

¿La reversión de un cambio lingüístico?: Pasado y presente de la variación en las
relativas oblicuas de lugar en español
The reversal of a linguistic change? Past and present of the variation among the
locative relatives sentences in Spanish
Heading title: ¿La reversión de un cambio lingüístico?

José Luis Blas Arroyo
Universitat Jaume I
blas@fil.uji.es

Abstract

The panchronic perspective adopted in this study shows an apparently contradictory outcome in a phenomenon of variation and change that has a long tradition in Spanish. The variationist analysis of several oral corpora representative of current Peninsular Spanish displays an apparent regression of the pronominal forms of the relative in locative sentences ('la casa en la *que* vivo' ['the house in which I live']), in favour of the adverbial ones ('la casa *donde* vivo' [the house where I live]). This result contrasts with a movement in the opposite direction, previously found in the analysis of a historical corpus composed of texts of communicative immediacy, between the sixteenth and the first half of the twentieth centuries. However, this apparent reversal of the change contrasts with the constraints attested in the internal grammar and the extralinguistic context, which reveal a robust persistence over time, and, in addition, with the same direction of effect. To sum up, mainly due to some irregularities between the corpora analyzed, rather than the reversion of a linguistic change, what the panchronic analysis reveals is a remarkable stability of the variation over more than five centuries of history. This result sheds light on variation not leading to language change, and confirms the need to transcend the most superficial aspects in the study of change.

Keywords: Language Variation and Change, Comparative Sociolinguistics, Panchronic Analysis, Relative sentence, Pronoun / adverb variation, Spanish.

1 Introducción

El presente estudio parte una revisión del principio de uniformidad laboveano, según el cual el conocimiento de los procesos de variación y cambio lingüístico en la historia puede resultar de gran utilidad para interpretar cabalmente algunos ejemplos de los que todavía somos testigos en la actualidad (Bergs 2012). Con ello pretendemos dar cuenta de un fenómeno de alternancia sintáctica que cuenta con una larga tradición en la historia del español y, que lejos de haber concluido, muestra todavía una notable vitalidad en la lengua contemporánea. Los fragmentos de (1) y (2) corresponden a sendos ejemplos del género epistolar privado, escritos en el siglo XVI, por parte de inmigrantes españoles en América. Por su parte, los de (3) y (4) pertenecen a muestras de habla oral del *Macrocorpus del español hablado en Castellón* (MCSCS), un conjunto de entrevistas sociolingüísticas realizadas a individuos representativos de esta comunidad de habla a comienzos del siglo XXI. Como se puede apreciar, en todos ellos el contexto sintáctico es similar, con la presencia de un mismo antecedente (*pueblo*), un sustantivo cuya referencia locativa es recuperada en ocasiones por el relativo adverbial *donde* (ejemplos 1 y 3) y en otras por el pronombre *que* (ejemplos 2 y 4).¹ Estos

fragmentos demuestran la pervivencia de este fenómeno de variación desde el español clásico hasta la actualidad:

- (1) ... pienso salir presto de él para yr a acabar lo que resta de la vida en ese **pueblo** donde pienso seruir y descansar mucho a vm. (*Vida y fortuna del emigrante navarro*, 1576)
- (2) ... y porque vuestra alteza sepa que obra es esta destos **pueblos** que ariba tocamos en que al presente estamos ocupados y el fruto que della se espera (*Textos del Caribe*, 1518)
- (3) ... pues... cómo te diría (2 s.) inspirada el... en un **pueblo**, en un **pueblo** donde... donde... pues se intenta reflejar quee que las chicas van a los sex shops y demás (MCSCS, 42)
- (4) ... pero que hay **pueblos** en los que no no tengo nada que ver y que también son muy bonitos de ver (MCSCS, 326)

En el pasado hemos realizado ya un análisis diacrónico detallado acerca de este fenómeno, incluidos los condicionantes que lo han regulado en diferentes periodos de la historia. Así, mediante las herramientas de la sociolingüística histórica, y a partir de un corpus integrado por textos escritos cercanos al polo de la inmediatez comunicativa (Oesterreicher, 2004), en un trabajo anterior tuvimos ocasión de advertir una evolución lenta, pero sostenida en el tiempo, a favor de las formas pronominales, que, a mediados del siglo XX, habían superado a las relativas adverbiales, históricamente más frecuentes en estos contextos locativos (Blas Arroyo y Velando, 2017).

¿Qué ocurre en la actualidad con este fenómeno sintáctico? ¿Ha seguido las mismas pautas que en el pasado, consolidando la sustitución progresiva de los relativos adverbiales por otros pronominales? ¿Son los condicionantes lingüísticos y extralingüísticos similares a los que observamos en periodos anteriores? O, por el contrario, ¿se aprecian movimientos de ruptura tanto en la gramática interna como en el eje sociolectal? Con el presente estudio pretendemos dar respuesta a estos interrogantes, al tiempo que ponemos en práctica un análisis comparativo y pancrónico de la variación y el cambio lingüístico con el que hacer realidad la premisa expuesta al comienzo de estas páginas: cómo el pasado puede ayudarnos a entender mejor las fuerzas que regulan estos fenómenos en el presente estadio de lengua.

2 La alternancia adverbio / pronombre en las oraciones de relativo locativas

Las oraciones de relativo locativas son subordinadas adverbiales que encierran el significado de ‘lugar en donde’, ya sea con antecedente explícito (‘Es el río *donde* / *en el que* – *cual* pesco) o implícito (‘Estamos *donde* queremos’). Ahora bien, como revela este último ejemplo, la alternancia entre el adverbio *donde* y los pronombres tan solo es posible en las relativas del primer tipo (‘*Estamos *en el que* queremos’), de ahí que el presente análisis se ciña exclusivamente a estas.

Las cuestiones relativas a esta variación se recogen en diferentes gramáticas del español (RAE, 1973: 533; Alcina y Blecua, 1991 [1975]: 1114, §8.3.5.5; RAE y ASELE, 2009: 1600--1601, entre otras). En todas ellas, se alude a la naturaleza del antecedente, que debe ser nominal o pronominal, pero nunca adverbial, ya que en estos casos no hay variación posible (‘Fue allí *donde* / **en el que* lo vimos’) (Brucart, 1999: 509, §7.5.6.1; Herrero Ruiz de Loizaga, 2005: 209--210). Asimismo, se subraya el hecho de que los relativos pueden referir también a antecedentes no prototípicamente locativos, aunque reinterpretados metafóricamente con este valor (‘Es la carta *donde* /

en la que comunicaba su despido’) y, más ocasionalmente, como un punto o lapso en el eje temporal (‘Es un momento de tensión en el que / donde todos nos ponemos nerviosos’) (Girón, 2009: 1550, §14.5.3.1; RAE y ASELE, 2009: 1600--1601, §22.8k, §22.8l y §22.8m).

Aunque la mayor parte de los estudios sobre relativos dan cuenta, pues, de esta alternancia entre pronombres y adverbios, son más escasas las referencias a su posible interpretación como una variable lingüística, esto es, como formas alternantes de las que se sirven los hablantes para expresar un mismo contenido referencial. Esta cuestión tiene que ver, inicialmente, con la potencial equivalencia (mejor que sinonimia) de estas formas, cuestión a la que se refiere, por ejemplo, Cortés (1986). Este autor da por sentada la paridad entre las variantes adverbiales y pronominales, dado su idéntico valor locativo, la misma función de complemento circunstancial y su presencia indistinta en oraciones explicativas y especificativas. Para otros, sin embargo, *donde* y *que* (o *cual*) no significan lo mismo. Así, en un estudio sobre este tema, a partir de un corpus de textos españoles y americanos de los siglos XVI, XVII y XVIII, González Díaz (2009: 43) afirma lo siguiente: “La existencia de usos particulares de *donde* y PAR² permite afirmar que *ambas formas no son totalmente equivalentes*, tal como se ha entendido tradicionalmente” (la cursiva es nuestra). Y en otro lugar, a propósito esta vez de una muestra de textos periodísticos actuales, esta misma autora (González Díaz 2006: 772) sostiene que, a diferencia de Venezuela, país en el que *donde* es la forma predominante en la mayoría de los contextos lingüísticos –“[y] por lo tanto, aparece en los contextos que le son propios y en los que se esperaría una forma PAR”–, en España: “los relativos *donde* y PAR presentan una distribución casi complementaria”. (de nuevo, las cursivas son nuestras). Esta complementariedad, esto es, el hecho de que cuando se emplea una forma no lo hace la otra, y viceversa, vendría a suponer que *donde* y *en (el) que – cual* no son (en España) formas alternantes de una misma variable lingüística (aunque sí, aparentemente, en Venezuela).

Resultados tan dispares como estos revelan potenciales diferencias dialectales, aunque con un denominador común en la mayoría de los estudios sobre el español actual: el reducido empleo de *donde*, frente a la utilización cada vez más extendida de *que*. En su trabajo sobre los relativos en una red social andaluza (Baeza, Jaén), Rascón (2005) destaca así la escasa utilización del adverbio *donde*, lo que, en su opinión, podría explicarse por la naturaleza más ‘coloquial’ de los pronombres (Rascón 2005:379) (en el mismo sentido, véase Fernández Ramírez, 1987 [1951]: 244; y para una visión opuesta, Blas Arroyo y Velando, 2017 y, más adelante, el apartado §4 de este mismo artículo). Por su parte, Herrera Santana (1994--95) indica que, en el español de Santa Cruz de Tenerife, los empleos de *donde* apenas sobrepasan el 5,1%, frente a un contundente 88% de *que* (el resto corresponde a *cual*). Ahora bien, aunque la autora dedica una sección concreta a la alternancia entre ambos pronombres, no hay un análisis específico del contexto variable, por lo que es imposible determinar cuáles son realmente los límites de esa variación y cómo se establecen las comparaciones. Y lo mismo ocurre con algunas cifras sobre el empleo de *donde* y *que* en el español de México, tanto en su variedad culta (7% / 88.5%) (Palacios de Sámano, 1983) como en el habla popular (7.4% / 90.2%) (Mendoza, 1984). Por su parte, Girón (2009: 1548), en un estudio diacrónico a partir de textos escritos en diversos momentos de la historia del español, confirma también que las formas pronominales han sido más frecuentes que las adverbiales, al menos desde el siglo XVII.

Nuestros datos, sin embargo, difieren en lo esencial de estos recuentos. A partir de un análisis preciso del contexto en que tiene lugar la variación (*envelope of variation*) entre las formas adverbiales y pronominales, y mediante el concurso de textos cercanos al

polo de la inmediatez comunicativa entre los siglos XVI y XX, en otro lugar hemos tenido ocasión de comprobar cómo, hasta tiempos recientes (1.^a mitad del XX), la variante preferida en los contextos locativos ha sido *donde* y no *que*, al menos en el español europeo (Blas Arroyo y Velando, 2017). Con todo, esos mismos datos parecían apuntar hacia un lento proceso de cambio lingüístico, por el cual las formas pronominales iban ganando terreno en detrimento de las adverbiales. Aun así, el estudio revelaba también una notable continuidad de las fuerzas estructurales y extralingüísticas entre los diferentes periodos, un dato relevante que pondremos también a prueba en el estudio sobre el español peninsular contemporáneo del que damos cuenta en estas páginas, y cuyos principales detalles metodológicos se exponen a continuación.

3 Metodología

3.1 Corpus

El análisis pancrónico de este fenómeno parte de dos corpus. El primero, ya examinado en trabajos previos, es un conjunto textual compuesto íntegramente por textos cercanos al polo de la inmediatez comunicativa (Oesterreicher, 2004), conceptualmente más cercanos a la oralidad que aquellos con los que ha trabajado tradicionalmente la lingüística histórica. Dentro de estos, destacan los textos epistolares, especialmente miles de cartas de naturaleza privada, en las que ventilan asuntos diversos, desde temas y relaciones familiares y personales en un extremo, a otras de carácter menos íntimo (comerciales, oficiales, etc.).³ Al mismo tiempo, los más de 2600 autores de estos textos se distribuyen por todo el espectro social, desde las élites a representantes de profesiones manuales no cualificadas, pasando por diversos niveles intermedios. Todo ello representa un corpus integrado por más de tres millones de registros (para más detalles sobre el proceso de codificación de este corpus histórico, véase Blas Arroyo y Velando, 2017).⁴ La distribución por siglos de todas estas magnitudes se puede ver en la tabla 1.

Tabla 1: Dimensiones del corpus por siglos (corpus histórico)

	Nº palabras	Nº escritores
Siglo XVI	1.083.034	889
Siglo XVIII	1.242.588	1345
Siglo XX	1.037.849	410
TOTAL	3.363.471	2644

Por su parte, el corpus oral contemporáneo se compone de diversos subcorpus sociolingüísticos, compilados entre finales del siglo XX y los primeros años del siglo XXI. Tres de ellos corresponden a otras tantas ciudades españolas integradas en el Proyecto PRESEEA, cuyos materiales se han publicado en diversos formatos durante los últimos años.⁵ Se trata de: Madrid (Barrio de Salamanca), Málaga y Valencia. A estas se añade Castellón, por medio del ya mencionado *Macrocorpus sociolingüístico de Castellón y sus comarcas (MCSCS)* (ver §1).

Los materiales de estos cuatro corpus son comparables no solo por las fechas similares en que fueron confeccionados, sino también por la metodología empleada en su elaboración: entrevistas semidirigidas con una duración media aproximada de 45--50 minutos. En su desarrollo se combinó el método antropológico de las historias de vida, junto a otros módulos temáticos destinados a obtener diferentes tipologías textuales

(fragmentos expositivos, narrativos, argumentativos...). Asimismo, todos los corpus se hallan estratificados inicialmente por cuotas de sexo, edad y nivel de estudios.

El resultado de todo ello es la obtención de muestras de habla de 315 informantes, representativos de diversas variedades del español peninsular. En la tabla 2 se puede ver su distribución por comunidades de habla.

Tabla 2: Número de informantes por comunidad (corpus oral, s. XXI)

Comunidades	N.º informantes
Castellón	123
Valencia	72
Madrid (Barrio de Salamanca)	54
Málaga	66
Total	315

3.2 Codificación

Mediante un programa de concordancias (Wordsmith 6.0) se localizaron en el corpus todas las ocurrencias de las formas adverbiales y pronominales. A continuación, y tras excluir de la matriz de datos aquellas que no presentan variación (ver lo señalado en §2), cada ejemplo se codificó de acuerdo con diversos factores lingüísticos, estilísticos y sociales. La relación y tipología de estos se resume a continuación en la tabla 3, dejando para más adelante (§4) el detalle de aquellos que se han revelado más importantes para explicar la variación.

Tabla 3: Breve resumen y tipología de los predictores considerados en el análisis

Predictores	Factores	Breve descripción	Tipo
Tipo de relativa	Explicativa Especificativa	Clasificación sintáctico – semántica de la oración encabezada por el pronombre	Categorico
Semántica	Lugar1 Lugar2 Otros	Sentido prototípico del antecedente	Categorico
<i>Priming 1</i>	Pronombre Adverbio Resto	Presencia de las categorías pronominal y adverbial en el cotexto previo	Categorico
<i>Priming 2</i>	Presencia Ausencia	Presencia / ausencia de la preposición <i>en</i> en el antecedente	Categorico
Género	Masculino Femenino	Género del antecedente	Categorico
Número	Singular Plural	Número del antecedente	Categorico
Función	C. Circunst. Otras	Función sintáctica del antecedente	Categorico
Adyacencia	Adyacencia Resto	Posición del relativo con respecto a su antecedente	Categorico
Distancia	N.º de palabras	Distancia del relativo con respecto a su antecedente, medida en número de palabras	Continuo
Tenor (solo corpus histórico)	Personal Distante	Tipo de comunicación, relacionada con el tema y/o la relación entre los interlocutores	Categorico
Estilo (solo c. oral)	Fases interacción	Interpretación laboveana del estilo a partir de la atención dispensada por el hablante a su habla. Para ello se establecen cinco etapas de la	Continuo

		interacción, medidas a partir del porcentaje que arrojan en el programa de concordancias (1--20%; 21--40%; 41--60%; 61--80%; 81--100%). La hipótesis es que, conforme avanza la conversación, el nivel de atención disminuye.	
Sexo	Hombre Mujer	Sexo del informante	Categorico
Estatus	Alto Medio Bajo	Nivel sociocultural de los informantes, medido a través de diversos parámetros, como la posición social heredada, el estatus profesional, etc. (corpus histórico) y el nivel de estudios (corpus oral)	Categorico
Edad1 (corpus oral)	Edad del informante	Edad del informante medida en cifras absolutas	Continuo
Edad2 (corpus histórico)	Jóvenes Adultos	Interpretación émica de la edad a partir de los contextos vivenciales por los que atraviesa el informante	Categorico
Lengua dominante	Español – Catalán – valenciano	Lengua dominante de los miembros de la muestra a partir de su testimonio en una breve encuesta al final de la entrevista (solo en el corpus oral de la Comunidad Valenciana)	Categorico
Antecedente	Lema	Lema del antecedente, normalizada a partir de la forma masculina en singular.	Aleatorio
Informante	Nombres	Identidad del hablante recogida a partir de su nombre (corpus histórico) o de un código específico (corpus oral)	Aleatorio

Las formas alternantes recogidas en el corpus son las siguientes: *donde* en el plano adverbial, y *que*, *el que* y *el cual* precedidas por la preposición *en*, en el paradigma pronominal. Dado que el interés de la investigación se centra en analizar la progresión de estas últimas en detrimento de las primeras en el español actual, al tiempo que se compara con la observada en periodos previos de la historia, para el análisis cuantitativo agrupamos todas ellas en un solo bloque, que consideramos como variante de referencia (*application value*) a los efectos del análisis de regresión binario, cuyos fundamentos se explican en el siguiente apartado.

3.3 *Análisis estadístico*

Para el estudio cuantitativo nos servimos inicialmente de una estadística descriptiva, que nos permite obtener una primera radiografía acerca de las tendencias de variación predominantes en forma de frecuencias y porcentajes. Con todo, la verdadera significación de estas medidas tan solo puede emprenderse a través de una estadística multivariable, asociada a los modelos de regresión logística (Tagliamonte y Baayen, 2012). Dentro de estos, en los últimos años han adquirido cada vez más relevancia los modelos de efectos mixtos, aquellos que combinan tanto factores categoriales fijos (tipo de relativa, sexo...) como otros de naturaleza continua (distancia al antecedente, edad...) y aleatoria (lema del antecedente, informante). Ello da lugar, generalmente, a resultados más conservadores, pero, como contrapartida, más fiables y consistentes. Para este análisis, en el presente estudio nos servimos del programa *Rbrul* (Johnson, 2009), cuya interfaz, sencilla e intuitiva, permite acceder a dos de las medidas fundamentales de todo estudio variacionista: a) la significación de los factores seleccionados como significativos, y b) la dirección explicativa en el seno de los correspondientes predictores.

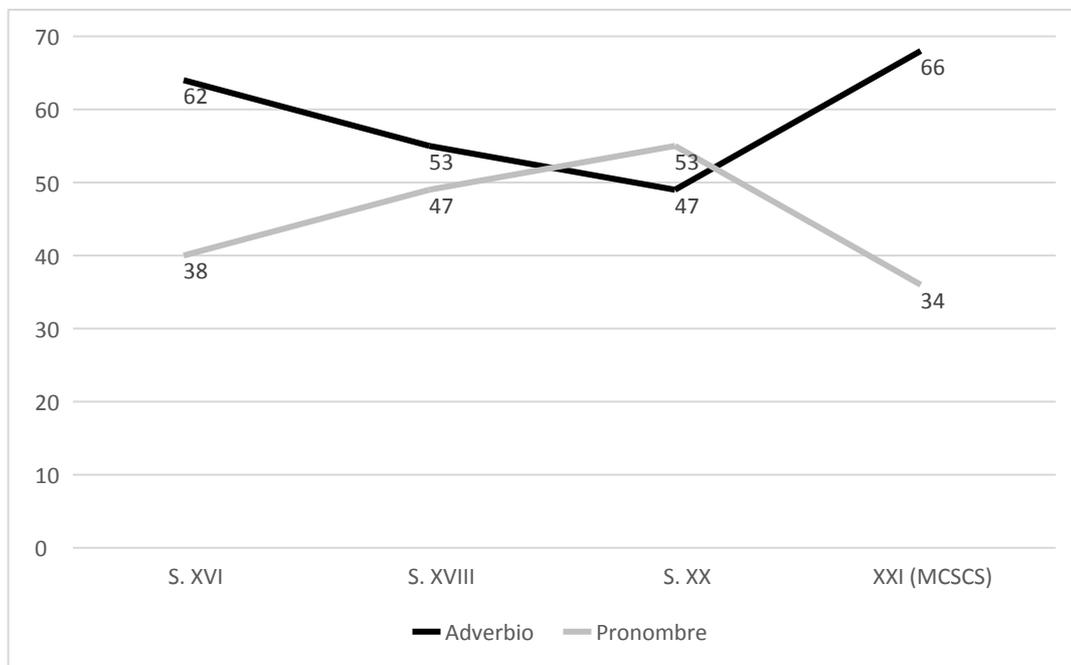
Sin embargo, pese a su potencial predictivo, los análisis de regresión logística no siempre permiten apreciar fácilmente la jerarquía explicativa de estos predictores y las sutiles y complejas interacciones que se producen frecuentemente en su interior, sobre todo cuando en el estudio se examina un número considerable de estos. Por este motivo, en los últimos años han adquirido carta de naturaleza algunos modelos no paramétricos de regresión que palían tales deficiencias, como ocurre con los árboles de inferencias condicionales (*conditional inferential trees*) o los bosques aleatorios (*random forest*), ambos posibles a través del entorno de R (Levshina 2015). Con ellos podemos comprobar no solo qué condicionantes resultan más decisivos que otros, sino también el modo en que trabajan en tándem (para más detalles acerca de la interpretación de estos, véase §4.2).

4 Resultados y análisis

4.1 Resultados generales

El gráfico 1 muestra la evolución de las variantes adverbiales y pronominales en cinco siglos de la historia del español. Como se puede apreciar, a comienzos del siglo XXI, el cambio favorable a las formas pronominales que advertíamos en el pasado, parece truncado. Aunque no cabe descartar que en las diferencias frecuenciales haya podido influir, en alguna medida, la naturaleza escrita (cartas) u oral de los textos de inmediatez comunicativa analizados, la entidad de aquellas es de suficiente magnitud como para presumir inicialmente la reversión de un cambio lingüístico histórico. De este modo, en los primeros años del siglo XXI, las formas pronominales descienden casi veinte puntos porcentuales, hasta un 34%, con respecto a las cifras obtenidas a mediados del XX (53%), regresando así a proporciones advertidas en el español clásico (38%). Sea como sea, estos resultados desmienten la pretendida inferioridad del adverbio *donde* con respecto a *que*, señalada previamente en la bibliografía (ver §2). A nuestro juicio, esta disparidad en los resultados podría obedecer al método utilizado en la comparación. Así, en el contraste entre las frecuencias absolutas de ambas formas en un corpus de español cualquiera, el pronombre *que* sale, sin duda, victorioso. Sin embargo, esto no ocurre cuando, como aquí, la comparación se establece en un marco de variación determinado.

Gráfico 1. Evolución de la variable lingüística entre los siglos XVI y XXI (%)



Por comunidades, tal como revela la tabla 4, tan solo una ciudad, Málaga, ofrece cifras significativamente más bajas de la variante pronominal (19%).⁶ De las demás, las ubicadas en la Comunidad Valenciana (Valencia y Castellón) se sitúan levemente por encima de la media, aunque en proporciones muy similares, al tiempo que Madrid lo hace ligeramente por debajo.

Tabla 4: Frecuencias de uso de la variante pronominal por comunidades de habla (corpus oral)

Comunidades	N	%	Σ
Castellón	197	41	485
Valencia	91	38	242
Madrid (Barrio de Salamanca)	77	30	251
Málaga	45	19	240

4.2 *Análisis del contexto variable en el español contemporáneo*

En el presente apartado reseñaremos los principales resultados que ofrece el análisis cuantitativo en el corpus oral contemporáneo, y dejaremos para más adelante el (§4.3) el estudio comparativo con etapas previas.

Como muestra la tabla 5, y a la luz de la muestra disponible (1218 ocurrencias de la variable), la selección de las formas pronominales es sensible a cinco predictores, tres de naturaleza lingüística (*semántica* y *función* del antecedente, además de una de las manifestaciones de *priming* estructural – *Priming* 1–) y dos de carácter extralingüístico (*región* y *nivel sociocultural*). Los demás factores exhiben diferencias menores (en algunos casos, prácticamente nulas), por lo que han sido descartados por el análisis de regresión. Con todo, en algún caso se advierten algunas tendencias que merecen nuestra atención, y que consideraremos también más abajo.

Tabla 5: Contribución de diferentes predictores a la selección de las formas pronominales en un corpus oral contemporáneo

	N	%	Σ	P. (A.)	P. (I.)	P. (C.)

Semántica (antecedente)						
Lugar 1	50	11	449	.24	.16	.22
Lugar 2	63	18	346	.36	.34	.38
Otros	297	70	423	.84	.91	.85
Función (antecedente)						
CC	105	22	477	.41	.39	.42
Otras	305	41	741	.59	.61	.58
Priming 1						
Pronombre	20	80	25	.85	.84	.88
Donde	5	5	109	.13	.13	.10
Resto	385	35	1084	.52	.54	.54
Priming 2						
Ausencia	335	36	924	–	–	–
Presencia	75	25	294	–	–	–
Adyacencia						
Adyacente	278	32	881	–	–	–
No	132	39	337	–	–	–
Distancia (continuo)						
Tipo de relativa						
Explicativa	136	32	423	–	–	–
Especificativa	274	35	795	–	–	–
<i>Rango</i>						
Género						
Masculino	194	29	664	–	–	–
Femenino	216	39	554	–	–	–
Número						
Singular	321	33	989	–	–	–
Plural	89	39	229	–	–	–
Sexo						
Hombre	197	31	627	–	–	–
Mujer	213	36	591	–	–	–
NSC						
Alto	218	40	545	.63	.61	.60
Medio	150	33	448	.52	.55	.54
Bajo	42	19	225	.36	.34	.36
Edad (continuo)						
Región						
Com. Valenciana	288	40	727	.62	.62	–
Madrid	77	31	251	.54	.58	–
Málaga	45	19	240	.33	.30	–
Lengua (solo CV)						
Castellano	126	37	342	–	–	–
Valenciano	162	42	385	–	–	–
Estilo (continuo)						
				–	–	–

Comenzando este repaso por los factores de naturaleza lingüística, cabe señalar, en primer término, la robustez de la *Semántica* del antecedente, que, como vimos (§3.2) atiende al sentido básico de este. Para su configuración, hemos dividido las ocurrencias de la variable en tres grupos. En el primero (*Lugar 1*), consideramos los referentes más prototípicamente locativos, aquellos que aluden a ciudades (*Madrid*), países (*España*), accidentes geográficos (*monte*), lugares al aire libre (*cuesta*), etc. Por su parte, los codificados como *Lugar 2* pueden interpretarse también como ubicaciones, pero de una manera menos prototípica. A diferencia de los anteriores, aluden a edificios (*casa*), y construcciones en general (*barco*), generalmente cubiertas, en contraste con los

sustantivos del grupo anterior. El resto de antecedentes se agrupan en la categoría de *Otros*, que da cuenta de los referentes cuya naturaleza semántica original no es locativa, pese a que puedan aparecer en contextos de este tipo (*lazo, carta, sociedad, etc.*), así como algunos cuyo sentido originario es temporal, pero pueden interpretarse, por traslación metafórica, como ‘posiciones’ en el tiempo (*momento, situación, temporada, etc.*)

Como se aprecia en la tabla 5, la selección de las formas pronominales es extraordinariamente sensible al carácter semántico de estos antecedentes, y en particular a la oposición entre las referencias locativas y no locativas, siendo estas últimas las que más contribuyen a su aparición. Así, en el 70% de los ejemplos con antecedentes no locativos (*Otros*), los hablantes seleccionan alguna de las formas pronominales,⁷ lo que se corresponde con uno de los pesos explicativos más elevados del estudio (.84).⁸ Por el contrario, los antecedentes locativos son mucho más desfavorables a estas formas, en especial los de naturaleza más prototípica (*Lugar 1*) (11%; .24), cuyas restricciones a la elección pronominal superan incluso a las de otros no tan prototípicos (*Lugar 2*) (18%; .36).

Un segundo factor estructural que condiciona también de manera independiente la variación atiende a la función sintáctica del antecedente. De esta manera, comprobamos que casos como los del fragmento (5), en los que el antecedente desempeña la función de *complemento circunstancial*, inhiben más la selección de una variante pronominal (22%; .41) que los demás contextos sintácticos (*Otras*) (41%; .59), como el ejemplificado en (6), donde el antecedente es complemento directo:

(5) ...porque aquí en cuanto se termina la naranja **en esta zona** en que vivo yo, en cuanto se termina la naranja, qué pasa? pues que el mal aumenta! (MCSCS, 380)

(6) ... cuando tenía dieciocho años, en mi casa, que teníamos **una fábrica de gaseosa** en la que trabajamos, pues resulta que teníamos una camioneta...(MCSCS, 413)

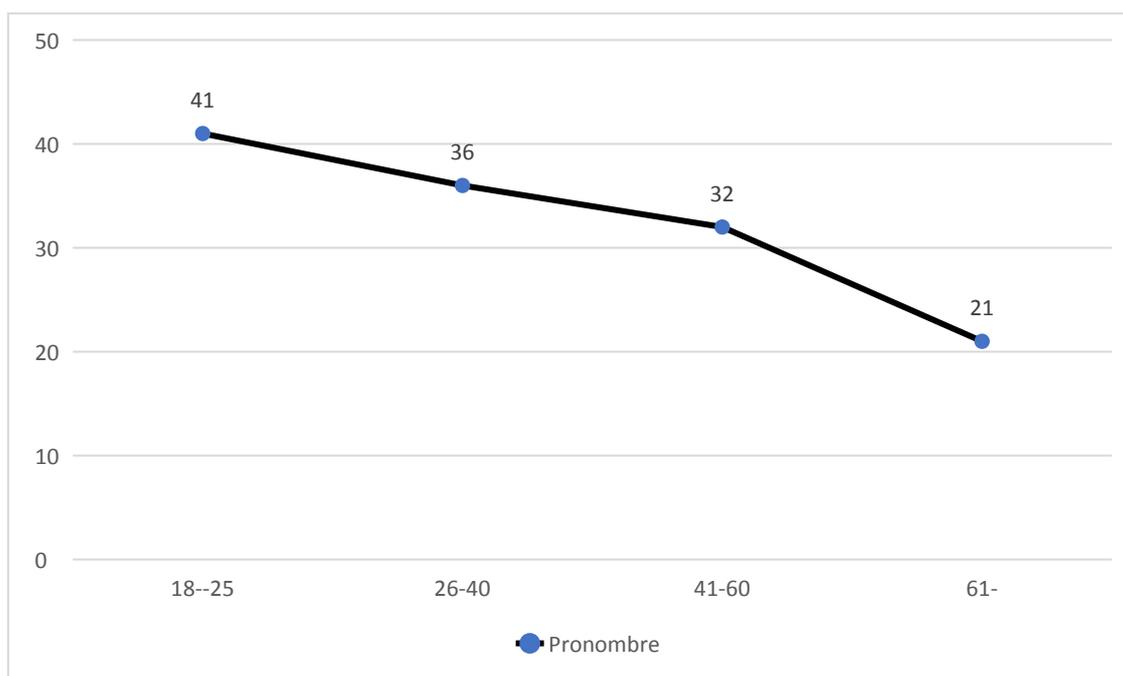
Finalmente, la variación es sensible también al *priming* estructural, y en particular a la presencia en el contexto previo inmediato de las mismas o diferentes formas.⁹ Así, la tabla 5 revela que aquellos casos en los que la variante pronominal está precedida por otra forma pronominal, como en (7), son muy proclives a la repetición (80%; .85). Por el contrario, cuando ese papel está desempeñado por una forma adverbial, las posibilidades de duplicación se reducen drásticamente (5%; .13):

(7) ...vacaciones en las **que** están los alumnos hay muchos niveles ee luego tienes... yo me suelo fie... fijar mucho en... (chasquido) en la situación familiar en la que están (VAL, Alto02)

Por lo que al eje socioestilístico se refiere, los resultados del estudio no arrojan apenas diferencias generolectales, así como entre las diversas fases de la interacción, criterio que, como se recordará hemos tomado como referencia para evaluar potenciales diferencias diafásicas (para más detalles sobre la configuración de este factor ver §3.2). Del mismo modo, tampoco en el caso de la Comunidad Valenciana hay diferencias relevantes entre castellanohablantes y valencianohablantes dominantes, lo que probablemente es un correlato de la ausencia de puntos de conflicto estructural entre español y catalán en esta área de la sintaxis.¹⁰ Finalmente, a la luz de la muestra disponible, tampoco la edad, considerada en el análisis de regresión como un factor

continuo, revela diferencias significativas. Con todo, el gráfico 2 muestra una distribución lineal en el uso de las variantes pronominales, tras su agrupamiento en cuatro bloques etarios: 18--25; 26--40; 41--60 y 61--). En él se puede apreciar cómo el uso de estas formas disminuye conforme avanza la edad de los informantes, con un descenso especialmente acusado en el grupo de edad más avanzada.

Gráfico 2: Distribución de las variantes pronominales por grupos de edad en el corpus oral (%)



Al margen de estas diferencias, para cuya interpretación no tenemos, sin embargo, suficiente fundamento empírico, la variación se muestra sensible al nivel sociocultural de los hablantes. De este modo, las mayores cifras de la variante se producen entre los representantes del grupo más alto (40%; .63), seguidos por los informantes del grupo medio (33%; .52), con proporciones cercanas a la neutralidad, y a considerable distancia de los hablantes con menor formación, que figuran entre los sectores menos proclives a su empleo (19%; .36). Estos resultados parecen contradecir la idea de que estas formas pronominales sean más ‘coloquiales’ que las adverbiales, como se ha propuesto en alguna ocasión (Rascón, 2005). El hecho de que sean impulsadas preferentemente por los individuos con mayor nivel sociocultural y, como veíamos en el párrafo anterior, por los grupos de edad más potencialmente afectados por las presiones del estándar tras su paso por el sistema educativo (18--25) o su entrada reciente en el mundo laboral (26--40), parece sugerir justamente lo contrario y, eventualmente, la posibilidad de que estemos asistiendo a un cambio desde arriba en el presente estadio de lengua (Labov, 2001).

Finalmente, la variación se ve condicionada también por la procedencia geográfica de los hablantes. Agrupados los datos de Castellón y Valencia – no solo por su proximidad dialectal, sino también por unas cifras singularmente parecidas – sus resultados se contraponen a los de las otras dos comunidades (Madrid y Málaga). El análisis de regresión arroja diferencias significativas entre las tres variedades, con las ciudades valencianas como principales adalides en el uso de la variante pronominal (40%; .62), seguidas por el madrileño Barrio de Salamanca (31%; .54) y a distancia de Málaga, con mucho la comunidad de habla menos proclive al empleo de las formas pronominales (19%; .33).

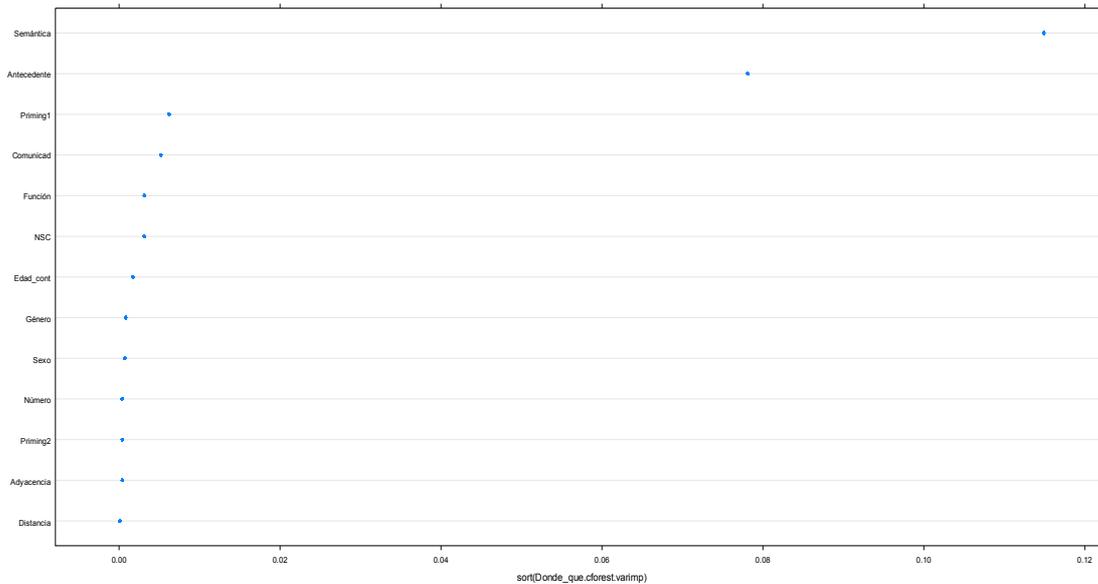
Hasta el momento hemos visto cuáles son los condicionantes lingüísticos y extralingüísticos a los que la variación se muestra sensible, y cuál es la dirección del efecto en el interior de cada uno de ellos. Ahora bien, ¿cómo es la jerarquía entre esos condicionantes? Y, sobre todo ¿cómo dar cuenta de ella? Para obtener una respuesta a

estos interrogantes hemos llevado a cabo un análisis de regresión no paramétrica, denominado *bosque aleatorio* (*random forest*), que permite evaluar con precisión esa jerarquía, incluso cuando se presentan muchos predictores al mismo tiempo y de diferente naturaleza (categóricos, continuos y aleatorios), superando con ello las insuficiencias de los tradicionales *rangos*. El gráfico 3 muestra cómo en cabeza de esos predictores, y a considerable distancia de los demás, se encuentra la *semántica* del antecedente, seguido a continuación por el *lema* de este. Estos datos confirman la relevancia del factor semántico que señalábamos más arriba (§4.2), pero también la existencia de una notable variación en el seno de algunos antecedentes.

De hecho, el análisis de este factor aleatorio – cada vez más frecuente en los estudios de variación, pero pocas veces interpretado – muestra algunos resultados reseñables. La mayoría de los antecedentes que aparecen en el corpus lo hacen aisladamente, en proporciones que se reducen en muchos casos a una o dos ocurrencias. En ellas, los hablantes se decantan por una u otra variante, siguiendo por lo general la regla mencionada a propósito de la semántica del antecedente. Fuera de estos, los ejemplos de selección categórica de una variante determinada son escasos, aunque encontramos algunos (entre paréntesis figuran las frecuencias correspondientes a las formas pronominales y adverbiales, respectivamente): *edificio* (0 / 4), *momento* (9 / 0), *hotel* (0 / 4), *local* (0 / 4). Como se puede comprobar, asistimos también aquí al seguimiento de la norma general impuesta por el sentido del antecedente: los sustantivos que refieren lugares de manera más nítida optan por la variante adverbial, mientras que los dotados de otro sentido, eligen un pronombre. Con todo, no faltan los ejemplos de lo contrario. Así, un sustantivo como *película*, conceptualizado dentro del grupo menos prototípicamente locativo (Otros), se asocia en solo una ocasión con el pronombre y, sin embargo, en seis con el adverbio. Aun así, son todavía más frecuentes los ejemplos de antecedentes con una alta recursividad en el discurso oral, en los que una clara preferencia por una u otra forma del relativo no impide, sin embargo, la existencia de variación (en ocasiones, incluso, en el habla de un mismo individuo). A continuación se exponen aquellos antecedentes que aparecen diez o más veces en el corpus, y en los que se puede apreciar bien lo que decimos: *fiesta* (4 / 6), *barrio* (4 / 36), *calle* (3 / 20), *casa* (13 / 33), *ciudad* (4 / 7), *colegio* (7 / 21), *empresa* (4 / 15), *época* (4 / 9), *lugar* (4 / 18), *país* (3 / 7), *piso* (2 / 9), *pueblo* (5 / 33), *punto* (2 / 9), *sitio* (8 / 144), *situación* (53 / 6), *sociedad* (10 / 3), *trabajo* (7 / 13), *zona* (12 / 41).

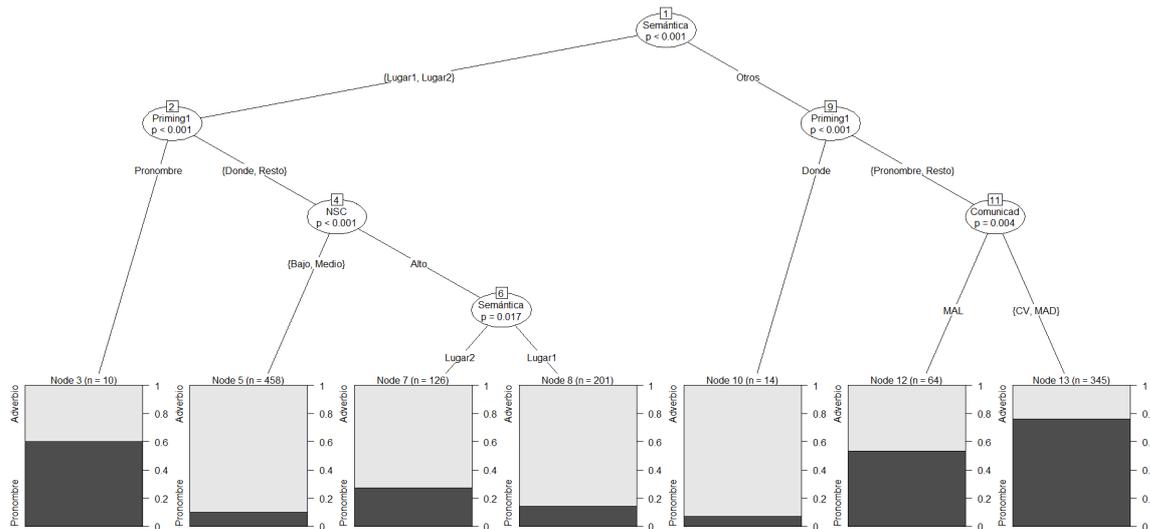
Finalmente, a los factores reseñados le siguen, por orden de relevancia, el paralelismo estructural (*Priming* 1), la comunidad lingüística, el nivel sociocultural, la función del antecedente y la edad, todos ellos (salvo la edad), seleccionados ya por el programa de regresión. El resto de predictores, alineados cerca del eje vertical confirman la nula relevancia explicativa que habíamos advertido ya en el análisis previo.

Gráfico 3: Jerarquía de los predictores en un bosque aleatorio (*random forest*) en la selección de la variante pronominal (C index of concordance: 0.95)



Para concluir esta sección, nos interesa comprobar también el modo en que los predictores más relevantes¹¹ actúan en tándem entre sí y la manera en que estos cruces pueden explicar más cabalmente la variación. Para ello, construimos un árbol de inferencias condicionales (*conditional inferential tree*) con el concurso de aquellos que se han revelado más determinantes, tanto en el análisis de regresión logístico como en el bosque aleatorio que acabamos de reseñar. El resultado puede verse en el gráfico 4.

Gráfico 4: Árbol de inferencias condicionales (*Conditional inference tree*) para la selección de la variante pronominal (C index of concordance: 0.83)



El gráfico vuelve a revelar la importancia decisiva del sentido del antecedente, cuyo nódulo inicial preside todo el conjunto. Al mismo tiempo, muestra cómo los cruces más explicativos difieren entre el grupo de antecedentes locativos agrupados (*Lugar 1* y *Lugar 2*) y el resto (*Otros*). Entre los primeros, el paralelismo estructural (*Priming 1*) establece una nueva oposición que afecta en primer lugar a los relativos pronominales precedidos por otro pronombre, contexto que arroja las mayores probabilidades de selección de la variante. En los demás casos, el nivel sociocultural entra en liza, con la oposición entre el grupo alto y el resto de la sociedad. Las cosas son bien distintas, sin embargo, entre los antecedentes no locativos. Como cabía esperar, en estos casos el cruce con el *priming* discursivo establece una dicotomía entre *donde* – siendo este el contexto menos propicio para la variante pronominal – y el resto. Finalmente, en este último nódulo nace una nueva escisión, relacionada con la comunidad lingüística: en línea con lo ya advertido anteriormente, los hablantes malagueños hacen un uso significativamente menor de la forma pronominal que los valencianos y madrileños.

4.3 *Análisis pancrónico de la variación y el cambio lingüístico*

Pese al cambio en la distribución de las variantes entre el periodo actual, que analizábamos en el apartado anterior, y otras etapas de la historia estudiadas con anterioridad (Blas Arroyo y Velando, 2017), el análisis comparativo que ofrece la tabla 6 muestra una considerable continuidad en el condicionamiento sociolingüístico del fenómeno. En este sentido, la variable exhibe notables ejemplos de persistencia en el tiempo, como se desprende del hecho de que los mismos predictores que actuaron en el pasado para interpretar la variación lo hagan también ahora, y, además, con la misma dirección explicativa en su interior. Si acaso, se aprecian algunas diferencias en la magnitud de esos pesos, que, con todo, y salvo alguna excepción, no implican un debilitamiento de tales predictores, sino más bien lo contrario.

Como hemos visto, en cabeza de todos ellos se sitúa la semántica del antecedente, con aquellos cuya referencia es más prototípicamente locativa (*Lugar 1*) como los menos propicios siempre al empleo de la variante pronominal, lo contrario que los no locativos (*Otros*), sus principales aliados. Por otro lado, el hecho de que sea en el momento presente cuando los pesos probabilísticos se hacen más extremos da cuenta de que, lejos de atenuarse y favorecer la extensión del cambio lingüístico a contextos tradicionalmente adversos, este factor se afianza en la dirección tradicional. En el mismo sentido cabe interpretar el valor del grupo intermedio, el de los antecedentes locativos menos prototípicos (*Lugar 2*): el peso de este factor decrece con el tiempo, desde valores cercanos a la neutralidad en los siglos XVI (.53) y XVIII (.47), a su conversión progresiva en un entorno cada vez más enemigo de las variantes pronominales, tanto en la primera mitad del siglo XX (.37) como en los primeros años del XXI (.36).

Lo mismo sucede con el *priming* discursivo. Si a lo largo de los cinco siglos que median entre los extremos temporales del presente estudio la presencia de un pronombre en el contexto previo inmediato ha sido siempre un acicate para su repetición posterior, esta presión estructural se acrecienta notablemente en los albores de la presente centuria, con un valor P. que figura entre los más altos del conjunto (.85). Y, complementariamente, la aparición de *donde* en los aledaños ha acentuado su efecto desfavorecedor hasta extremos casi categóricos (.13), desde posiciones que, aun siempre hostiles a la variante pronominal, lo eran de manera más atenuada (XVI: .36; XVIII: .34; XX: .38).

Aunque la función sintáctica continúa también con la misma lectura interpretativa a lo largo del tiempo – con las funciones circunstanciales como las menos propicias al empleo de la variante pronominal –, su impacto aparece mitigado esta vez. Así, el efecto inhibitor del complemento circunstancial en el corpus contemporáneo (.41) figura entre los menores de toda la serie histórica (XVI: .29; XVIII: .30), continuando así una senda que ya se adivinaba entre los datos del siglo XX (.42).

Esta misma continuidad entre los dos periodos más recientes se aprecia también en el plano sociolectal. Aunque a lo largo de la historia las élites han liderado siempre las variantes pronominales – desmintiendo nuevamente su carácter más ‘coloquial’¹², la estratificación social tan solo alcanza el umbral de significación estadística a partir de la primera mitad del siglo XX. En este marco, lo que muestran los datos del XXI es un afianzamiento de las diferencias sociales, que se traduce en un patrón de distribución lineal, en el que cada nivel sociocultural se sitúa en un punto diferente en la selección de la variante. De este modo, los niveles altos incrementan su significación favorable a las formas pronominales (.63; frente a .57 en la 1.ª mitad del XX), el grupo medio se queda en una posición intermedia, cercana a la neutralidad (.52), mientras que el estrato más bajo acentúa su alejamiento de aquellas (.36).

Por último, cabe destacar también la coincidencia que supone la no selección de los mismos factores lingüísticos (*Priming 2*, *tipo de relativa*, *género*, *número*) y extralingüísticos (*sexo*) en todos los periodos considerados en el presente estudio pancrónico. Mención especial merece, sin embargo, la *edad*. Aunque, como vimos el factor no ha sido seleccionado como significativo en el corpus oral contemporáneo, la agrupación etaria de los informantes en cuatro bloques muestra una correlación inversa entre la edad y el grado de uso de la variante pronominal, siendo especialmente acusadas las diferencias entre los grupos extremos (18--25 años; 61 en adelante). Esta correlación es, sin embargo, inversa a la que advertimos en el corpus histórico. Si bien el factor solo fue seleccionado como significativo en la muestra del XVIII, las diferencias frecuenciales en el XX apuntaban en la misma dirección: en ambas etapas, la

variante pronominal era favorecida por los más adultos, mientras que los más jóvenes eran menos proclives a su empleo.

5 Discusión y conclusiones

La perspectiva pancrónica adoptada en el presente estudio permite arrojar algo de luz sobre el desenlace de un fenómeno de variación y cambio lingüístico que continúa muy activo en español tras varios siglos de desarrollo y que, a la vista de los datos analizados, parece estar lejos de completarse. Al mismo tiempo, los resultados muestran algunos desenlaces aparentemente contradictorios que resumimos a continuación.

En el plano frecuencial, los datos orales correspondientes a la sincronía actual muestran inicialmente un importante retroceso en el uso de las variantes pronominales (*que, el que, el cual*) en las oraciones de relativo locativas, en beneficio de la forma adverbial *donde*. Este resultado difiere del movimiento en la dirección contraria que habíamos advertido previamente en el análisis de un corpus histórico compuesto por textos de inmediatez comunicativa, entre los siglos XVI y la primera mitad del XX.

Aun así, esta aparente reversión del cambio contrasta con los movimientos que se advierten tanto en el interior de la gramática interna como en el eje extralingüístico. Y es que, contrariamente a lo que es habitual en los procesos de cambios consolidados, en los que el avance de las frecuencias de la variante novedosa va generalmente asociado a un debilitamiento de los factores que la han condicionado en el pasado o el surgimiento de otros condicionantes que no operaban antaño, en el presente caso, el examen detenido del contexto variable revela una robusta persistencia de los mismos predictores, y, además, con la misma dirección explicativa en su interior.

En este sentido, hemos comprobado cómo, en toda la serie histórica (siglos XVI, XVIII, XX y XXI), la selección del pronombre es alentada cuando el antecedente tiene una naturaleza semántica no locativa, mientras que, en el extremo opuesto los referentes más prototípicamente locativos inhiben la variante e impulsan en la misma proporción la elección de *donde*. En el mismo sentido actúa la influencia del *priming* estructural: desde el siglo XVI hasta nuestros días, la presencia de otro pronombre en los alrededores de la variable favorece su empleo, todo lo contrario que cuando ese papel lo desempeña el adverbio *donde*, uno de los escenarios más adversos para la selección de la variante pronominal. También en el plano estructural, la variación se ve condicionada en todo momento por el estatus sintáctico del antecedente, con la función circunstancial como terreno menos propicio para el empleo de los pronombres. Finalmente, aunque en el plano sociolectal las condiciones muestrales entre los corpus histórico y oral no son en muchos casos comparables (debido, sobre todo, a insuficiencias muestrales en el primero), entre los resultados se advierte también una correlación persistente en el tiempo entre el estatus social de los informantes y el uso de *que*, de tal manera que – sobre todo, a partir del siglo XX – cuanto mayor es ese nivel, mayor es también la selección de esta forma. Si acaso, en esta comparativa tan solo cabría destacar algunas diferencias en la magnitud de los efectos que, salvo alguna excepción aislada, traslucen un afianzamiento general de las restricciones lingüísticas y extralingüísticas en el momento actual, en lugar de su debilitamiento.

Este seguimiento de las mismas presiones estructurales y sociolectales en periodos de tiempo tan dilatados choca con escenarios más habituales del cambio lingüístico. Un ejemplo de esto último lo tenemos, por ejemplo, en un fenómeno de variación que afecta también a las oraciones de relativo, y que hemos tenido ocasión de analizar recientemente a partir de materiales similares a los implicados en este estudio (Blas

Arroyo, en prensa). Así, en la alternancia entre las formas con artículo y sin artículo del pronombre *que* en relativas oblicuas ('la casa en *(la) que* vivo'), hemos confirmado el avance de la primera desde cifras reducidas en el siglo XVIII (17%), hasta otras mucho más prominentes en la presente centuria (53%), pasando por algunos hitos intermedios (XIX: 24%; XX: 29%). Con todo, lo más relevante fue comprobar cómo, junto a algunas manifestaciones de persistencia – habituales, por lo demás, en los procesos de cambio –, en el presente estadio de lengua predominan los escenarios de generalización, esto es, de neutralización de las diferencias antaño explicativas, e incluso de cambios en la dirección de los efectos.

¿Cómo interpretar, entonces, los movimientos aparentemente contradictorios que se aprecian en el fenómeno analizado en estas páginas? ¿Cómo es posible que la pronunciada reducción frecuencial vaya de la mano con unos condicionantes que, en esencia, continúan siendo los mismos y que, en el mejor de los casos, no han hecho más que afianzarse? En la tradición variacionista no faltan ejemplos de compatibilidad entre ambos escenarios, como han visto, por ejemplo, Tagliamante y D'Arcy (2007) a propósito del incremento de frecuencias del marcador *be like* en la región canadiense de Ontario en un breve periodo de tiempo (incluso en el habla de los mismos individuos), hecho que, sin embargo, no era incompatible con la mantenimiento en lo esencial de la misma gramática subyacente. Ahora bien, en nuestro caso las razones podrían ser más prosaicas y depender del tipo de materiales sobre los que se asienta la variación, como se desprende de los datos de la tabla 7. En ella se da cuenta de la distribución en los diferentes corpus de los dos principales contextos semánticos en que hemos dividido el predictor más explicativo de todos, la semántica del antecedente pronominal: locativos (agrupados) y no locativos.

Tabla 7: Distribución de los contextos locativos y no locativos en el corpus de cuatro periodos históricos

	XVI		XVIII		XX		XXI	
	%	Σ	%	Σ	%	Σ	%	Σ
Semántica								
Lugar	68	589	57	540	48	327	65	795
Otros	32	280	43	404	52	355	35	423
Total		869		944		682		1218

Como se puede apreciar, hay una singular coincidencia muestral entre los dos puntos extremos del eje temporal. Así, tanto en el siglo XVI como en el XXI, el corpus tiene una sobrerrepresentación de antecedentes locativos (XVI: 68%; XXI: 65%), al tiempo que los no locativos suponen una minoría (XVI: 32%; XXI: 35%). No así en los otros dos periodos, en los que la representación está más equilibrada. ¿Podrían explicar estas diferencias muestrales las contradicciones a las que aludíamos al comienzo de este apartado? Creemos que una respuesta afirmativa a este interrogante es plausible. En primer lugar, hemos visto cómo las frecuencias globales de uso de la variante en estos periodos son singularmente parecidas. Pero también hemos comprobado una notable persistencia del factor semántico, el más relevante de todos los analizados en el corpus, y cuya robustez, lejos de debilitarse se ha fortalecido con el tiempo. Los usos de la forma pronominal en los contextos locativos no solo no se han incrementado, sino que, por el contrario, han disminuido. Por otro lado, tampoco se han movido prácticamente los empleos en la esfera de los sustantivos no locativos, en las que las proporciones continúan siendo singularmente parecidas.

En conclusión, antes que la reversión de un cambio lingüístico, lo que el análisis pancrónico revela es una notable estabilidad del fenómeno de variación a lo largo de más de cinco siglos de historia. Las fluctuaciones frecuenciales pueden enmascarar esta realidad, pero no ocultarla ante un análisis variacionista como el emprendido en estas páginas, en el que se toman en consideración no solo los aspectos más superficiales (frecuencias) de la variación y el cambio lingüístico, sino también otros más profundos.

Agradecimientos

El presente estudio forma parte del Proyecto de investigación ‘Dimensiones estructurales, sociales e idiolectales del cambio lingüístico: nuevas aportaciones desde la sociolingüística al estudio del español’, financiado por el Ministerio de Economía y Competitividad (2018--2021) (Ref. FFI2017--86194–P) y la Universitat Jaume I (Ref. UJI–2017–01). Quisiera agradecer a Carme Barberà, Maria Chiara Marullo, Jordi Ayza y Javier Lara por su colaboración en el Laboratorio de Sociolingüística de la Universitat Jaume I a lo largo de los últimos años. Asimismo, agradezco los perspicaces comentarios de cuatro evaluadores anónimos a una versión previa del artículo. Ni que decir tiene que cualquier error que se detecte en el texto es de mi entera responsabilidad.

Sobre el autor

José Luis Blas Arroyo is professor of Spanish linguistics at the Universitat Jaume I (Castellón, Spain), where he teaches Sociolinguistics and Pragmatics in the Faculty of Arts. His main research areas are devoted to variationist and sociopragmatic topics, as well as to bilingual matters related to the Spanish in contact with other languages. He has published a number of different books (*Sociolingüística del español*, *Sociolingüística histórica del español*, *Lenguas en contacto*, *Políticos en conflict...*), and many articles on these subjects in international journals and research monographs. Since 1998, he leads the ‘Sociolinguistic Laboratory’ at the University Jaume I, and from 2011 to the present a research project on Historical Sociolinguistics of Spanish.

Referencias

- Alcina, J. y Bleca, J. M. (1991 [1975]) *Gramática española*. Barcelona: Ariel (8ª ed.).
- Bergs, A. (2012) The uniformitarian principle and the risk of anachronisms in language and social history. In J. M. Hernández–Campoy y J. M. Conde–Silvestre (eds.) *The Handbook of Historical Sociolinguistics* 80--98. Malden, MA: John Wiley y Sons Ltd.
- Blas Arroyo, J. L. (en prensa) Traces of the past in a lengthy change (still) in progress: persistence and generalization in prepositional relative clauses in Peninsular Spanish. In M. Díaz-Campos (ed.) *Handbook of Variationist Approaches to Spanish*, Abingdon: Taylor and Francis.
- Blas Arroyo, J. L. y Velando, M. (2017) Sobre puntos de la gramática particularmente resistentes al cambio: cuatro (largos) siglos en la evolución de las relativas oblicuas de lugar. Un análisis variacionista. *Borealis. International Journal of Hispanic Linguistics* 6(1): 139--174.
- Brucart, J. M. (1999) La estructura del sintagma nominal: las oraciones de relativo. In: I. Bosque y V. Delmonte (eds.) *Gramática descriptiva de la lengua española* 395--522. Madrid: Espasa-Calpe.
- Cortés, L. (1986) Alternancia de los relativos *donde: que / el que: el cual* en el español hablado. *Revista española de lingüística aplicada* 2: 9--22.
- Fernández Ramírez, S. (1987 [1951]) *Gramática española*. Madrid: Arco / Libros.

- Girón, J. L. (2009) Las oraciones de relativo II. Evolución del relativo compuesto *el que, la que, lo que*. In C. Company (dir.) *Sintaxis histórica de la lengua española. Segunda parte: La frase nominal 1477--1590*. México: UNAM, Fondo de Cultura Económica.
- González Díaz, C. (2006) Uso de *donde* y PAR (*preposición + artículo determinado + relativo que / cual*): análisis en el español escrito de Venezuela y España. In Milka Villayandre (ed.) *Actas del XXXV Simposio Internacional de la Sociedad Española de Lingüística* 754--773. León: Universidad de León.
- González Díaz, C. (2009) Combinatoria sintáctico-semántica de *donde* y otros relativos. *Boletín de Lingüística* 21: 25--46.
- Herrera Santana, J. L. (1994--95) *Estudio sociolingüístico de los relativos en el español de Santa Cruz de Tenerife*. La Laguna: Universidad de La Laguna (Tesis doctoral).
- Herrero Ruiz de Loizaga, F. J. (2005) *Sintaxis histórica de la oración compuesta*. Madrid: Gredos.
- Johnson, D. E. (2009) Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1): 359--383.
- Labov, W. (2001) *Principles of linguistic change. Social factors*. Oxford: Blackwell.
- Levshina, N. (2015) *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins.
- Mendoza, J. F. (1984) Sintaxis de los relativos en el habla popular de la ciudad de México. *Anuario de Letras* 22: 65--77.
- Oesterreicher, W. (2004) Textos entre inmediatez y distancia comunicativas: el problema de lo hablado escrito en el Siglo de Oro. In R. Cano (coord.), *Historia de la lengua española* 729--769. Barcelona: Ariel.
- Palacios de Sámano, M. (1983) *Sintaxis de los relativos en el habla culta de la ciudad de México*. México: UNAM.
- RAE (Real Academia Española) (1973) *Esbozo de una nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- RAE (Real Academia Española) y ASELE (Asociación de las Academias de la lengua española) (2009) *Nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- Rascón, M.^a F. (2005) *Estudio de la variación sociolingüística de los relativos en una red social de Baeza (Jaén)*. Granada: Universidad de Granada (Tesis doctoral).
- Tagliamonte, S. A. y D'Arcy, A. (2007) Frequency and variation in the community grammar: Tracking a new change through the generations, *Language Variation and Change* 19(2): 1--19.
- Tagliamonte, S. y Baayen, H. R. (2012) Models, forests and trees of York English: *Was / were* variation as a case study for statistical practice, *Language Variation and Change* 24(2): 135--178.

Notas

¹ Durante largos periodos, la forma más frecuente fue *que*, en alternancia con el relativo compuesto *el cual*, gramaticalizado desde finales del siglo XV. Con todo, a partir del siglo XVIII, ambas formas comenzaron a competir con el mismo pronombre precedido de artículo (*el que*) (Girón 2009, Blas Arroyo y Velando 2017).

² Formas pronominales precedidas de preposición.

³ El corpus se completa con una representación menor de otros géneros autobiográficos, como diarios y memorias.

⁴ El lector interesado puede obtener la relación completa de las obras que integran el corpus en la siguiente dirección electrónica: <http://sociolingüisticawe.wix.com/sociolingüisticauji>.

⁵ Aun así, quisiera agradecer muy sinceramente a los coordinadores de esos proyectos por su amabilidad al facilitarme el acceso a los materiales necesarios para el estudio.

⁶ Estas cifras contrastan de nuevo con las ofrecidas por algunos autores a propósito de otras variedades meridionales, tanto andaluzas (Cortés 1986, Rascón 2005) como canarias (Herrera Santana 1994--95).

⁷ Preferentemente *el que* ($n= 242$; 20%) y, en menor medida, *que* ($n= 98$; 8%) y *el cual* ($n=70$; 6%).

⁸ Esta cifra corresponde al peso probabilístico obtenido en el análisis en que tomamos con variable aleatoria el lema del antecedente. La tabla 5 ofrece también los pesos correspondientes a otros dos análisis, los realizados con las variables *Informante* y *Comunidad*, respectivamente, y que, como se puede observar, presentan magnitudes similares. Con todo, a los efectos de una necesaria economía expositiva, en lo que sigue mencionamos tan solo los datos correspondientes al lema del antecedente.

⁹ Para la codificación de este factor, consideramos la aparición de estas formas en las 25 palabras anteriores a la variante analizada, contadas a partir del texto de la transcripción.

¹⁰ Como se recordará (ver tabla 3), para el estudio de este predictor restringimos el análisis a las comunidades de habla valencianas (Castellón y Valencia), por lo que hemos llevado a cabo un segundo análisis de regresión, en el que tan solo se consideran sus datos. Por otro lado, recordamos que para la operacionalización de este factor tomamos en consideración la respuesta de los informantes a una pequeña encuesta sociológica, al final de la entrevista.

¹¹ En este caso, debemos descartar el lema del antecedente, dado que el elevado número de factores – 349 palabras distintos – haría imposible el dibujo de un árbol de estas características.

¹² En el mismo sentido apuntan las diferencias relacionadas con el *tenor* de los escritos, de tal manera que tanto en el siglo XVI como en el XVIII y en el XX, los pronombres han encontrado un mejor acomodo en los textos más distantes y formales, frente a la mayor difusión del adverbio *donde* en los escritos más familiares y personales. Una nueva prueba, en suma, que vendría a desmentir la naturaleza más espontánea de los primeros.