pero estas son extremadamente infrecuentes en español.

¹⁶ La revisión manual asegura que todos los ejemplos considerados sean efectivamente ejemplos de concordancia de sentido, pero no permite recuperar aquellos ejemplos de dicha concordancia que el lematizador no reconociera adecuadamente.

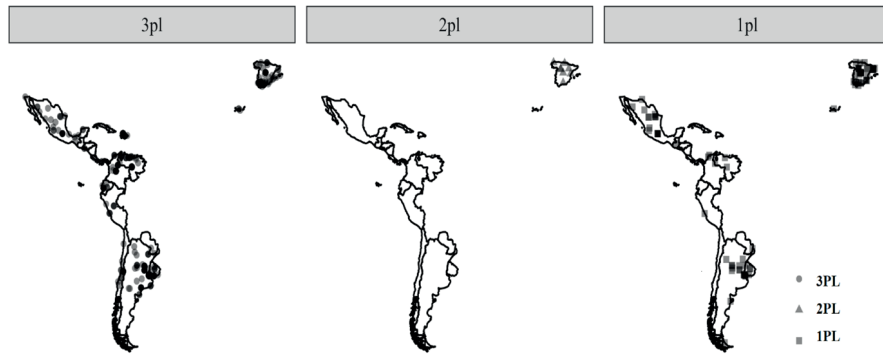
el verbo en 3PL sin que se observe una diferencia sustancial¹⁷. Sin un estudio de la frecuencia de la posición del verbo con cada uno de estos sujetos en 3PL (la más frecuente), es difícil asegurar si la posición del verbo ejerce un efecto en la concordancia, especialmente porque, cuando se observan diferencias, estas no van siempre en la misma dirección.

		<i>alguien</i>	<i>nadie</i>	<i>la gente</i>	<i>la familia</i>	<i>todo el mundo</i>
1PL	ant	0 (0 %)	4 (4,3 %)	28 (13,3 %)	7 (29,2 %)	6 (11,8 %)
	pos	0 (0 %)	18 (19,6 %)	12 (5,7 %)	7 (29,2 %)	10 (19,6 %)
2PL	ant	0 (0 %)	2 (2,2 %)	1 (0,5 %)	0 (0 %)	2 (3,9 %)
	pos	0 (0 %)	1 (1,1 %)	1 (0,5 %)	0 (0 %)	1 (2 %)
3PL	ant	9 (22,75 %)	24 (26,1 %)	114 (54 %)	8 (33,3 %)	14 (27,5 %)
	pos	31 (77,5 %)	43 (46,7 %)	55 (26,1 %)	2 (8,3 %)	18 (35,3 %)
Totales		40 (100 %)	92 (100 %)	221 (100 %)	24 (100 %)	51 (100 %)

Tabla 2. Concordancia *ad sensum* en Twitter

En el mapa 1 se observa que estos ejemplos de concordancia *ad sensum* se encuentran repartidos por todo el territorio hispanohablante, como se espera de un universal vernáculo. La comparación entre la proporción de tuits totales y la proporción de casos de concordancia *ad sensum* por país confirma esta impresión (figura 1). Como puede observarse, ambas proporciones son muy parecidas (con diferencias que rondan el 1-2 %), lo que sugiere que no hay diferencias regionales importantes. España es el país en que esta diferencia es más llamativa (5 %), lo que podría indicar que la concordancia de sentido es más frecuente, pero las cifras sugieren un efecto bastante moderado.

¹⁷ Aunque no hemos analizado los más de 180 000 ejemplos encontrados de *nadie* y *alguien* en el corpus, ambos aparecen en proporciones muy similares, lo que sugiere que sí podemos comparar sus frecuencias absolutas, aunque otra vez debamos ejercer la cautela, ya que podrían aparecer en contextos distintos.



Mapa 1. Concordancia plural en el verbo en Twitter

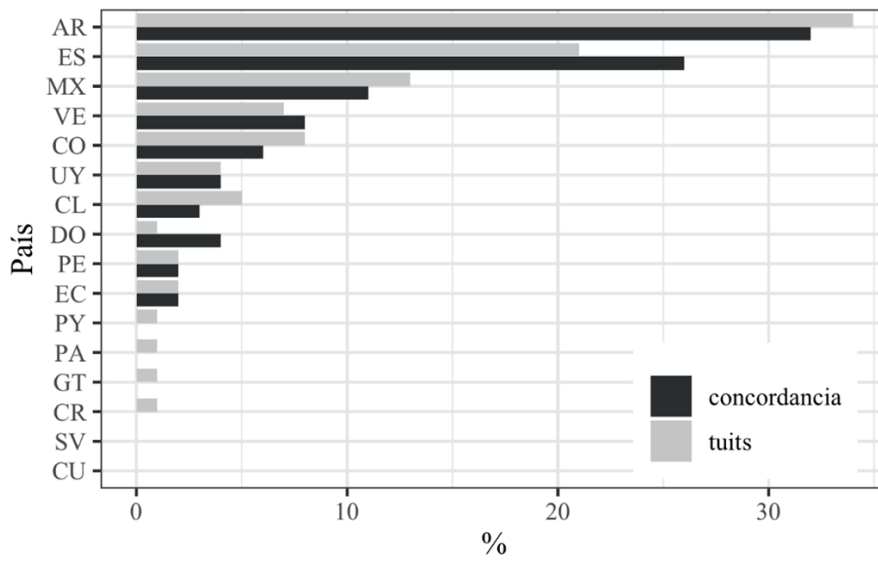


Figura 1. Número de tuits y casos de concordancia *ad sensum* por países

4. La concordancia *ad sensum* en el COSER

Para estudiar con más detalle algunos de los factores lingüísticos y geográficos que afectan a la concordancia *ad sensum*, acudimos al *Corpus Oral y Sonoro del Español Rural* (COSER). Si bien este corpus se limita al español de España, presenta indudables ventajas para este estudio, puesto que contiene habla oral, de distintas regiones del país, en una muestra de hablantes homogénea (hablantes rurales poco móviles, de edad avanzada y bajo nivel de escolarización, encuestados en su lugar de procedencia). Además, por su temática (la vida en el mundo rural) se presta a la presencia de los sustantivos y pronombres de determinación poco definida aquí estudiados¹⁸.

Las búsquedas realizadas se dirigieron a encontrar todas las ocurrencias de *gente*, *familia*, *todo el mundo*, *alguien* y *alguno/a* en los informantes del corpus¹⁹. Los 5855 resultados se analizaron manualmente para descartar aquellos casos en que el sustantivo o pronombre no es sujeto y anotar los restantes según una serie de parámetros lingüísticos. Solo hemos considerado los casos en que el sustantivo o pronombre es sujeto del verbo de la oración en que se encuentra, y sujeto del verbo de una oración subordinada a aquella en la que se encuentra o del verbo de la principal si se encuentra explícito en una

¹⁸ La casualidad ha querido que este trabajo se solape así con el de Barrio de la Rosa (2020), que fue publicado unos días después de que este fuera entregado (agradezco a los editores del volumen haberme permitido hacer cambios para referirme a dicho artículo). Barrio de la Rosa (2020) también emplea los datos del COSER para estudiar la concordancia del sustantivo *gente* (y también lo relaciona con el concepto de los universales vernáculos). Afortunadamente, el objeto de análisis del presente trabajo es más amplio y las técnicas estadísticas empleadas permiten hallazgos nuevos, por lo que el solapamiento no significa repetición.

¹⁹ Las formas se buscaron en la consulta simple, por búsqueda exacta, y únicamente en el habla de los informantes (Barrio de la Rosa 2020 incluye también la de los encuestadores y observa diferencias entre el habla de estos y el de los informantes, lo cual sugiere diferencias de tipo diastrático o diafásico). Para el indefinido buscamos la forma masculina *alguno* y la femenina *alguna*. Puesto que *todo* presenta formas reducidas en el habla oral, que pueden fusionarse con el artículo que sigue, buscamos las formas gráficas *todo el mundo*, *to el mundo*, *to'l mundo*, *tol mundo* y *todo'l mundo*. Debido a un error, por el cual buscamos en primer lugar *la gente* y *la familia*, acompañados del artículo, completamos esas búsquedas con las siguientes por expresiones regulares: “[^IL][^a] gente\b”, “[^IL][^a] familia\b”, “[IL][^a] familia” y “[IL][^a] gente”, que permitieron encontrar todos los resultados de *gente* y *familia* no precedidos por *la*. Aunque se buscaron también *nadie* y *ninguno/a*, fueron finalmente excluidos, pues, al tratarse de un corpus oral y teniendo en cuenta que frecuentemente la 3PL solo se diferencia de la del singular en la *-n* final, cuando el verbo está antepuesto al pronombre resulta prácticamente imposible distinguirlos.

subordinada. Hemos excluido los casos en que el sujeto tenía un complemento partitivo (como *alguno de ellos*) o un complemento que especificara la composición del colectivo (*una familia de veterinarios*).

Respecto de los verbos cuyo sujeto se halla en la misma oración, hemos anotado los siguientes parámetros²⁰:

1. Persona gramatical del verbo.
2. Posición del verbo respecto del sujeto (antepuesto o pospuesto).
3. Adyacencia entre sujeto y verbo. Para el verbo pospuesto, consideramos adyacentes también aquellos separados del sujeto por clíticos de objeto o los adverbios de polaridad *no*, *ya* o *sí* (cuando estos últimos no se encuentran entre pausas), pues esta es su posición natural en la oración y, además, son palabras muy breves. Igualmente, no consideramos que la expansión del sujeto (por un adjetivo o un SP) impida la contigüidad, pues la entendemos en términos estructurales. En el caso del verbo antepuesto, no consideramos que su aparición en una perífrasis verbal o un tiempo compuesto impida la adyacencia, aunque la forma concordante no sea estrictamente contigua al sujeto.
4. En caso de que, estrictamente a continuación del verbo (si está pospuesto) o del sujeto (si el verbo está antepuesto) haya una oración yuxtapuesta o coordinada que comience con un verbo cuyo sujeto sea correferencial con el que nos interesa, también hemos anotado la persona gramatical de este:
 - (3) a. *Se sentaba la gente y se lo pasaba* mu bien (Narros del Puerto, Ávila, COSER-614).
 - b. Todos, todos igual, *todo el mundo trabajaba, se iba a arar* con las vacas (Almázcara, Congosto, León, COSER-2637).

Para los verbos que se hallan en una oración diferente, pero en relación de dependencia (por ser subordinada o principal), hemos anotado únicamente la persona gramatical del verbo. La mayoría de estos casos son oraciones de relativo. De esta manera recogimos 2478 sujetos y 2627 verbos concordantes con ellos, repartidos como se indica en las tablas siguientes. Como se puede observar, la gran mayoría de los casos recogidos presenta el sujeto y el verbo en la misma oración (tabla 3). Además, *gente* es el sujeto más frecuente (tabla 4).

²⁰ Barrio de la Rosa (2020) también anota la presencia o ausencia de determinantes en el SN del sujeto, pero no observa diferencias significativas entre una u otra posibilidad.

Verbo	Ocurrencias
Misma oración	2047 (77,9 %)
Oración coordinada o yuxtapuesta	54 (2,1 %)
Relación de dependencia	526 (20 %)

Tabla 3. Distribución de contextos verbales

Sujeto	Ocurrencias
<i>alguien</i>	75 (3 %)
<i>alguno</i>	212 (8,6 %)
<i>familia</i>	156 (6,3 %)
<i>gente</i>	1708 (68,9 %)
<i>todo el mundo</i>	327 (13,2 %)

Tabla 4. Distribución de los sujetos

La tabla 5 recoge la proporción de las distintas formas de concordancia verbal documentadas. Lo primero que se observa es que, como es de esperar, la concordancia formal (en 3SG) es mayoritaria (78,2 % de los casos). Pero el mismo dato muestra que la concordancia de sentido no es anecdótica. La más común es la 3PL (17,1 %), seguida de la 1PL, que ya solo aparece en un 4,2 % de nuestros ejemplos. La concordancia de 2PL es muy infrecuente (0,3 %), algo previsible en el contexto de las entrevistas semidirigidas, que no favorecen la referencia a los interlocutores. Por último, es destacable la existencia de 5 ejemplos (0,2 %) con el verbo en 2SG. Esta concordancia ni es formal ni es de sentido *stricto sensu*, pues tiene que ver con la lectura impersonal de esta 2SG gramatical. Como se notó en el apartado 2, este tipo de sustantivos e indefinidos producen lecturas similares, por lo que parece que nos hallamos ante un tipo de construcción mixta. El sujeto expreso aporta información sobre el tamaño del grupo indefinido que se interpreta como agente en la impersonal (4)²¹. En lo que sigue, se excluyen estos ejemplos por no ser comparables con los demás.

²¹ Si bien puede tratarse de una cuestión anecdótica, no queremos dejar de mencionar que todos estos ejemplos se localizan en el castellano oriental (Huesca, Zaragoza, Teruel y Guadaluajara). Es precisamente esta zona donde Benito Moreno (2015) encuentra también construcciones impersonales mixtas, con *se* y un sujeto flexivo distinto de la tercera persona, aunque

- (4) a. Pos eso, como daban chocolate, claro a lo mejor *te juntabas mucha gente*, mucha (Yebra, Guadalajara, COSER-1921).
- b. Y *según cuánta gente ibas*, pues entonces o cuántas personas, entonces se echaba más pan y la fiambra más grande y llena de, de todo (Fuentes Claras, Teruel, COSER-4117).
- c. Y sabes cuándo pa la Se-, la Semana Santa había una, si-, sie-, siete domingos de san José que decían que no había baile ya, ya, y *ya en misa te juntabas alguna*, al salir de misa (San Mateo de Gállego, Zaragoza, COSER-4719).

Persona	Ocurrencias
3SG	2054 (78,2 %)
3PL	449 (17,1 %)
1PL	111 (4,2 %)
2PL	8 (0,3 %)
2SG	5 (0,2 %)

Tabla 5. Frecuencia de las distintas concordancias verbales

Debido a las asimetrías en las frecuencias absolutas de los contextos estudiados, a continuación, comparamos el comportamiento de los cinco sujetos solo en los tres contextos más frecuentes: con el verbo adyacente dentro de la misma oración, tanto antepuesto como pospuesto, o con relación de dependencia. De la figura 2 se desprenden varias observaciones interesantes. Por un lado, la concordancia *ad sensum* es sustancialmente menos frecuente cuando el sujeto es un indefinido que cuando es un sustantivo. Las diferencias entre *alguien* y *alguno/a* no son muy notables, a pesar de que solo el segundo admite con plena facilidad un complemento partitivo que admitiría un pronombre plural (contextos excluidos de esta investigación). Debido al escaso número de ocurrencias en 2PL es difícil saber si el hecho de que todas se documenten con *alguno/a* y ninguna con *alguien* es indicativo de una diferencia real en su

también en pequeña proporción. Solo un estudio más minucioso y voluminoso podrá determinar si efectivamente estas construcciones son más frecuentes en esta zona. Por otra parte, estos ejemplos recuerdan a las construcciones del portugués de Madeira que combinan tanto el sujeto expreso *a gente* con la marca de impersonalidad *se* (Martins 2009; Henriques 2019), de productividad mucho mayor.

comportamiento o es producto del azar (en 1PL solo encontramos un ejemplo con cada uno de los indefinidos)²².

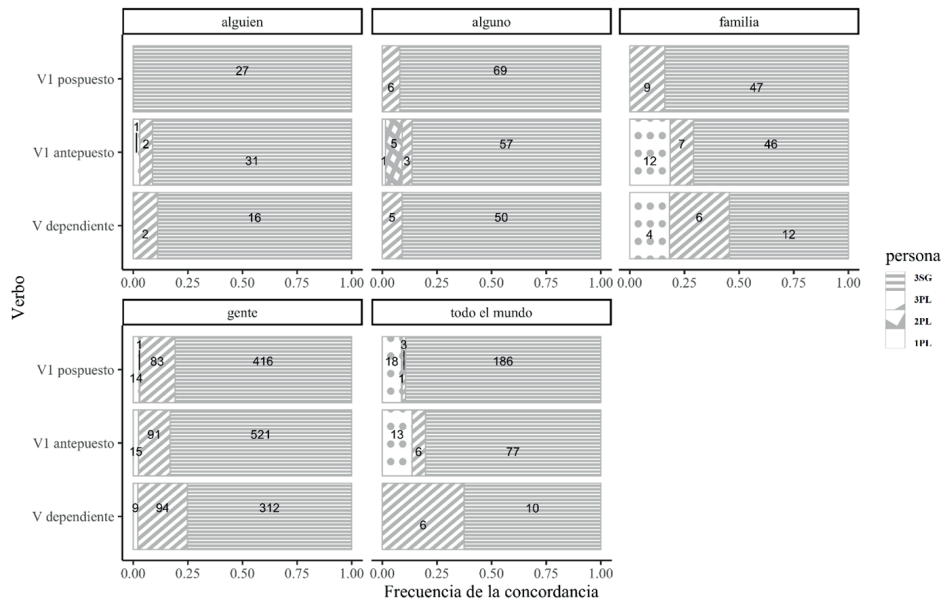


Figura 2. Concordancia verbal por sujeto en los contextos más frecuentes

Por otro lado, la posición del verbo no parece tener un efecto importante en la concordancia (frente a lo que establece Martínez García 1999), pues las diferencias son pequeñas y la dirección de estas no es constante. Los bajos números encontrados en algunos casos dificultan evaluar el papel de la distancia estructural, aunque sí parece que, cuando sujeto y verbo aparecen en oraciones distintas (aunque con relación de dependencia), la concordancia de sentido es más frecuente. Lo mismo ocurre con las diferencias entre sustantivos, ya que las frecuencias absolutas son demasiado bajas como para certificar que la 1PL sea más frecuente con *familia* y *todo el mundo* que con *gente*. Si así fuera, la razón no es evidente, pues *todo el mundo* y *gente* son semánticamente más similares entre sí, frente a *familia*.

²² En total, de los ocho casos de concordancia en 2PL, seis aparecen con *alguno/a*; uno con *gente* y otro con *todo el mundo*.

A continuación, nos centramos únicamente en *gente*, que, por ser el sustantivo más frecuente, permite observar el contraste entre todos los contextos estudiados. Excluimos los dos casos de concordancia en 2SG/2PL, pues no aportan mucha información. La figura 3 muestra que la concordancia de sentido con la 1PL es residual en comparación con la 3PL.

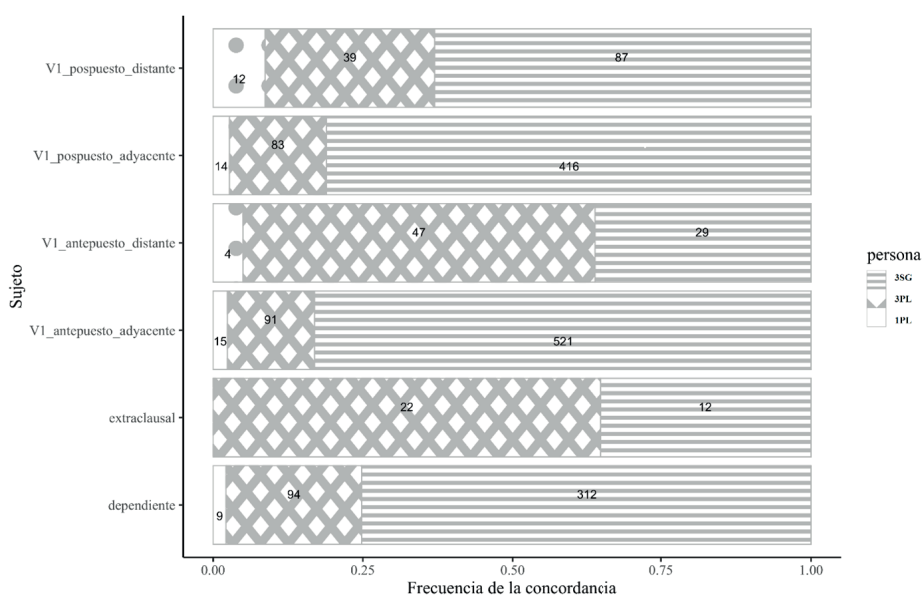


Figura 3. Concordancia de *gente* según el contexto

Además, estos datos confirman que la concordancia *ad sensum* aumenta cuando la distancia estructural y lineal crece. El efecto de la distancia lineal se observa en el hecho de que, cuando el verbo se encuentra en la misma oración que el sujeto, el plural es siempre más frecuente si no hay contigüidad entre ambos. El papel de la distancia estructural se ve en la elevada frecuencia del plural (llega a ser mayoritaria) en oraciones coordinadas o yuxtapuestas. Por último, sí se observa un efecto de la posición del verbo respecto del sujeto, pero solo cuando no hay adyacencia: la concordancia *ad sensum* es sustancialmente más frecuente cuando el verbo está antepuesto (en contra de los

señalado por Martínez García 1999)²³. Esto sugiere que en estos casos el sujeto puede no haberse decidido todavía cuando se pronuncia el verbo.

Estas observaciones se ven confirmadas en el modelo lineal generalizado mixto presentado en la tabla 6, que toma como variable dependiente la concordancia en plural (aunando todas las personas) o singular, en función de cuatro variables independientes (adyacencia verbal, tipo de oración, posición del verbo e interacción entre la posición verbal y su adyacencia respecto al verbo)²⁴. La localidad COSER se ha incluido como efecto aleatorio, para controlar la variabilidad individual de cada entrevista. Los coeficientes del modelo se dan en forma de razón de momios o probabilidades, más intuitivos que la escala logarítmica habitual. Así, los valores mayores de 1 indican mayor probabilidad de la concordancia plural en el contexto incluido en la tabla frente al valor de referencia. Baste un ejemplo de esa interpretación: la probabilidad de la concordancia plural en oraciones coordinadas o yuxtapuestas frente a otras es de 10 a 1. Por cautela, evitamos el uso de valores *p* y, en cambio, ofrecemos los intervalos de confianza, que permiten hacerse una mejor idea de la incertidumbre sobre el tamaño del efecto (Nuzzo 2014).

	Coeficientes	Error estándar	CI 95 % (menor)	CI 95 % (mayor)
Intersección	0,167	0,0276	0,121	0,231
Adyacencia verbal: no	2,75	0,649	1,74	4,37
Oración: relación de dependencia	1,04	0,326	0,56	1,92
Oración: coordinada o yuxtapuesta	10,0	4,35	4,27	23,5
Posición verbal: posposición	0,856	0,148	0,611	1,20
Interacción: V antepuesto no adyacente	3,32	1,20	1,64	6,75
COSERid (efecto aleatorio)	1,07	NA	NA	NA

Tabla 6. Concordancia verbal con *gente* en el COSER: modelo lineal generalizado mixto

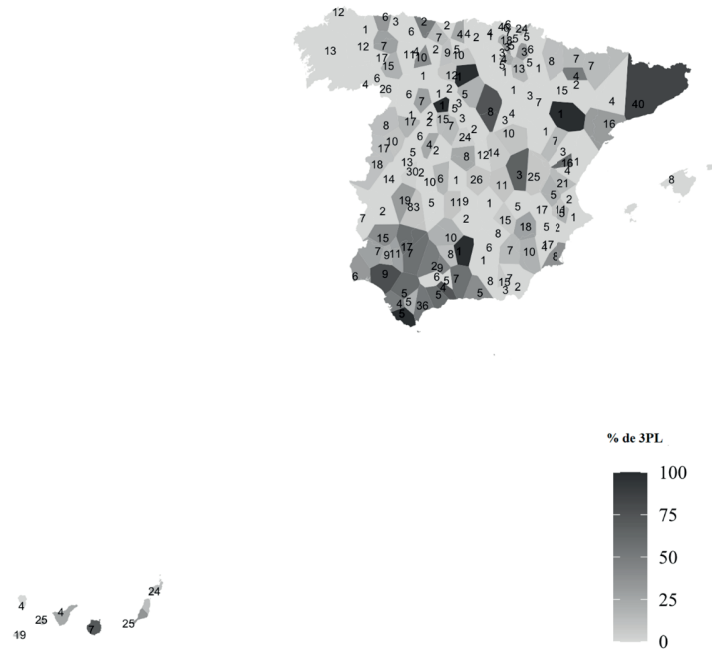
²³ Curiosamente, Barrio de la Rosa (2020) tampoco observa un efecto de la posición del sujeto, pero esto seguramente se deba a que no mide la distancia entre sujeto y verbo —aunque sí la menciona como un factor relevante—: el efecto lo detectamos como una interacción entre ambos factores.

²⁴ El modelo se calculó con la librería *lme4* (Bates *et al.* 2015) de R.

Queda un aspecto por tratar de los datos del COSER. Como se dijo más arriba, el origen de esta investigación está en la observación subjetiva de que *gente* desencadena la concordancia de sentido con más frecuencia de lo habitual en Andalucía. El mapa 2 representa la frecuencia de la concordancia de 3PL con *gente* cuando el verbo es adyacente (antepuesto o pospuesto) o cuando está en una oración con relación de dependencia respecto de la del sujeto, es decir, en los casos en los que esta concordancia es globalmente menos frecuente. Los tonos representan la frecuencia (en porcentaje) de la concordancia semántica frente a la formal, mientras que los números representan el número de casos (plural y singular) totales recogidos en cada lugar, para que estas frecuencias puedan ponerse en perspectiva. Por un lado, el mapa muestra que la concordancia *ad sensum* en estos contextos se encuentra, con bajas frecuencias, repartida por todo el país. Por otro lado, se observa una mayor concentración de esta en dos lugares: Andalucía occidental y Cataluña. Por lo que se refiere a Cataluña, la cautela es aconsejable, ya que solo disponemos de tres entrevistas en esta región. La entrevista realizada en la población barcelonesa de La Serra (La Torre de Claramunt) es responsable de la gran área roja, que incluye Gerona únicamente por el método matemático empleado para la agregación espacial (diagramas de Voronoi), pero es indudable que este hablante presenta una concordancia muy regular de *gente* en plural (34/40 casos, 85 %) ²⁵. En Andalucía occidental las frecuencias del plural no son tan elevadas, pero la multitud de puntos en los que se documenta sugiere que este fenómeno, propio de la oralidad concepcional en otros lugares, conforma un rasgo diatópico de esta región ²⁶.

²⁵ El estatus de la concordancia de sentido con nombres colectivos en catalán es similar al que tiene en español (lengua en que fue realizada la entrevista): el singular es más frecuente y el plural se ve favorecido por la distancia con el sujeto (Bel 2008; Martí i Girbau 2008; IEC 2016). El IEC (2016) nota una especial frecuencia de la concordancia plural con *gent* frente a otros sustantivos.

²⁶ Barrio de la Rosa (2020), de los mismos datos, interpreta lo contrario, es decir, que no existen diferencias dialectales. Pero si los datos son los mismos, no lo es la información visualizada, lo que explica estas diferencias. Barrio de la Rosa muestra los enclaves en que se documenta la concordancia de sentido y, aunque nota que hay más presencia de enclaves meridionales, lo interpreta como probable fruto de la casualidad. Al añadir la frecuencia relativa de dicha concordancia se observa no solo un mayor número de enclaves en Andalucía occidental, sino también una productividad más alta de la concordancia *ad sensum*. Para dilucidar completamente la cuestión haría falta una base empírica mucho más amplia de la que nos ofrece el COSER por ahora, que permitiera aplicar técnicas de estadística inferencial.



Mapa 2. Frecuencia de la concordancia de 3PL con *gente* en los tres contextos más frecuentes (COSER)

5. Conclusiones

Los datos aquí considerados muestran, por un lado, que la concordancia verbal *ad sensum* con sujetos colectivos o indefinidos singulares se documenta en todas las variedades del español, como es esperable de un rasgo universal (que en este caso es propio de la oralidad concepcional). Esta universalidad se refleja también en la existencia de un efecto de la distancia (lineal y estructural) en la frecuencia de dicha concordancia, efecto detectado ya numerosas veces tanto para el español como para otras lenguas (Corbett 2006): los factores subyacentes son universales. Por último, los datos del COSER sugieren que la concordancia plural, al menos con el sustantivo *gente*, es particularmente frecuente en Andalucía occidental: este es, claro, un rasgo situado en el nivel histórico, puesto que, como se vio en el apartado 1, los rasgos universales se

materializan en una lengua histórica concreta, donde podrán seguir caminos particulares. Este hecho nos parece relevante en la discusión sobre los universales vernáculos, pues subraya la necesidad de separar ambos planos (el universal y el histórico) y de ser cautelosos a la hora de deducir la universalidad de la amplia distribución geográfica: los rasgos universales se materializan necesariamente en rasgos históricos y estos son susceptibles de difusión.

Bibliografía

- AL-AMAR, Abdullah, y ABABNEH, Mohammad (2017): “La concordancia verbal desde dos perspectivas: normativa y pragmática, el sujeto colectivo en el discurso oral como ejemplo”, en *Jordan Journal of Modern Languages and Literature*, 9/2, pp. 179-195.
- ARESES, Irene (2019): “Concordancia verbal plural de sujetos partitivos que presentan en la coda nombres que designan colectividades”. Trabajo de fin de grado: Alcalá de Henares: Universidad de Alcalá.
- BARRIO DE LA ROSA, Florencio del (2020): “La transparencia de la pluralidad. Notas sobre la concordancia semántica de *gente* en el español rural europeo”, en *Ars & Humanitas*, 14/2, pp. 227-240.
- BATES, Douglas, MAECHLER, Martin, BOLKER, Ben, y WALKER, Steve (2015): “Fitting Linear Mixed-Effects Models Using lme4”, en *Journal of Statistical Software*, 67/1, pp. 1-48.
- BEL, Aurora (2008): “Les funcions sintàctiques”, en Joan Solà, Maria-Rosa Lloret, Joan Mascaró y Manuel Pérez Saldanya (eds.), *Gramàtica del Català Contemporani*, vol. 2. Barcelona: Empúries, pp. 1075-1147.
- BELLO, Andrés (1988 [1847-1860]): *Gramática de la lengua castellana destinada al uso de los americanos. Con las notas de Rufino José Cuervo*. Madrid: Arco/Libros.
- BENITO MORENO, Carlota de (2015): “Pero se escondíamos como las ratas: Syncretism in the Reflexive Paradigm in Spanish and Catalan”, en *Isogloss*, 1, pp. 95-127.
- (2020): “Reflexiones sobre la ‘lengua vulgar dialectal’ y el vulgarismo”, en Inés Fernández-Ordóñez (ed.), *El legado de Ramón Menéndez Pidal (1869-1968) a principios del siglo XXI*. Madrid: CSIC (Anejos de la *Revista de Filología Española*, 108), pp. 419-456.
- BOSQUE, Ignacio (1999): “El nombre común”, en Ignacio Bosque y Violeta Demonte (dirs.), *Gramática descriptiva de la lengua española*, vol. 1. Madrid: Espasa Calpe, pp. 3-75.

- CHAMBERS, Jack K. (2000): “Universal Sources of the Vernacular”, en *Sociolingüística*, 14, pp. 11-15.
- (2004): “Dynamic Typology and Vernacular Universals”, en Bernd Kortmann (ed.), *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*. Berlin/New York: Mouton de Gruyter, pp. 127-146.
- CORBETT, Greville G. (2006): *Agreement*. Cambridge: Cambridge University Press.
- COSERIU, Eugenio (1958): *Sincronía, diacronía e historia*. Montevideo: Universidad de la República.
- (1973): *Sincronía, diacronía e historia*. Madrid: Gredos.
- ESTRADA ARRÁEZ, Ana, y BENITO MORENO, Carlota de (coords.) (2016): *Variación lingüística e internet: cayendo en la red*, en *Revista Internacional de Lingüística Iberoamericana*, 28.
- FÄLT, Gunnar (1972): *Tres problemas de concordancia verbal en el español moderno*. Uppsala: Almqvist and Wiksell.
- FERNÁNDEZ-ORDÓÑEZ, Inés (dir.) (2005-2021): *Corpus Oral y Sonoro del Español Rural (COSER)*. Madrid: Universidad Autónoma de Madrid. <<http://www.corpus-rural.es>> (28-10-2020).
- GALLEGO, Ángel J. (dir.) (2015-2017): *Proyectando la variación lingüística de Internet (ProValn)*. Barcelona: Universidad Autónoma de Barcelona/Fundación BBVA.
- GARCÍA DE DIEGO, Vicente (1978): *Manual de dialectología española*. Madrid: Centro Iberoamericano de Cooperación.
- GILI GAYA, Samuel (1961): *Curso superior de sintaxis española*. Barcelona: Bibliograf.
- HECHT, Brent, y STEPHENS, Monica (2014): “A Tale of Cities: Urban Biases in Volunteered Geographic Information”, en *Proceedings of the International Conference on Web and Social Media*. Palo Alto: The AAAI Press, pp. 197-205.
- HENRIQUES, Yoselin (2019): *Particularidades morfosintáticas em variedades rurais do português falado na Ilha da Madeira*. Zürich: Universität Zürich.
- INSTITUT D’ESTUDIS CATALANS (2016): *Gramàtica de la llengua catalana*. Barcelona: Institut d’Estudis Catalans.
- KABATEK, Johannes (2004): “A diferenciación diafásica do galego actual”, en Rosario Álvarez Blanco, Francisco Fernández Rei y Antón Santamarina (eds.), *A lingua galega: historia e actualidade*. Santiago de Compostela: Instituto da Lingua Galega, pp. 379-388.
- KOCH, Peter, y OESTERREICHER, Wulf (2007 [1990]): *Lengua hablada en la Romania: español, francés, italiano*. Madrid: Gredos.
- LEVIN, Magnus (2001): *Agreement with Collective Nouns in English*. Lund: Lund University.
- (2006): “Collective Nouns and Language Change”, en *English Language and Linguistics*, 10/2, pp. 321-343.

- LOPES, Célia Regina dos Santos (2003): *A inserção de a gente no quadro pronominal do português*. Madrid/Frankfurt: Iberoamericana/Vervuert.
- LÓPEZ SERENA, Araceli (2002): “Reseña de Peter Koch y Wulf Oesterreicher (1990): *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*”, en *Lexis*, XXVI/I, pp. 255-271.
- MARTÍ I GIRBAU, Núria (2008): “El SN: els noms”, en Joan Solà, Maria-Rosa Lloret, Joan Mascaró y Manuel Pérez Saldanya (eds.), *Gramàtica del Català Contemporani*, vol. 2. Barcelona: Empúries, pp. 1281-1335.
- MARTÍNEZ GARCÍA, José Antonio (1999): “La concordancia”, en Ignacio Bosque y Violeta Demonte (dirs.), *Gramática descriptiva de la lengua española*, vol. 2. Madrid: Espasa Calpe, pp. 2695-2786.
- MARTINS, Ana M.^a (2009): “Subject Doubling in European Portuguese Dialects: The Role of Impersonal *Se*”, en Enoch Oladé Aboh, Elisabeth van der Linden, Josep Quer, Petra Sleeman, *Romance Languages and Linguistic Theory: Selected Papers from ‘Going Romance’ Amsterdam 2007*. Amsterdam/Philadelphia: John Benjamins, pp. 179-200.
- NUZZO, Regina (2014): “Statistical Errors”, en *Nature*, 506/7487, pp. 150-152.
- OGGIANI, Carolina (2013): “Reseña de María Ángeles Soler Arechalde (2012): *La concordancia de número en español. Cuatro casos de alternancia*”, en *Linguística*, 29/2, pp. 335-338.
- ORTEGA, Gonzal, y MORERA, Marcial (1982): “La concordancia numérica de los colectivos: un caso de silepsis”, en *Archivum: Revista de la Facultad de Filosofía y Letras*, XXXI-XXXII, pp. 645-656.
- PARODI, Claudia, y SANTA ANA, Otto (1997): “Tipología de comunidades de habla: del español rural al estándar”, en *Nueva Revista de Filología Hispánica*, XLV/2, pp. 305-320.
- PAVALANATHAN, Umashanthi, y EISENSTEIN, Jacob (2015): “Confounds and Consequences in Geotagged Twitter Data”, en Lluís Màrquez, Chris Callison-Burch y Jian Su (eds.), *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisboa: Association for Computational Linguistics, pp. 2138-2148.
- PRESEEA = *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. <<http://preseea.linguas.net>> (10-05-2020).
- REAL ACADEMIA ESPAÑOLA Y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA: *Corpus del español del siglo XXI (CORPES)*, [versión 0.9]. Madrid: RAE. <<http://www.rae.es>> (10-05-2020).
- (2009): *Nueva gramática de la lengua española*. Madrid: Espasa.
- REY QUESADA, Santiago del (2021): “Lo marcado y lo no marcado en la cadena de variedades”, en Klaus Grübl, Teresa Gruber, Katharina Jacob y Thomas Scharinger, *Was bleibt von kommunikativer Nähe und Distanz?* Tübingen: Narr.

- SAN JULIÁN SOLANA, Javier (2018): “La Concordancia (*ad sensum*) con sustantivos cuantificadores en español”, en *Verba*, 45, pp. 67-106.
- SÁNCHEZ AVENDAÑO, Carlos (2007): “‘Para que la gente se enteren’. La concordancia *ad sensum* en español oral”, en *Filología y Lingüística*, XXXIII/2, pp. 205-226.
- THOMAS, Alan R. (2008 [1994]): “English in Wales”, en Robert Burchfield (ed.), *The Cambridge History of the English Language. Volume V. English in Britain and Overseas: Origins and Development*. Cambridge: Cambridge University Press, pp. 94-147.
- TRUDGILL, Peter (1999): “Standard English: What It Isn’t”, en Tony Bex y Richard J. Watts (eds.), *Standard English: The Widening Debate*. London/New York: Routledge, pp. 117-128.
- (2009): “Vernacular Universals and the Sociolinguistic Typology of English Dialects”, en Markku Filppula, Juhani Klemola y Heli Paulasto (eds.), *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*. London: Routledge, pp. 304-322.
- WIFFELS, Jan (2020): “udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the ‘UDPipe’ ‘NLP’ Toolkit”. R package version 0.8.4-1. <<https://CRAN.R-project.org/package=udpipe>>
- WEINREICH, Uriel, LABOV, William, y HERZOG, Marvin I. (1968): “Empirical Foundations for a Theory of Language Change”, en Winfred P. Lehmann y Yakov Malkiel (eds.), *Directions for Historical Linguistics. A Symposium*. Austin/London: University of Texas Press, pp. 95-188.