

Círculo de Estudios Lingüísticos Hispánicos de Tokio. 2024-10-26

Posición del pronombre átono español en variación histórica, geográfica y social

Problemas y métodos de análisis de datos lingüísticos

Hiroto Ueda

(2024)

enclisis-es 2024-11-11

1. Introducción¹

Este estudio aborda los cambios y variaciones bajo los parámetros del tiempo, el espacio y la sociedad de las cuatro relaciones posicionales entre pronombres átonos y verbos en español:

- (1) <pronombre + verbo conjugado> (p + vc: *lo+tiene*)
- (2) <verbo conjugado - pronombre> (vc-p: *tiene-lo*)
- (3) <verbo no conjugado - pronombre> (vn-p: *tener-lo*)
- (4) <pronombre + verbo no conjugado> (p+vn: *lo+tener*)
- (5) <verbo imperativo - pronombre> (vi-p: *ten-lo*)

De esta manera, se muestran ejemplos que utilizan *tener* y *lo* para todas las relaciones morfológicas y posicionales (*lo+tiene*, *tiene-lo*, *ten-lo*, *tener-lo*, *lo+tener*). Dado que solo tratamos con pronombres átonos, los *pronombres débiles* se denominan simplemente "pronombres". Los verbos son conjugados (vc), no conjugados (vn) o imperativos (vi). Los imperativos son afirmativos y no incluyen los negativos que utilizan formas subjuntivas. El grupo de las formas no conjugadas incluyen el infinitivo (if), gerundio (g) y participio pasado (pv: participio verbal) e imperativo positivo (ip: imperativo), por la coincidencia sintáctica en la combinación con el pronombre. Cuando un pronombre precede a un pronombre, agregamos un '+' antes del verbo y cuando un pronombre le sigue, lo conectamos con un '-', por considerar que el pronombre pospuesto forma parte de una palabra como posteriormente discutiremos.

En español moderno, los pronombres se colocan delante del verbo conjugado (*le miraba*, *se detuvo*), pero históricamente con cierta frecuencia se han colocado después del verbo conjugado, por ejemplo, *mirábale*, *detúvose*. Cuando se coloca antes y aparte del verbo, se llama 'proclisis', y cuando se coloca inmediatamente después del verbo, se llama 'enclisis'. Por ejemplo, la enclisis aparece con cierta frecuencia en las obras de Benito Pérez Galdós (1843-1920), al lado del uso normal de la proclisis. La aparición de proclisis y enclisis está determinada principalmente por las condiciones intralingüísticas -la forma del verbo (conjugado / no conjugado / imperativo) y la posición dentro de la oración (inicial y medial de la oración)- y las extralingüísticas de año, región y sociedad.

¹ Agradezco de todo corazón la ayuda prestada por Inmaculada Martínez en la preparación de este estudio.

Lo que el lector encontrará a continuación (capítulo 2) es el abordaje a la posición de los pronombres cuando se toman en consideración las condiciones lingüísticas y extralingüísticas mencionadas. Utilizamos los datos ofrecidos por el corpus CODEA² y en todo momento se sigue a Ueda (2024). Por último, es preciso señalar que los detalles de los métodos estadísticos se discuten en la sección 6. Apéndice.

2. Método

Las citas siguientes son de Hanssen (1913), Menéndez Pidal (1976, 1980), Lapesa (1982), Nakaoka (1993) y Penny (2006), donde se menciona el orden de <verbo conjugado-pronombre> (*tiene-lo*) predominante en el español medieval. Algunos autores afirman que no existe <pronombre + verbo conjugado> (*lo+tiene*), en la posición inicial de oración ni detrás de conjunciones *e* (*et*) y *mas* (subrayado [...]: autor).

Hanssen (1913)

En antiguo castellano, los pronombres inacentuados no pueden ocupar el primer lugar de la frase: se dice *díxome*, no se puede decir *me dixo*. Esta regla vale para todas las lenguas neolatinas, pero las lenguas modernas no la observan. (194)

Origen de la Regla. (...) el pronombre inacentuado se apoyaba en latín bíblico enclíticamente en la palabra que precedía: *ille me videt, videt me*. (197)

Menéndez Pidal (1976)

El pronombre personal átono es generalmente enclítico. (402)

Si el verbo encabeza la proposición, o va precedido de la conjunción *e* o *mas*, el pronombre se le pospone, como sucedía aún en el siglo XVI. (403)

Si al verbo preceden otras partes del discurso, el pronombre se le antepone, apoyándose en la palabra que precede inmediatamente al verbo. (404)

² <Corpus de documentos españoles anteriores a 1900>. Grupo de Investigación Textos para la Historia del Español [GITHE]: CODEA+2022 (Corpus de documentos españoles anteriores a 1900) [en línea], doi: <https://doi.org/10.37536/CODEA.2015> [15 de septiembre de 2014]

Menéndez Pidal (1980)

El pronombre átono va después del verbo, si este encabeza la frase o va precedido solo de la conjunción *et*: *faca nos Deus...* GlEmil 89; *kate nos* 129; *et ouisti me a dare centu solidos* 1092 León. (379)

Lapesa (1982)

El pronombre átono, esencialmente enclítico entonces [en español arcaico], no podía colocarse ante el verbo después de pausa, ni cuando precedieran sólo las conjunciones *e* o *mas* [...] (218)

Nakaoka (1993: 35, traducción nuestra)

En el español medieval, la forma débil del pronombre personal era un enclítico. Por lo tanto, se colocó inmediatamente después de la pausa y no se podía colocar al comienzo de una oración o frase. En otras palabras, el orden moderno de las palabras "Lo vi todo" no existía en el lenguaje medieval.

En el español medieval, la posición del pronombre personal débil es la siguiente:

(A) Colocado inmediatamente después del verbo cuando aparece al principio de una oración o inmediatamente después de una pausa. *Partios' de la puerta* [...] (Cid 51) / *Armaronse e tañeron vna trompeta. Al son della, oyéronlo* *los de la gran galera* [...] ([F. González Ollé, *Lengua y literatura españolas medievales, textos y glosario*. Ariel. Barcelona, 1980], 318; Victorial)

(B) Cuando una palabra acentuada precede a un verbo, se coloca antes del verbo. [...] *Decidme uostros nombres, nom los querades celar* [...] (*Auto RM*. 81)

Penny (2006)

En español medieval, los pronombres personales átonos eran normalmente enclíticos [...]. (163) [...] el pronombre aparecía tras el verbo (conjugado o no), a no ser que este fuera precedido en la misma cláusula por otra palabra tónica [...]: *ascóndense de mio Cid, pero non lo desafié, aquel que gela diesse*. (163)

Las investigaciones previas sobre la historia del español parecen haber tratado las obras literarias, pero el lenguaje literario a menudo sigue antiguas y tradicionales normas. Por otro lado, nuestra investigación trata documentos notariales que reflejan la realidad de sociedades lingüísticas pasadas. ¿Podría ser que <pronombre + verbo> (*lo + tiene*) estuviera limitado bajo ciertas condiciones

en los documentos notariales del lenguaje medieval, o es posible que la misma forma se usara básicamente en muchos casos al igual que en los tiempos modernos? Esta investigación utiliza una gran cantidad de materiales no literarios para aclarar esta cuestión.

Aquí, el <verbo-pronombre> se llama "enclisis del pronombre". En otras palabras, generalmente se acepta que un pronombre sigue a un verbo, pero en el período moderno temprano (1500-), se dice que se convierte en un <pronombre+verbo> (proclisis del pronombre). Por tanto, se supone el siguiente cambio histórico:

Cambio histórico: [Edad Media] <Enclisis> → [Edad Moderna] <Proclisis>

En cuanto a la posición de los pronombres en la lengua medieval, Penny (2006) no distingue entre las formas de los verbos conjugados y los no conjugados. En cambio, este estudio aborda toda la historia y distingue entre ambos tipos de verbos. El grupo de las <formas no conjugadas> incluye infinitivos, gerundios y participios pasados. La forma imperativa positiva es una forma de flexión dentro del sistema morfológico, pero a nivel de morfología, las formas irregulares especiales se utilizan solo en la forma imperativa positiva y no en la forma imperativa negativa (*ven, no vengas; vete, no te vayas*, etc.). Además, en términos de sintaxis (relación posicional con los pronombres), como veremos a continuación, muestra los mismos cambios y variaciones que las formas no flexionadas (infinitivo, participio presente, participio pasado).

En primer lugar, establecemos la siguiente hipótesis³:

Hipótesis: desde la Edad Media hasta los primeros tiempos modernos y hasta la actualidad, <pronombre + verbo conjugado> era el orden no marcado de las palabras. El orden de <verbo conjugado - pronombre>, heredado del latín, es un orden marcado de palabras y ha sobrevivido hasta hace poco en determinadas condiciones. Por otro lado, el orden de <verbo no conjugado - pronombre> (*tener-lo*) se ha desarrollado a manera diferente.

Al verificar la validez de esta hipótesis, intentaremos resolver las siguientes preguntas de investigación:

³ Ésta es la conclusión a la que se llegó en este estudio, pero creemos que el siguiente desarrollo quedará más claro si la misma conclusión se presenta al principio como una hipótesis.

Pregunta-1: ¿En las lenguas medievales (fuentes textuales: 1200-1500), la enclisis (*tiene-lo*) era más común y la proclisis (*lo+tiene*) se limitaba a ciertas condiciones?

Pregunta-2: ¿Ha habido un cambio en la conjugación de los verbos de enclisis (*tiene-lo*) a proclisis (*lo+tiene*) a lo largo de la historia (en los documentos de 1200-1800)?

Pregunta-3: o, por el contrario, ¿ha habido un cambio en el verbo no conjugado de los verbos a lo largo de la historia, de proclisis (*lo+tener*) a enclisis (*tener-lo*)?

Pregunta-4: ¿Cuál es la razón histórica del imperativo positivo enclítico (*ten-lo*), a diferencia del imperativo negativo?

Las siguientes son las combinaciones más frecuentes y sus frecuencias totales:

1. Pronombre + verbo conjugado (*lo+tiene*): se contiene (422), se á (340), lo firmó (302), se ha (213), me obligo (179), se an (162), la fiz (155), me á (154), se llama (153), la confirma (148), vos damos (141), se han (116), se halla (110), se puede (109), le dio (102), se dio (101), se haga (100), se sirva (98), nos obligamos (97), se pueda (97), ... [Total: 35,986].

2. Verbo conjugado + pronombre (*tiene-lo*): tóvelo (47), dámosvos (41), dóvos (40), llámase (37), danle (35), tovímoslo (35), obligome (34), firmólo (33), fiole (29), pidióme (29), fuéle (28), leyósele (25), áse (24), confírmoles (22), confirmámoslo (22), obligámosnos (18), abrióse (15), sépase (15), otorgámosles (15), entróse (14), ... [Total: 2,135].

3. Verbo no conjugado + pronombre (*tener-lo*): darle (76), averse (63), hallarse (43), haviéndose (40), haverse (39), hallándose (38), aviéndose (35), llamarse (35), darme (33), mandarme (33), dándole (29), hacerlo (29), diciéndola (25), siéndolo (25), serbirte (24), haberse (23), dándoles (22), averle (21), darla (21), siéndole (21)... [Total: 5,842].

4. Pronombre + verbo no conjugado (*lo+tener*): lo fazer (53), vos fazer (53), les fazer (44), vos dar (37), les ir (30), lo vender (22), los tener (21), la fazer (20), vos defender (18), lo dar (16), los dar (16), lo aver (15), nos ayuntar (15), gela quebrantar (13), les passar (12), lo tener (12), le dar (11), les dar (11), lo cumplir (11), vos redrar (11)... [Total: 1,536].

5. Verbo imperativo + pronombre (*ten-lo*): dátgela (19), quédate (13), dále (12), sírvase (12), dádgela (11), díme (7), sírbase (7), déle (6), póngase (6), dálos (5), dásele (5), llámese (4), quédese (4), véase (4), diémosgelo (3), dáme (3), dénos (3), dígame (3), díle (3), llámenle (3)... [Total: 303].

Las siguientes abreviaturas se utilizan para categorías gramaticales:⁴

a: adjetivo, ab: abreviatura, ar: artículo, av: adverbio, c: conjunción, cj: conjugado, cl: clítico (pronombre átono), cl.cl: clítico+clítico, d: demostrativo, df: definido, g: gerundio, id: indefinido, if: infinitivo, ij: interjección, ip: imperativo, ir: interrogativo, n: nombre, nombre, nm: numeral, np: nombre de persona, p: preposición, pa: participio adjetival, pn: pronombre tónico, ps: personal, psa: posesivo átono, pst: posesivo tónico, pt: puntuación, pv: participio verbal, r: relativo, t: topónimo, v: verbo.

A continuación, mostramos los resultados del análisis equilibrado (\rightarrow 6) realizado en el orden anterior de (1) a (4), donde utilizamos una combinación de las siguientes variables. Los elementos anteriores de '<' son las variables objetivas y los posteriores, la variable de referencia:

(a) R:A<S: La región (R) y el año (A) son variables objetivas y la sociedad (S) es la variable de referencia.

(b) S:A<R: La sociedad (S) y el año (A) son variables objetivas y la región (R) es la variable de referencia.

En los puntos del gráfico lineal, el grado de significatividad se indica mediante colores según los siguientes criterios (s: grado de significatividad) (\rightarrow 6. Apéndice).

- 1.azul: $s > 0.99$ (significativamente muy grande)
- 2.azul claro: $0.95 < s \leq 0.99$ (significativamente grande)
- 3.blanco: $0.05 \leq s \leq 0.95$ (no significativo)
- 4.rojo claro: $0.01 \leq s < 0.05$ (significativamente pequeño)
- 5.rojo: $s < 0.01$ (significativamente muy pequeño)

El "análisis equilibrado" cruza la frecuencia combinada de dos variables objetivas con la variable de referencia ("variable de referencia") y calcula el

⁴ En este artículo, p indica un pronombre, y en los materiales, indica una preposición. cl (clítico) en el documento se reemplaza por p en el texto.

promedio ("media mayoritaria) de las dos variables objetivas descompuestas (→ 6. Apéndice).

Todos los números se ajustan al valor promedio de las mediciones reales (frecuencia absoluta de aparición), por lo que están en la misma escala que la frecuencia de aparición.

En el último corpus histórico (4023 documentos) de *CODEA* (<Corpus de documentos Españoles Anteriores a 1900>) hay documentos reducidos en *CODEA* hasta el 1100, por lo que nos centraremos en los documentos del 1200 al 1800: a1200, a1300, ..., donde el número representa el año del inicio; por ejemplo, a1200 incluye los años desde: 1200 hasta 1299.

Las variaciones regionales se clasifican en demarcaciones tradicionales⁵: AN: Andalucía, AR: Aragón, CN: Castilla la Nueva, CV: Castilla la Vieja, EX: Extremadura, LE: León.

⁵ León, Castilla la Nueva, Castilla la Vieja son los antiguos nombres de regiones administrativas utilizados hasta 1983. Castilla la Nueva incluye principalmente la actual comunidad autónoma de Castilla La Mancha y Madrid. Castilla la Vieja y León corresponden a la actual comunidad autónoma de Castilla y León. Los nombres regionales antiguos son más adecuados para investigar la dialectología histórica española que los que designan las comunidades autónomas actuales.

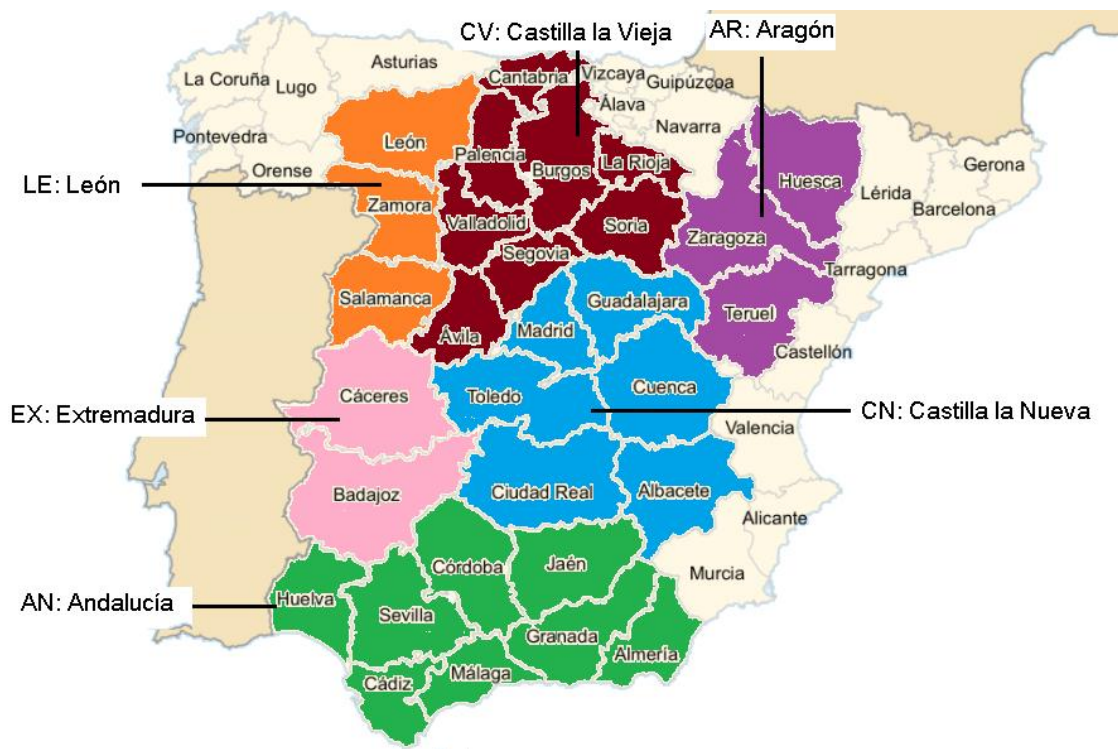


Fig. 1. Demarcaciones tradicionales

Las clasificaciones sociales (en torno a los documentos) son estas:

- C: documento canclleresco
- E: documento eclesiástico
- J: documento judicial
- M: documento municipal
- P: documento particular.

3. Análisis

3.1. Pronombre + verbo conjugado (*lo+tiene*)

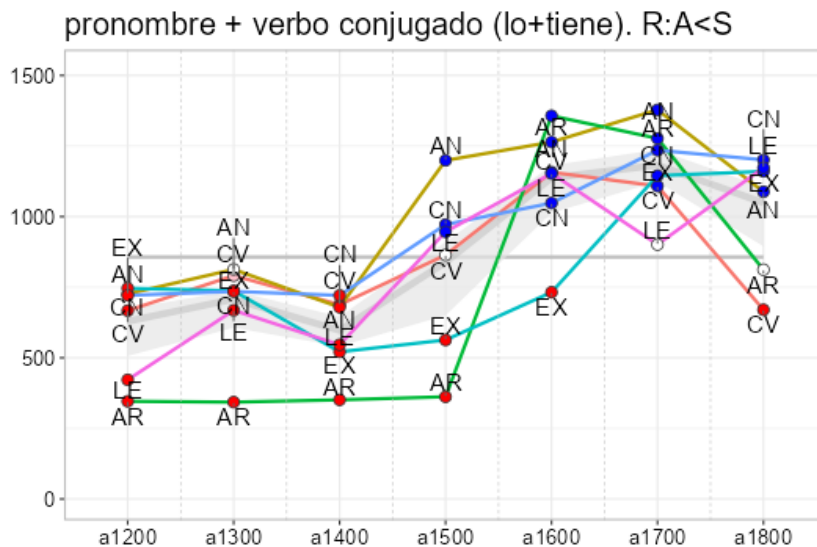
Observación

En primer lugar, observaremos la forma <pronombre + verbo conjugado> (*lo+tiene*) que es la más frecuente de todas. Realizamos el análisis equilibrado-1 (variables objetivas = región:edad, variable de referencia = sociedad) y el análisis de equilibrio-2 (variables objetivas = sociedad:edad, variable de referencia = región). Para la visualización de datos, preparamos un gráfico de líneas que trate con los cambios a lo largo del tiempo y un gráfico de variación que muestre la

distribución de valores numéricos. En el gráfico de líneas, el valor promedio general se muestra como una línea horizontal gris, el valor de la media mayoritaria para cada año (→ 6. Apéndice) se muestra como una línea gris oscura y, por último, se muestra el área con color gris claro para los promedios de desviaciones superior e inferior.

Formas correspondientes (en orden de frecuencia): se contiene (frecuencia: 422), se á (340), lo firmó (302), se ha (213), me obligo (179), se an (162), la fiz (155), me á (154), se llama (153), la confirma (148), vos damos (141), se han (116), se halla (110), se puede (109), le dio (102), se dio (101), se haga (100), se sirva (98), nos obligamos (97), se pueda (97)... [Total: 35,986].

Análisis equilibrado-1: variables objetivas = región:año, referencia=sociedad.



pronombre + verbo conjugado (lo+tiene). R:A<S

	a1200	a1300	a1400	a1500	a1600	a1700	a1800
CV	667 -0.293	791 -0.023	688 -0.248	864 0.108	1157 0.617	1108 0.532	670 -0.287
AN	726 -0.165	813 0.020	679 -0.267	1199 0.690	1264 0.802	1378 1.000	1088 0.497
AR	346 -0.993	343 -1.000	351 -0.983	362 -0.959	1357 0.964	1277 0.825	812 0.018
EX	746 -0.121	737 -0.141	521 -0.612	563 -0.520	732 -0.152	1145 0.596	1160 0.622
CN	721 -0.176	735 -0.145	721 -0.176	972 0.296	1048 0.428	1235 0.752	1201 0.693
LE	422 -0.828	668 -0.291	546 -0.557	946 0.251	1153 0.610	901 0.173	1173 0.644

Fig.2a, b. *Pronombre + verbo conjugado. Región:Año<Sociedad.(Media: 857)*

(AN: Andalucía, AR: Aragón, CN: Castilla la Nueva, CV: Castilla la Vieja, EX: Extremadura, LE: León)

En general, se presentan altas frecuencias. La literatura anterior ha señalado que, en la Edad Media, <verbo conjugado + pronombre> (tiene-lo) se usaba comúnmente, y <pronombre + verbo conjugado> (*lo + tiene*) era relativamente raro; sin embargo, en este estudio, como muestran las figuras anteriores, se ha registrado un número bastante grande de <pronombre + verbo conjugado> (*lo + tiene*). Como se explicará más adelante, <pronombre + verbo conjugado> (*lo+tiene*) se han registrado abrumadoramente a lo largo de la historia

(→3.6). Sin embargo, si solo nos fijamos en <pronombre + verbo conjugado>, es seguro que habría relativamente pocos en la época medieval en comparación con la época moderna temprana. Por tanto, la posposición de pronombres puede reconocerse como una tendencia general (no una regla definitiva) en la Edad Media.

Análisis equilibrado-2: variables objetivas = sociedad:año, referencia=región.

Veamos el resultado del análisis equilibrado de Sociedad y Año, con referencia de la Región (S:A<R):

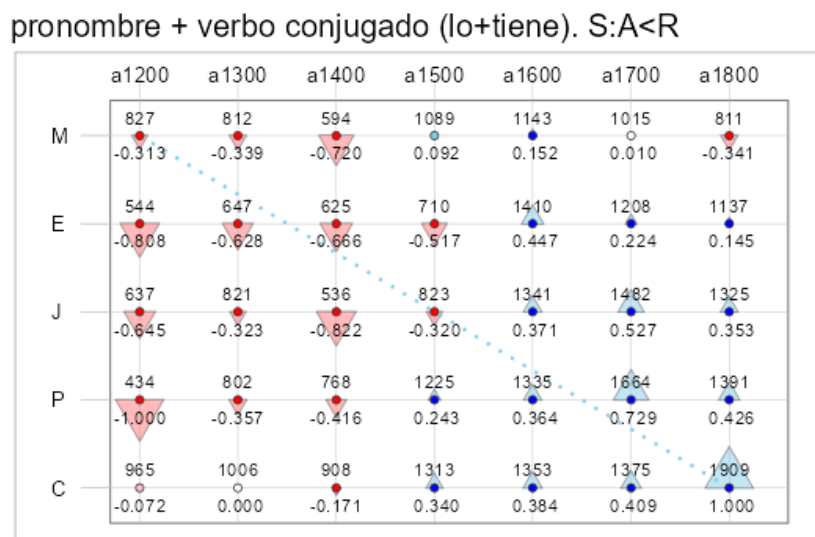
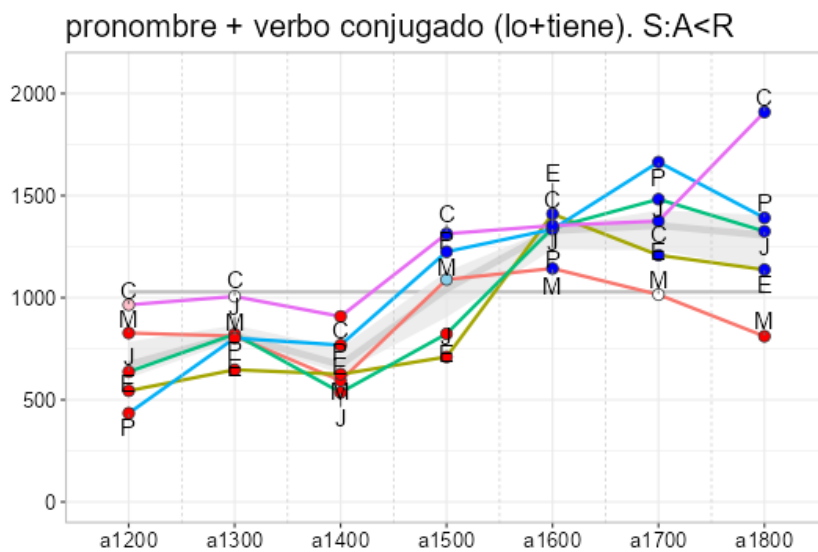


Fig.3a, b. *Pronombre + verbo conjugado. Sociedad:Año<Región..(Media:1028)*

(C: documento cancilleresco, E: d. eclesiástico, J: d. judicial, M: d. municipal, P: d. particular)

Se observan las variables sociales con la misma tendencia que las variables

regionales. El uso de pronombres es relativamente bajo en la Edad Media en general, pero significativamente más común en los documentos cancillerescos (C). Después de a1500, hubo un aumento general.

Consideración

(1) Palabra que precede a <pronombre + verbo conjugado>

A continuación, calculamos el valor contrastivo (r) de <pronombre + verbo conjugado> (*lo + tiene*) (y) comparando las condiciones (x): a comienzo de oración (o después de una pausa) (#) y en medio de la oración (=). El resultado es el que aparece en la Tabla 1. El número total de palabras es la suma de las combinaciones de pronombres y verbos relacionados (los datos se extraen de los dos órdenes de palabras *lo+tiene* y *tiene-lo* de a1200-a1400, total: 10955).

s. x	s. y	x	y	a	b	c	d	s	r
#	c1, v_cj	36	9,819	19	17	9,800	1,119	10,955	0.056
=	c1, v_cj	10,919	9,819	9,800	1,119	19	17	10,955	0.795

Tabla 1. *Relación entre el comienzo de una oración (#), la posición medial de una oración (=) y <pronombre + verbo conjugado>*

Este resultado muestra que es relativamente menos probable que <pronombre + verbo> (*lo+tiene*) aparezca cuando la condición está al principio de una oración (#), pero es más probable que aparezca cuando la condición está en medio de una oración (=). Sin embargo, el orden de las palabras <pronombre + verbo> (*lo+tiene*) era aún más común al comienzo de las oraciones (=19) que el orden inverso (*tiene-lo*, =17).

La siguiente es una lista de <pronombres + verbo conjugado> (*lo+tiene*) que aparecen al principio de una oración (o después de una pausa) (#).

<pronombre + verbo conjugado>: # l'ayades, # l'dé, # la complierdes, # la daremos, # la fiz, # La fiz, # la juró, # la signe, # le dio, # le fiz, # lo fiz, # lo vendiemos, # lo verán, # m'confirma, # m'diestes, # me parto, # se hallavan, # Vos rogamos. [Total: 19]

El número de esta forma (a = 19) es mayor que el número de formas de <verbo conjugado - pronombre> (*tiene-lo*) que aparecen en la misma posición (al principio de una oración o después de una pausa) (b=17). Esto es significativamente diferente del punto de vista de investigaciones anteriores ("los pronombres no aparecen antes de los verbos al comienzo de una oración o después

de una pausa": →2).

<verbo conjugado - pronombre>: # Agola, # Áse, # concertámoslo, # Danles, # Fágovos, # mandámosla, # Mándome, # obligome, # Vendémosle, # vendémosvos, # Vendiónos, # Vendívos, # Véndovollo, # Véndovos. [Total:17]

A continuación, examinaremos si el punto de Nakaoka (1993: 35) con respecto al lenguaje literario, "cuando una palabra acentuada precede a un verbo, se coloca antes del verbo" se aplica al documento notarial que aquí se trata.

s. x	s. y	x	y	a	b	c	d	s	r
a	cl, v_cj	7, 111	9, 819	6, 330	781	3, 489	355	10, 955	0. 780
t	cl, v_cj	3, 844	9, 819	3, 489	355	6, 330	781	10, 955	0. 815

Tabla 2. *Relación entre palabras antecedentes débiles (s.x=a) y acentuadas (s.x=t) y <pronombre + verbo conjugado>*

Es cierto que cuando la palabra antecedente es tónica (s.x=t), suele ir precedida de un pronombre (a:6,330). Sin embargo, a la inversa, incluso cuando la palabra antecedente es átona (s.x=a), suele ir precedida por un pronombre también (a:3,489). El valor contrastivo (r) es ligeramente mayor cuando la palabra antecedente es una palabra acentuada (s.x=t), pero la diferencia es pequeña. Por lo tanto, a partir de este material se puede decir que, de hecho, los pronombres a menudo se colocaban delante de ambas oraciones, independientemente de si la palabra antecedente era débil o acentuada.

A continuación, examinamos la categoría gramatical de la palabra anterior como condición (s.x). Las palabras anteriores recogen palabras de alta frecuencia⁶:

s. x	s. y	x	y	a	b	c	d	s	r
av_ir	cl, lv_cj	93	9, 819	93	0	9, 726	1, 136	10, 955	1. 000
pn_id	cl, lv_cj	56	9, 819	56	0	9, 763	1, 136	10, 955	1. 000
av_r_T	cl, lv_cj	35	9, 819	35	0	9, 784	1, 136	10, 955	1. 000
pn_r	cl, lv_cj	1, 797	9, 819	1, 795	2	8, 024	1, 134	10, 955	0. 998
av_r	cl, lv_cj	150	9, 819	149	1	9, 670	1, 135	10, 955	0. 987
av	cl, lv_cj	884	9, 819	859	25	8, 960	1, 111	10, 955	0. 943
pn_d	cl, lv_cj	162	9, 819	151	11	9, 668	1, 125	10, 955	0. 864
v_if	cl, lv_cj	145	9, 819	135	10	9, 684	1, 126	10, 955	0. 862
v_cj	cl, lv_cj	159	9, 819	146	13	9, 673	1, 123	10, 955	0. 836
n	cl, lv_cj	1, 116	9, 819	996	120	8, 823	1, 016	10, 955	0. 785
np	cl, lv_cj	279	9, 819	248	31	9, 571	1, 105	10, 955	0. 778
pn_ps	cl, lv_cj	476	9, 819	412	64	9, 407	1, 072	10, 955	0. 731

⁶ Esto se debe a que incluso si una palabra de baja frecuencia tiene un valor contrastivo alto, su significado es bajo.

c	cl, 1v_cj	3, 884	9, 819	3, 294	590	6, 525	546	10, 955	0. 696
a	cl, 1v_cj	162	9, 819	134	28	9, 685	1, 108	10, 955	0. 654

Tabla 2. *Relación entre la categoría gramatical de la palabra anterior y <pronombre + verbo conjugado>*

De esta manera, muchas categorías gramaticales eligen <pronombre + verbo conjugado> (*lo+tiene*) y evitan <verbo conjugado - pronombre> (*tiene-lo*). Como se discutirá más adelante, dado que no existe ninguna parte del discurso que seleccione preferentemente <verbo conjugado - pronombre> como forma subsiguiente, se puede decir que el orden de las palabras <pronombre + verbo conjugado> fue el general, sin ninguna distinción particular entre las categorías gramaticales. Como vimos anteriormente (→ 2.) y como se explicará más adelante (→4.1.), el orden de las palabras <pronombre + verbo conjugado> es el orden más común entre los cuatro tipos de órdenes de palabras, por lo que no es marcado. Si se trata de un término no marcado, puede que no tenga mucho sentido buscar las condiciones bajo las cuales aparece el orden de las palabras <pronombre + verbo conjugado>. Más bien, debemos prestar atención a la posición al principio de la palabra, donde es poco probable que aparezca el término no marcado <pronombre + verbo conjugado>.

Menéndez Pidal (1976: 403) y Lapesa (1982: 218) toman formas verbales específicas *e* y *mas* y afirman que inmediatamente después de ellas, el pronombre viene después del verbo. La lista de categorías gramaticales que se mostró anteriormente incluye pronombres relativos y conjunciones, por lo que seleccionamos las palabras relativamente frecuentes, *como*, *do*, *e*, *mas*, *porque*, *que* y *si*, y examinamos la relación entre estas palabras y el orden de <pronombre + verbo conjugado>. Encontramos los siguientes resultados:

	s. x	s. y	x	y	a	b	c	d	s	r
como	cl, 1v_cj		77	9, 819	77	0	9, 742	1, 136	10, 955	1. 000
si	cl, 1v_cj		122	9, 819	122	0	9, 697	1, 136	10, 955	1. 000
do	cl, 1v_cj		46	9, 819	46	0	9, 773	1, 136	10, 955	1. 000
porque	cl, 1v_cj		58	9, 819	58	0	9, 761	1, 136	10, 955	1. 000
que	cl, 1v_cj	3, 424	9, 819	3, 421	3	6, 398	1, 133	10, 955	0. 998	
e	cl, 1v_cj	1, 504	9, 819	946	558	8, 873	578	10, 955	0. 258	
mas	cl, 1v_cj		8	9, 819	0	8	9, 819	1, 128	10, 955	-1. 000

Tabla 3. *Relación entre la palabra anterior y <pronombre + verbo conjugado>*

De esta manera, la posición después de *mas* excluye <pronombre + verbo conjugado>, por lo que los resultados de la búsqueda anterior son consistentes con las opiniones de Menéndez Pidal (*ibid.*) y Lapesa (*ibid.*). Sin embargo, existen

muchas excepciones (a=946) con respecto a la conjunción *e*, por lo que no coinciden la opinión de ambos autores y el resultado de nuestra búsqueda.

Las conjunciones subordinantes (*como, si, do, porque, que*) forman oraciones subordinadas. El detallado estudio de Castillo Lluch (1996: 142-197) se refiere al hecho de que los pronombres se colocan antes de los verbos en la mayoría de los casos de oraciones subordinadas en los textos medievales, lo cual es acorde con la preferencia casi absoluta de las conjunciones subordinadas hacia el orden no marcado de <pronombre + verbo conjugado>.

(2) Sintagma nominal

En esta sección (la sección 3.1), vimos que el <pronombre + verbo conjugado> (*lo + tiene*) se usaba con frecuencia. Como se discutirá más adelante (→3.6), esta forma era abrumadoramente más frecuente que el <verbo conjugado - pronombre> (*tiene-lo*).

La razón por la cual la estructura <pronombre + verbo conjugado> (*lo + tiene*) representó la mayoría es que el orden de <palabra átona + palabra tónica> se ve a menudo en todo el sistema del idioma español. En otras palabras, la secuencia <palabra débil + palabra acentuada> se encuentra en <artículo definido / posesivo + nombre / adjetivo>, <preposición + nombre / adjetivo>, <conjunción + nombre / adjetivo / verbo / adverbio> y <relativo + nombre / adjetivo>, etc.

Entre estas <palabras inacentuadas + palabras acentuadas>, los artículos definidos y los posesivos, en particular, son morfológicamente similares a los pronombres: *lo, la, los, las, mi-me, tu-te, su-se*. Por ello, parece que la forma (*lo+tiene*) se ha convertido en la preferida mayoritariamente. Además, <pronombre (paradigma) + conjugación (flexión)> y <artículo definido (paradigma) / posesivo (paradigma) + nombre (flexión)> tienen en común la presencia del cambio morfológico (paradigma y flexión).

Además, la forma de *-s* frecuente en los elementos flexivos de ambos es común (el plural de los nombres y la segunda persona singular de los verbos). Las estructuras <pronombre + verbo conjugado> y <artículo definido / posesivo + nombre (flexión)> comparten el patrón de <palabra funcional + palabra de contenido>. En ambos casos, las palabras funcionales limitan el significado de las palabras de contenido.

En cuanto al significado, por ejemplo, los sintagmas nominales - *tu gusto*,

su casa- y los sintagmas verbales *-te gusta, se casa-* tienen funciones gramaticales diferentes, pero son semánticamente paralelos. Si no reconocemos la relación entre ambos, esto significaría que sus formas y significados fueran casualmente similares, pero esto es inconcebible. Este parecido deriva históricamente del latín y de una posible raíz común del protoindoeuropeo. Este modelo, así, se transmitió de manera continua al castellano medieval, donde sus hablantes utilizarían el idioma basándose en principios subyacentes comunes sin ser particularmente conscientes de las categorías gramaticales de sintagmas nominales y verbales según la gramática normativa moderna.

Aquí, conviene señalar que <pronombre + verbo conjugado> es morfológica, estructural y semánticamente similar a <palabra débil + palabra acentuada> y especialmente a <artículo definido / posesivo + nombre> como sigue (Ueda 2024):

Sintagma	Artículo definido / posesivo	Nombre
nominal	[paradigma: el, los, la, las, lo; mi, tu, nuestro, vuestro, su]	[Flexión: -s]
Sintagma	Pronombre	Verbo conjugado
verbal	[paradigma: lo, los, la, las, le, les, me, te, nos, os, se]	[Flexión: -s]

En las palabras átonas, incluimos determinadas categorías gramaticales⁷: Todas las palabras distintas a estas son palabras acentuadas:

Artículo definido (el, los, la, las, lo); relativo (*adonde, cual, cuala, cuales, cuan, cuando, cuanta, cuantas, cuanto, cuantos, cuya, cuyas, cuyo, cuyos, do, donde, que, quien, quienes*); conjunción (*aunque, ca, como, conforme, conformes, confuso, conque, desque, e, empero, maguer, maguera, mas, mientras, ni, nin, o, pero, porque, pues, que, si, sino, siquiera, u, y*); clítico (*la, las, le, les, lo, los, me, mi, nos, os, se, se la, se las, se les, se lo, se los, te*), preposición (*a, ante, bajo, bajos, cabe, con, conlos, contra, de, desde, durante, en, enta, entre, entro, excepto, hacia, hasta, mediante, para, por, salvo, salvos, sin, so, sobre, tras, troa*), preposición-artículo definido (*al,*

⁷ La preposición / adverbio relativo *según* es una palabra acentuada (Quilis 1981: 315-316). Las palabras átonas incluyen los siguientes nombres, adjetivos y palabras numéricas: *don Antonio, doña Ana, fray Pedro, san Juan, santo Tomás, santa Teresa, María José, dos mil* (Quilis 1981: 315-317), *cuarenta y siete* (Alcina Franch y Bleuca 1975: 443). Dalbor (1969: 215) considera *señor Martínez* una palabra acentuada, pero Stirling (1935: 60) la considera una palabra átona. En este estudio, solo se consideran palabras débiles las formas de títulos que históricamente han perdido el sonido final (*don, doña, san, santa, fray*).

del), adjetivo posesivo antepuesto (*lur, lures, mi, mis, nuestra, nuestras, nuestro, nuestros, su, sus, tu, tus, vuastra, vuestras, vuestro, vuestros*), signo⁸ (-, -, ,, ;, :, !, i, ?, ¿, .., ..., ·, ', ' , “, ” , <, (,), [,], {, }, §, ¶, /, #, ·, ^, +, =, |, —, |)

Lo que sigue es la frecuencia total de palabras acentuadas y palabras débiles con sus correspondientes porcentajes:

Palabras tónicas (1.023.061: 57%), palabras átonas (765.156: 43%)
[Total: 1.788.217].

La siguiente es una lista que muestra todas las combinaciones consecutivas de dos palabras y sus frecuencias en todos los materiales:

- (1) <palabra átona + palabra tónica>: la dicha (6,858), el dicho (6,001), los dichos (3,643), del dicho (3,160), las dichas (1,691), que no (1,596), a vós (1,565), en esta (1,526), que yo (1,197), la dita (1,187), al dicho (1,176), del mes (1,165), que es (1,137), e yo (1,114), nuestro señor (1,098), de esta (1,078), la otra (1046), y dos (1,032), la villa (971), en ella (960)... [Total: 534,830].
- (2) <palabra tónica + palabra átona>: villa de (1,612), todos los (1,359), mes de (1,341), todo lo (1,027), Juan de (953), días del (937), todas las (880), no se (864), carta de (857), María de (825), dixo que (687), días de (649), manera que (630), lugar de (622), obispo de (588), yo el (582), año de (557), era de (550), rey don (549), monesterio de (546)... [Total: 425,960].
- (3) <palabra tónica + palabra tónica>: esta carta (2,243), dicha villa (1,342), dicho es (1,213), dicho monesterio (1,026), esta villa (938), escrivano público (824), esta dicha (802), seis reales (747), cuatro reales (662), dicho señor (639), ocho reales (632), dos reales (631), otra parte (525), dicha cibdad (516), muchos años (515), tres reales (514), cada uno (499), vieren cómo (490), más una (480), aquí adelante (479)... [Total: 361,697].
- (4) <palabra átona + palabra átona>de la (13,967), en la (5,565), de los (5,255), e de (4,373), en el (4,080), que se (3,786), lo que (3,378), a la (3,354), de las (3,014), e a (2,599), a los (2,522), e por (2,487), de su (2,332), para que (2,134), por la (2,051), y que (1,955), e en (1,856), que en (1,785), que la (1,752), y de (1,751)... [Total: 215,352].

⁸ En el lugar de un signo, suponemos una pausa.

De esta manera, dado que (1) <Palabras átona + palabra tónica> muestra la frecuencia más alta, se considera a esta secuencia de dos palabras como la más importante.

El siguiente gráfico muestra las tendencias históricas en la frecuencia de aparición de las cuatro combinaciones anteriores (a: palabras átonas, t: palabras tónicas).

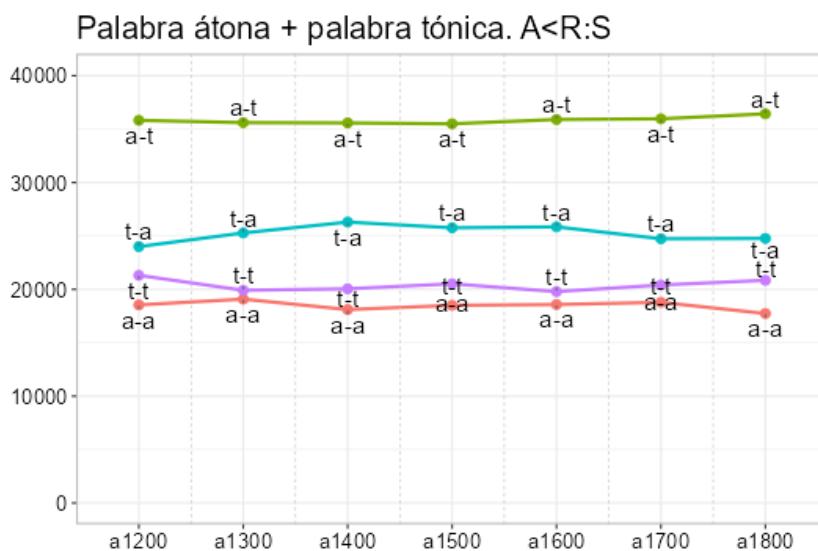


Fig. 4a,b. Cambios en combinaciones de palabras átonas (a) y palabras tónicas (t)

El diagrama anterior muestra que, a lo largo de la historia, <palabra átona + palabra tónica> (a+t) siempre ha sido la combinación más frecuente.

El siguiente es el resultado de examinar el desglose de categorías gramaticales de dos palabras consecutivas⁹:

n+p (107,905), p+n (60,604), p+ar_df (57,418), **ar_df+n (56,889)**, pt+c (53,131), n+c (45,753), a_nm+n (40,130), v_cj+p (35,910), **cl+v_cj (35,731)**, **psa+n (34,872)**, c+p (32,048), pt+p (29,859), c+v_cj (28,264), pt+n (27,783), c+n (27,397), np+np (25,006), a_pa+n (23,326), p+t (22,063), n+a (21,720), ar_df+a_pa (21,411)... [Total: 1,788,217]

A partir de los datos de frecuencia anteriores, nos damos cuenta de que el

⁹ ar: artículo, av: adverbio, c: conjunción, cl: clítico (pronombre átono), cl.cl: clítico+clítico, n: nombre, np: nombre de persona, p: preposición, pn: pronombre, psa: posesivo átono, pt: puntuación, t: topónimo, v: verbo.

orden de las palabras <pronombre + verbo conjugado> (cl+v_cj) coincide con el patrón más frecuente <palabra átona + palabra tónica> y, entre estos, especialmente con <artículo definido + nombre> y <posesivo + nombre>.

La siguiente es una lista de las formas más frecuentes de <artículo definido + nombre>, <posesivo + nombre> y <pronombre + verbo conjugado> en todos los materiales (a1200-a1800).

<artículo definido + nombre>: la villa (971), el señor (645), el rey (642), el día (579), la gracia (490), la ley (463), la iglesia (452), la casa (425), los bienes (417), el escrivano (396), la carta (388), la ciudad (386), las cartas (376), la orden (364), las partes (349), la iglesia (347), la ciudat (292), la manera (281), la moneda (280), la parte (279)... [Totsl: 56,889].

<posesivo + nombre>: nuestro señor (1098), vuestra señoría (943), su magestad (659), vuestra merced (654), vuestra excelencia (633), vuestra magestad (509), mi carta (465), su muger (426), nuestra carta (376), su nombre (362), nuestra señora (332), mis bienes (318), su merced (274), mi señor (244), su casa (241), mi señora (230), su tierra (226), su madre (224), mio signo (217), su signo (209)... [Totsl: 34,872].

<pronombre + verbo conjugado>: se contiene (422), se á (340), lo firmó (302), se ha (213), me obligo (179), se an (162), la fiz (155), me á (154), se llama (153), la confirma (148), vos damos (141), se han (116), se halla (110), se puede (109), le dio (102), se dio (101), se haga (100), se sirva (98), nos obligamos (97), se pueda (97)... [Totsl: 35,986].

Al observar esta lista, se puede ver que los patrones de <palabra corta + palabra larga> son muy similares en general, con algunas escasas excepciones. Esta es también la razón por la que asumimos una correlación entre <artículo definido / posesivo + nombre> y <pronombre + verbo conjugado>.

El siguiente gráfico examina las tendencias cronológicas de sintagmas nominales y verbales seleccionadas de <palabra átona + palabra tónica>: <artículo definido + nombre> (ar+n), <posesivo + nombre> (ps+n), <pronombre + verbo conjugado> (p+v.cj):

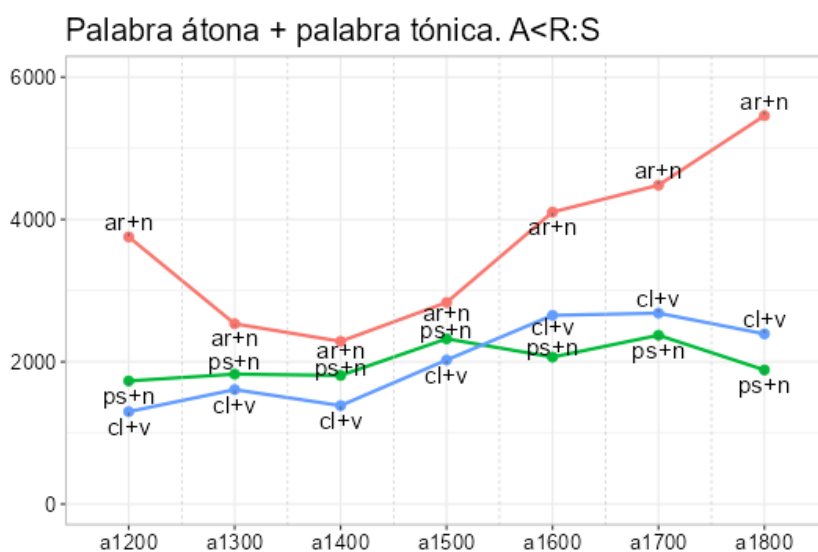


Fig. 5a,b. Cambios cronológicos en frases nominales y verbales de <palabra átona + palabra tónica>

El gráfico anterior muestra que tanto <artículo definido / posesivo + nombre> (ar+n, ps+n) como <pronombre + verbo conjugado> (cl+v) tuvieron una frecuencia consistentemente alta.

De esta manera, <pronombre + verbo conjugado> (*lo + tiene*), que ocupa la abrumadora mayoría entre las cuatro combinaciones de pronombres y verbos, se ajusta al patrón general de <palabra átona + palabra tónica>. En particular, se puede decir que nos encontramos ante una estructura no marcada influida por la analogía con los frecuentes patrones de <artículos definidos / posesivos + nombres>, ya que el patrón, la forma, la estructura, el significado y la frecuencia de aparición son similares.

En general, las siguientes pueden ser las razones por las que la estructura <palabra átona + palabra tónica> muestra la frecuencia más alta. Las colocaciones de acento más frecuentes para las formas de palabras en español son:

... 'CVC(V[s,n])# (C: consonante, V: vocal) (Ueda 2021)

La aparición de (V[s,n]#) en estos paréntesis es libre, y la C anterior está casi completamente limitada a /r, s, n, l, d, z, y/, que pertenecen en todos los casos a una articulación dentoalveolar. En casos distintos a los sonidos dentoalveolares, es casi obligatoria la presencia de una vocal (*capa, cabo, coche, tomo, rico, pago*)¹⁰.

¹⁰ Tales extranjerismos o latinismos como *club, reloj, álbum, déficit* son excepciones.

De esta manera, una característica de las palabras acentuadas es que básicamente tienen acento al final de la palabra.

Cuando se combinan palabras para formar un grupo de palabras, la terminación siempre debe ser una palabra acentuada. Navarro Tomás (1972: 277-305, 278) distingue cuatro tipos de notación fonética de la duración de la pausa: |||, ||, |, '. Por ejemplo, seleccionamos la siguiente parte:¹¹

Lucero ', que, según he sabido después |, tiene ya la costumbre de hacer piernas 'cuando pasa por delante de la casa de Pepita, || empezó a retozar | y a levantarse un poco de mano. |||

Todas las palabras que están delante de diversas pausas en este párrafo son acentuadas. A diferencia del inglés, las palabras débiles no aparecen al final de una frase (por ejemplo, *Where are you from?*, *He is the friend I was thinking of*).

De esta manera, el español tiene la característica de la 'acentuación final', por la que siempre se acentúa el final de una palabra, grupo de palabras u oración. El hecho de que esta característica sea compartida también por <artículo definido / posesivo + nombre> (ar+n, ps+n) y <pronombre + verbo> (cl+v) puede ser una de las razones por la que han ocurrido cambios históricos similares entre ellos.

Como es bien sabido, <artículo definido + nombre> no existe en el latín clásico, y las lenguas romances construyen esta secuencia a partir de <pronombre demostrativo + nombre> *ille homine(m)* en el latín clásico. Este pronombre demostrativo llevaba acento, pero luego en español se gramaticalizó como *illu(m) homine* > *elo homne* > *el hombre*, y se convirtió en un artículo definido débil. La proposición posesiva también era una palabra acentuada (*mió / mío fiijo*) en el español temprano, pero gradualmente se fue debilitando y se convirtió en la forma presente (*mi hijo*). De la misma manera, los pronombres en la estructura <pronombre + verbo conjugado> eran inicialmente fuertes como pronombres demostrativos, pero luego se convirtieron en pronombres personales átonos.

Si estas tres estructuras paralelas (<artículo definido + nombre>, <posesivo + nombre> y <pronombre + verbo>) y sus cambios no pueden deberse a coincidencia casual, entonces se puede suponer una corriente subyacente común entre ellas. Una posible razón para todo esto es que una forma se fusiona con

¹¹ El texto citado no es la transcripción fonética, sino el texto original de Juan Valera, *Pepita Jiménez*, con las pausas añadidas en la transcripción fonética de Navarro Tomás.

formas similares a través de la analogía de otra forma. Si consideramos el estado de supervivencia y la distribución de frecuencia de las formas arcaicas en la Edad Media, la confluencia de estas gramaticalizaciones puede ser en el orden de <artículo definido + nombre> → <posesivo + nombre> → <pronombre + verbo>. Sin embargo, en realidad, se puede suponer que se forma un patrón común como resultado de una analogía mutua, y que ese patrón gradualmente se vuelve más fuerte, se vuelve mayoritario y se convierte en términos no marcados.

(3) Pronombres personales latinos

Aunque está fuera del alcance de este estudio, que trata de la historia del español, conviene indagar aquí el origen del <verbo conjugado - pronombre> (*tiene-lo*) con referencia a investigaciones previas y nuestras.

Pronombres personales latinos como *ego, mei:, mihi, me:, tu:, tui:, tibi, te:, no:s, nostri:, nostrum, no:bi:s, vo:s, vestri:, vestrum, vo:bi:s* y pronombres demostrativos como *es, ea, id, e:ius, ei:, eum, eam, id, eo: ea:, ei:, eae, ea, eo:rum, ea:rum, ei:s, eo:s, ea:s, ea* se derivan de palabras acentuadas protoindoeuropeas (Palmer 1974: 257-259). En latín, hay momentos en que unas palabras están acentuadas y otras inacentuadas, como se muestra a continuación.

La siguiente oración de ejemplo está incluida en el ejercicio del libro de texto de latín clásico (Higuchi y Fujii 1963: 44).

(1) *domina ea in conclave cum liberis servisque se recepit*

"La dama se retiró a la habitación con sus hijos y esclavos."

(2) *cur vos sic languidos video, pueri?* "¿Por qué os veo tan cansados, muchachos?"

(3) *fuitne Marcus tibi amicus?* "¿Fue Marcus un amigo para ti?"

En (1), *se recepit* es un verbo reflexivo, y el pronombre reflexivo *se* se coloca inmediatamente antes del verbo *recepit*. La posición de este pronombre reflexivo es la misma que en el español moderno. En (2), el pronombre personal acusativo *vos* viene antes del verbo *video*, pero se ubica en un lugar distante. En (3), el pronombre personal dativo *tibi* viene después del verbo y está separado del verbo. Se cree que los pronombres personales se colocaron en una posición libre de verbos de esta manera, puesto que los pronombres personales estaban acentuados.

Una de las posiciones libres para este pronombre era al comienzo de una oración. Según Luraghi (2016), los pronombres al comienzo de oraciones como la

siguiente en las *Guerras de las Galias* de César (100-44 a. C.) estaban enfatizados. Si miramos la traducción al inglés, ciertamente es un pronombre demostrativo.

(4) Flumen est Arar, [...]. **Id** Helvetii ratibus ac lintribus iunctis transibant. "There is a river [called] the Arar, [...]. **This** the Helvetii were crossing by rafts and boats joined together." (Caes. Gal. 1.12.10).

Por otro lado, a medida que transcurrió el tiempo, en el *Evangelio de Marcos* del *Nuevo Testamento* (Vulgata, a finales del siglo IV), la enclisis de pronombres (de tercera persona), como la que se muestra en las oraciones de los ejemplos (5) y (6), se volvió muy común (Luraghi, *ibid.*). Sin embargo, la oración del ejemplo (7), que proviene de las *Cartas a Atticus* (Luraghi, *ibid.*) de Cicerón (106-43 a. C), muestra una posición enclítica bastante rara en el latín clásico (Pinkster 1995: 221). Se cree que el pronombre que sigue inmediatamente a este verbo era una palabra átona: de ahí la diferencia.

(5) et adprehendens eum Petrus coepit increpare **eum**. "Peter took him, and began to rebuke **him**." (Mark 8.32).

(6) Conspuent **eum** et flagellabunt **eum** et interficiunt **eum**. "They will spit on **him**, scourge **him**, and kill **him**." (Mark 10.34).

(7) Etenim vel subire **eas** videor mihi summa cum dignitate vel declinare nulla cum molestia posse. "For I think I can either confront **them** with perfect dignity or decline them without any embarrassment." (Cic. Att. 2.19.1).

La parte verbal del <verbo-pronombre> (enclisis) anterior puede contener formas conjugadas (*conspuent*, *flagellabunt*, *interficiunt*) y formas no conjugadas (infinitivo: *in*crepare, *subire*).

Luraghi (*ibid.*) se ocupa únicamente de los pronombres de tercera persona, pero Yamamoto (1997) examina la posición de los pronombres de todas las personas en las traducciones latina, griega e inglesa antigua del *Nuevo Testamento* (*Evangelio de Marcos*). Al sumar la parte latina de este recuento, obtenemos 247 ejemplos de pronombres que vienen después de los verbos y 52 ejemplos de pronombres que anteceden a los verbos, por lo que los pronombres que vienen después de los verbos son, con gran diferencia, los más numerosos. En las oraciones de los siguientes ejemplos (8) y (9), notamos que el pronombre se coloca inmediatamente después del verbo, como se muestra a continuación.

(8) et angeli minisrabant **illi** "and the angels ministered to him" (Mk 1: 13)

(9) Ego baptizaui **uos** aqua "I have baptized you in water" (Mk 1: 8)

En (8) arriba, para el pronombre de tercera persona del singular, *ille, illi*: 'aquel' se usa en lugar de *is, ei*: "él". En latín clásico, se usaban habitualmente *is, ei*: y, en cambio, *ille, illi*: constituían la forma enfática (Lausberg 1973: 164). Por otro lado, si observamos de cerca este mismo ejemplo (8), parece que en el latín bíblico se usaba esta forma generalmente para significar 'él'. Con el tiempo, *illi*: 'a él' se convirtió en *le* en español. Menéndez Pidal (1968: 253) afirma que la vocal inicial de *illi*: fue eliminada por el uso de formas conjugadas como *cantarunt-(i)llu, non-(i)llu*. Dado que se trata de formas latinas, se cree que esto estuvo influenciado por el hecho de que en latín había muchas formas enclíticas (*cantarunt-(i)llu*).

Según las investigaciones de Luraghi (ibid.) y Yamamoto (ibid.), se puede suponer que en el latín clásico la posición del pronombre era libre, pero en el latín bíblico pasó a colocarse preferentemente después del verbo. Por lo tanto, para averiguar la situación del latín bíblico, que es particularmente importante en el estudio de la historia del español, hemos examinado todo el *Antiguo y Nuevo Testamento (Vulgata, finales del siglo IV)* traducido al latín bíblico limitándonos al verbo con frecuencia sumamente alta, como es *dicere* 'decir'. Como resultado de la investigación (un total de 6.181 palabras), hemos encontrado 938 ejemplos de pronombres pospuestos al verbo (a). Además, se encontraron 462 casos de *ad* + nombre (b), 77 casos de nombre en caso dativo (c) y 43 casos de nombres propios en caso dativo después de los verbos (d). En cambio, solo hubo 39 casos de pronombres que preceden a los verbos (e). Por lo tanto, aunque este es solo un ejemplo de *dicere* en la *Vulgata*, es posible que en el latín bíblico los objetos dativos y los nombres con *ad* se colocaran comúnmente después de los verbos. Esto puede haber influido en el hecho de que el pronombre dativo se colocara después del verbo.

La siguiente es una lista de ejemplos típicos (Weber und Gryson 1969, traducción moderna de Reina y Valera 1960):

(a) dixitque ei ego Deus omnipotens [...] "También le dijo Dios: Yo soy el Dios omnipotente: [...]" (Génesis: 35,10). / multi dicent mihi in illa die Domine Domine nonne in nomine tuo prophetavimus [...] S. (Mateo: 7, 22) "Muchos me dirán en aquel día: Señor, Señor, ¿no profetizamos en tu nombre, [...]"

(b) dixit ad Balac sta hic iuxta holocaustum tuum [...] "Entonces él dijo a Balac: Ponte aquí junto a tu holocausto, [...]" (Números: 23, 15) / dixitque ad illos videte et cavete ab omni avaritia [...] "Y les dijo: Mirad, y guardaos de

toda avaricia; [...]" (S. Lucas: 12, 15)

(c) ideo dic filiis Israhel ego Dominus qui educam vos de ergastulo Aegyptiorum [...] "Por tanto, dirás a los hijos de Israel: Yo soy JEHOVÁ; y yo os sacaré de debajo de las tareas pesadas de Egipto, [...]" (Éxodo: 6, 6). / quae cum exisset dixit matri suae quid petam et illa dixit caput Iohannis Baptistae [...] "Saliendo ella, dijo a su madre: ¿Qué pediré? Y ella le dijo: La cabeza de Juan el Bautista. [...]" (S. Marcos: 6, 24).

(d) et dixit Hieremias ad eos sic dicetis Sedeciae [...] "Y Jeremías les dijo: Diréis así a Sedequías: [...]" (Jeremías: 21, 3). / qui conversus dixit Petro vade post me Satana [...] "Pero él, volviéndose, dijo a Pedro: ¡Quítate de delante de mí, Satanás!; [...]" (S. Mateo: 16, 23).

(e) et illis dixit audiente me transite per civitatem sequentes eum et percutite [...] "Y a los otros dijo, oyéndolo yo: Pasad por la ciudad en pos de él, y matad; [...]" (Ezequiel: 9, 5). / hoc enim vobis dicimus in verbo Domini [...] "Por lo cual os decimos esto en palabra del Señor: [...]" (1ª a los Tesalonicenses: 4, 15).

Este uso frecuente de la posposición del pronombre fue heredado por las lenguas romances (Lausberg 1973: 171-208), y parece haber sido transmitido al español temprano ya en el siglo XIII (a1200) como un remanente de la forma latina (Penny 2006: 161). Por lo tanto, suponemos que se encuentran aún más enclisis anteriores a la lengua española temprana del siglo XIII (a1200), sobre la cual existe una gran cantidad de literatura, y que hubo un gran número de enclisis durante el período de transición del latín bíblico a las lenguas romances comunes. Este proceso disminuyó gradualmente, mientras en el español temprano del siglo XIII (a1200), como se confirma en este estudio, <pronombre + verbo>, es decir, la proclisis se situaba en la abrumadora mayoría.

3.2. Verbo conjugado + pronombre (*tiene-lo*)

Observación

El orden de <verbo conjugado - pronombre> (*tiene-lo*), que se retoma en las investigaciones anteriores sobre la historia de la lengua española, se observa normalmente con una alta frecuencia en los materiales medievales. Sin embargo,

en realidad en los documentos notariales es mucho más reducida esta secuencia que <pronombre + verbo conjugado> (*lo + tiene*).

Formas correspondientes (en orden de frecuencia): tóvelo (47), dámosvos (41), dóvos (40), llámase (37), danle (35), tovímoslo (35), oblígome (34), firmólo (33), fióle (29), pidióme (29), fuéle (28), leyósele (25), áse (24), confirmámoslo (22), confírmoles (22), obligámosnos (18), abrióse (15), otorgámosles (15), sépase (15), confírmoles (14)... [Total: 2,135].

Análisis equilibrado-1: variables objetivas = región:año, referencia=sociedad.

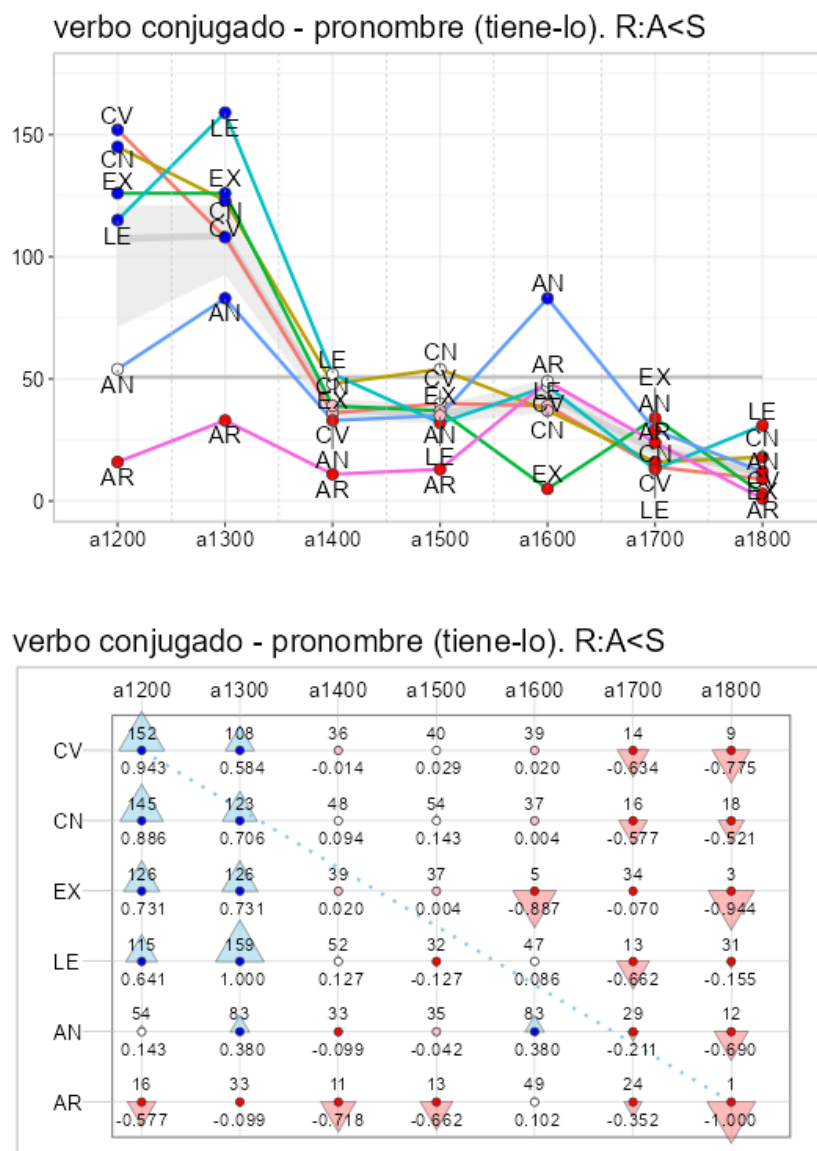


Fig.6a, b. Verbo conjugado+pronombre. PRegión:Año<Sociedad. (Media: 51)
(AN: Andalucía, AR: Aragón, CN: Castilla la Nueva, CV: Castilla la Vieja, EX: Extremadura, LE: León)

La estructura de <verbo conjugado - pronombre> (*tiene-lo*) era reducida, pero relativamente más frecuente en la Edad Media (a1200-a1300) que en la Moderna (a1500-). Sin embargo, disminuyó después de a1400 y desapareció en todas las regiones hacia a1800. Dado que el <verbo conjugado - pronombre> es un orden de palabras tradicional basado en el uso latino, se usaba a menudo en CV, CN y EX medievales. El uso de AN y AR en a1600 y a1700 puede, por tanto, tener un cierto estilo retrógrado.

Análisis equilibrado-2: variables objetivas=sociedad:año, referencia=región.

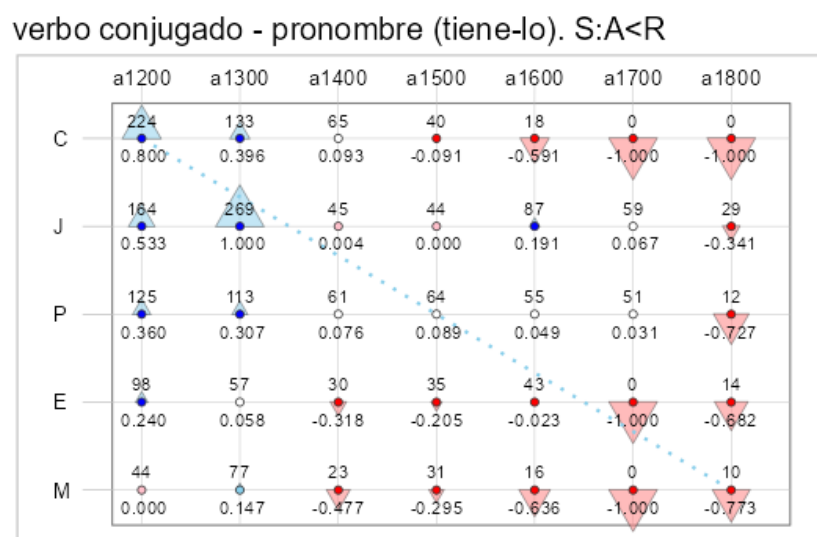
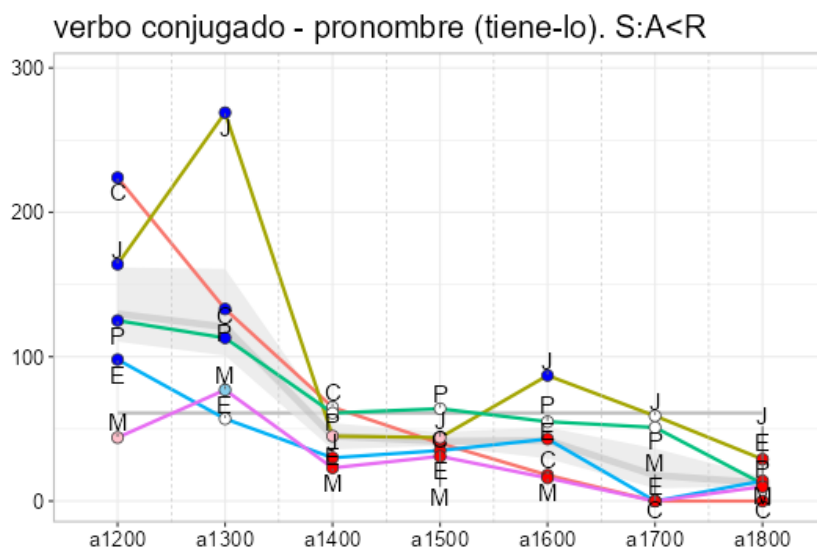


Fig.7. Verbo conjugado+pronombre. Sociedad:Año<Región. (Media: 61)
(C: documento cancelloresco, E: d. eclesiástico, J: d. judicial, M: d. municipal, P: d. particular)

La estructura de <verbo conjugado - pronombre> (*tiene-lo*) era común en C, J y E en a1200-a1300, pero disminuyó después de a1400 y cesó en todos los documentos en a1800, lo que muestra las peculiares características del estilo de escritura tradicional.

Consideración

(1) Palabra que precede al <verbo conjugado - pronombre>

Investigaciones anteriores han señalado que, en la Edad Media, el orden de las palabras <verbo conjugado - pronombre> (*tiene-lo*) se observaba al comienzo de las oraciones. De hecho, según la siguiente tabla, el valor contrastivo que representa la relación (r) es cercano a 0 al comienzo de la oración (#) de la condición antecedente (x), y este es el caso de <verbo conjugado - pronombre> (*tiene-lo*) y <pronombre + verbo conjugado> (*lo-tiene*), en que se usa de manera equivalente. Por otro lado, <verbo conjugado - pronombre> (*tiene-lo*) rara vez se usa en medio de oraciones (=). De esta manera, la posición al inicio de una oración (#) no condiciona fuertemente el pronombre conjugado (*tiene-lo*), pero en comparación con la posición media en la oración (=), el orden de <verbo conjugado - pronombre> se usa con mayor frecuencia.

s. x	s. y	x	y	a	b	c	d	s	r
#	v_cj_cl	36	1,136	17	19	1,119	9,800	10,955	-0.056
=	v_cj_cl	10,919	1,136	1,119	9,800	17	19	10,955	-0.795

Tabla 4. *Relación entre el comienzo de la oración (#) / el medio de la oración (=) y <verbo conjugado - pronombre>.*

Con respecto a la categoría gramatical de la palabra anterior, como se muestra en la siguiente tabla, no existe ninguna categoría gramatical que seleccione específicamente <verbo conjugado - pronombre>:

s. x	s. y	x	y	a	b	c	d	s	r
pst	v_cj_cl	22	1,136	4	18	1,132	9,801	10,955	-0.636
a	v_cj_cl	162	1,136	28	134	1,108	9,685	10,955	-0.654
a_pa	v_cj_cl	88	1,136	14	74	1,122	9,745	10,955	-0.682
c	v_cj_cl	3,884	1,136	590	3,294	546	6,525	10,955	-0.696
pn_ps	v_cj_cl	476	1,136	64	412	1,072	9,407	10,955	-0.731
np	v_cj_cl	279	1,136	31	248	1,105	9,571	10,955	-0.778
n	v_cj_cl	1,116	1,136	120	996	1,016	8,823	10,955	-0.785
v_cj	v_cj_cl	159	1,136	13	146	1,123	9,673	10,955	-0.836

v_if	v_cj_cl	145	1,136	10	135	1,126	9,684	10,955	-0.862
pn_d	v_cj_cl	162	1,136	11	151	1,125	9,668	10,955	-0.864
p	v_cj_cl	32	1,136	2	30	1,134	9,789	10,955	-0.875
a_id	v_cj_cl	22	1,136	1	21	1,135	9,798	10,955	-0.909
av	v_cj_cl	884	1,136	25	859	1,111	8,960	10,955	-0.943
cl	v_cj_cl	460	1,136	3	457	1,133	9,362	10,955	-0.987
av_r	v_cj_cl	150	1,136	1	149	1,135	9,670	10,955	-0.987
pn_r	v_cj_cl	1,797	1,136	2	1,795	1,134	8,024	10,955	-0.998
av_ir	v_cj_cl	93	1,136	0	93	1,136	9,726	10,955	-1.000
pn_id	v_cj_cl	56	1,136	0	56	1,136	9,763	10,955	-1.000
av_r_T	v_cj_cl	35	1,136	0	35	1,136	9,784	10,955	-1.000

Tabla 5. *Relación entre la categoría gramatical de la palabra anterior y <verbo conjugado - pronombre>*

Por tanto, se puede decir que el <verbo conjugado + pronombre> (*tiene-lo*), característico del español medieval, aparecía especialmente al inicio de las frases (después del inicio y los signos de puntuación). En esta posición, el pronombre débil parece permanecer en la posición de la conjugación del verbo (acentuación) porque no hay ninguna palabra acentuada precedente que lo respalde. Sin embargo, hubo muchas excepciones.

El siguiente gráfico muestra los cambios cronológicos cuando el orden de <verbo conjugado - pronombre> (*tiene-lo*) aparece al principio de una oración (#) y cuando aparece en medio de una oración (=).

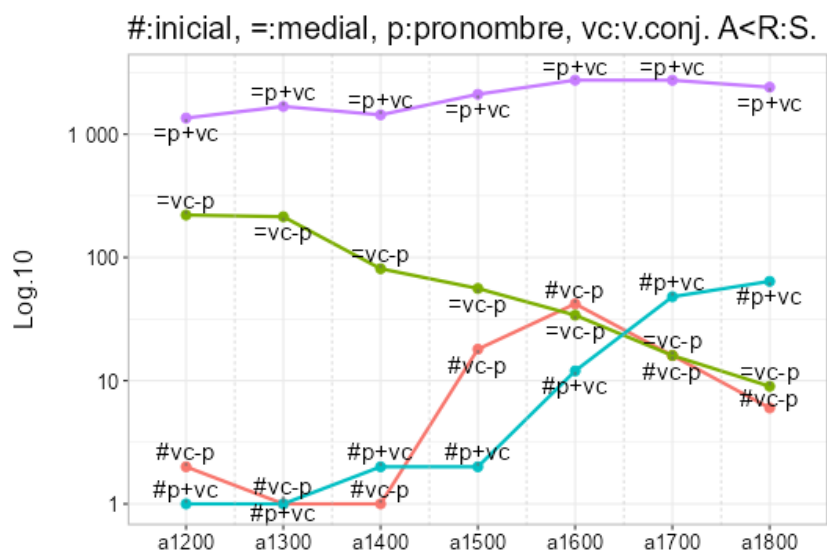


Fig. 8. <Forma conjugada - pronombre> en comienzo de oración (#) y en medio de oración (=)

De esta manera, el <verbo conjugado - pronombre> (*tiene-lo*) disminuye monótonamente en medio de una oración (=) y casi desaparece en a1800, pero al comienzo de una oración (#) regresa en a1500-a1700, especialmente en a1600, puesto que ha aumentado notablemente. Esta es una característica del estilo retrógrado vista en Andalucía (AN) como variable regional y en documentos judiciales como variable social. Casi todo este proceso terminará en torno a a1800.

Por ejemplo, cuando computamos manualmente las posiciones de los inicios de las narraciones, a excepción de conversaciones, en Benito Pérez Galdós (1843-1920), en concreto en *Marianela* (1878), encontramos que el número de <verbo conjugado - pronombre> (*tiene-lo*) es 48, y <pronombre + verbo conjugado> es 7. En cambio, en las oraciones conversacionales, la proporción se invierte: el número de <verbo conjugado - pronombre> (*tiene-lo*) es 9, y el número de <pronombre + verbo conjugado> (*lo+tiene*) es 65 (Ueda 2024). De todo ello se deduce que <verbo conjugado - pronombre> representa al estilo literario y que <pronombre + verbo conjugado> es más propio del estilo coloquial. Esto significa que los rasgos tradicionales de la Edad Media persistieron en las obras literarias hasta finales del siglo XIX.

(2) Razones de la disminución de <verbo conjugado - pronombre>

La razón del abandono de la secuencia <verbo conjugado - pronombre> (*tiene-lo*) parece estar en que los términos no marcados representaban la abrumadora mayoría en la construcción claramente en competencia <pronombre + verbo conjugado> (*lo+tiene*). Esto se debe a que el cambio de un término marcado a un término no marcado es natural.

Aquí conviene recordar lo ya abordado más arriba, que <pronombre + verbo conjugado> (*lo + tiene*) es similar a <artículo definido / posesivo + nombre> (*la / mi casa*), que tiene la estructura <palabra átona + palabra tónica>, y que siempre tuvo y tiene una frecuencia alta. El siguiente diagrama compara las frases nominales y verbales correspondientes (eje logarítmico, ar+n: <artículo definido + nombre>, ps+n: <posesivo + nombre>, p+vc: <pronombre + verbo conjugado>, vc-p: <verbo conjugado-pronombre>):

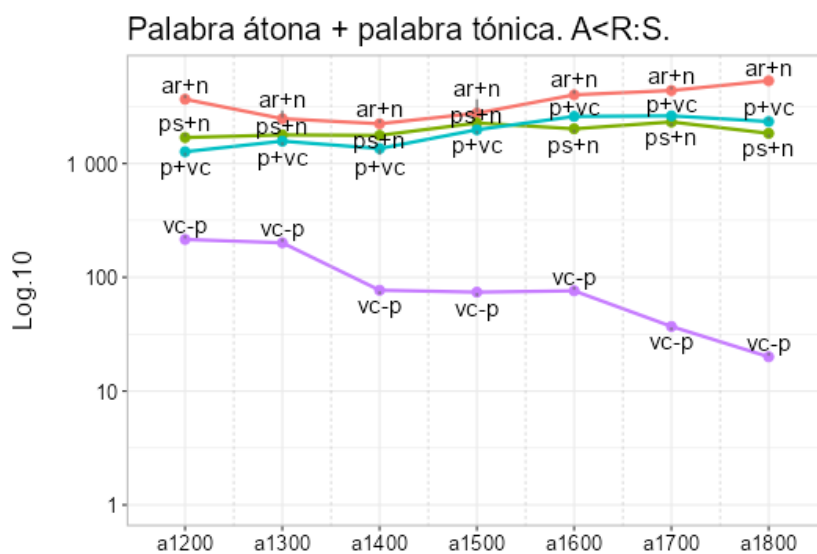


Fig. 9. Cambios cronológicos en <palabras átonas + palabras tónicas> y <formas conjugadas - pronombres>.

Como muestra el diagrama anterior, <artículo definido + nombre> (ar+n: *la casa*) y <posesivo + nombre> (ps+n: *mi casa*) se utilizaron con una frecuencia similarmente alta a lo largo de los siglos. Lo mismo ocurre con <pronombre + verbo conjugado> (p+vc: *lo+tiene*). Por otro lado, la tendencia en el orden inverso de <verbo conjugado - pronombre> (vc-p: *tiene-lo*) es bastante diferente, con una

frecuencia general baja y una disminución pronunciada a lo largo de los años. Sin embargo, no desapareció del todo y permaneció hasta el año 1800.

La siguiente es una lista de formas de palabras frecuentes de <artículo definido + nombre>, <posesivo + nombre> y <pronombre + verbo conjugado> en la Edad Media (a1200-a1400):

<artículo definido + nombre>: el rey (502), la iglesia (451), la gracia (388), la moneda (262), la villa (256), la carta (254), la ciudad (252), las partes (233), la orden (203), la pena (187), el día (179), la ley (173), el convento (170), la manera (155), la paga (153), las cartas (149), la tenencia (148), los bienes (142), los testigos (139), los clérigos (138)... [Total: 18,139].

<posesivo + nombre>: nuestro señor (467), mi carta (347), mis bienes (218), nuestra carta (191), mio signo (183), su signo (183), nuestros bienes (175), sus entradas (143), su nombre (136), su muger (133), mi signo (122), mio padre (108), sus derechos (105), su tierra (91), nuestro salvador (90), mi muger (89), mi padre (85), mi señor (82), mio sello (82), mis regnos (81)... [Total: 12,270].

<pronombre + verbo conjugado>: se contiene (207), la confirma (148), la fiz (139), vos damos (123), me obligo (101), vos vendo (90), lo fiziesse (85), vos vendemos (80), vos mando (72), nos obligamos (71), me tornaría (55), la mostrare (45), lo oviere (45), vos pechemos (45), se cumple (43), lo vala (40), me otorgo (39), vos fuere (39), la mando (38), vos será (38)... [Total: 9,819].

Estas secuencias de palabras comúnmente incluyen el patrón de acentuación <palabra débil + palabra acentuada>, el patrón de longitud de palabra <palabra corta + palabra larga> y el patrón <cambio paradigmático + flexión> (→3.1.). Por otro lado, la forma de <verbo conjugado - pronombre> (*tiene-lo*) tratada en esta sección presenta un patrón completamente diferente como se muestra a continuación:

<verbo conjugado - pronombre>: dámosvos (40), dóvos (37), tóvelo (37), obligóme (33), pidióme (28), confirmámoslo (22), confírmoles (22), obligámosnos (18), otorgámosles (15), confírmoles (14), embiáronnos (14), embiáronme (13), otórgovos (12), pidiéronnos (12), díxome (10), pidiónos (10), toviémoslo (10), tovimoslo (10), véndovoslo (10), otorgámosvos (9)... [Total: 1,136]

De esta manera, <pronombre + verbo conjugado> (*lo + tiene*) mostró una absoluta mayoría, en línea con <artículo definido / posesivo + nombre>. Por otro lado, pensamos que el <verbo conjugado + pronombre> (*tiene-lo*) tenía una estructura significativamente diferente de <artículo definido / posesivo + nombre>, por lo que era relativamente raro, gradualmente disminuyó y, consecuentemente, quedó marcado.

Por otro lado, el <verbo conjugado - pronombre> (*tiene-lo*) se usó de manera relativamente estable en la Edad Media, aunque con menos frecuencia que el <pronombre + verbo conjugado> (*lo+tiene*). Creemos que la razón de esto es que el verbo conjugado y el pronombre se combinaron para formar una sola palabra. Esto se debe a que una sola palabra podría sobrevivir por sí sola, sin verse afectada por el patrón de acentuación de dos palabras de <palabra débil + palabra acentuada>. Sin embargo, fue difícil escapar de la influencia de la mayoría que representaba la estructura <pronombre + verbo conjugado> (*lo + tiene*), por lo que fue decayendo gradualmente.

Penny (2006: 163) sostiene que la omisión de la /e/ al final de pronombres añadidos después de una palabra acentuada, como *metiól en el mayor az, pagós mio Cid*, se debe a la diferencia entre la palabra anterior (principalmente un verbo) y afirma que esto se debe a que formaban una "una sola unidad fonética". Esto también indica que *metiol* y *pagós* eran palabras únicas¹².

Sin embargo, el grado de finalización de la monolexicalización del <verbo conjugado - pronombre> (*tiene-lo*) fue menor que el del <verbo no conjugado - pronombre> (*tener-lo*). La razón parece ser que la parte conjugada incluye flexión. Esta flexión es completamente funcional en <verbo conjugado - pronombre> (*tiene-lo*) y cambia según la persona, el modo y el tiempo. Además, en algunos casos, como en *tiene-lo*, la vocal acentuada de la tercera sílaba desde el final de la palabra está diptongada (*e>ie*). Esto es algo que no existe en la palabra única normal. Por otro lado, el <verbo no conjugado - pronombre> (*tener-lo*) tiene una forma fija, porque el verbo no conjugado no está flexionado, y tiene el mismo patrón que, por ejemplo, *parte, perla, puerta, colmo, Carlos*, etc., y es casi

¹² Penny (2006: 159) afirma que un pronombre átono forma una sola palabra con una palabra acentuada tanto cuando va antes como después de dicha palabra acentuada. En cambio, nosotros consideramos <pronombre + verbo conjugado> como dos palabras, y cuando viene después de una palabra acentuada, se convierte en una sola palabra: <verbo conjugado - pronombre>.

completamente unificado, por lo cual se convirtió en una palabra. Se cree que esta monolexicalización incompleta es la causa del declive del <verbo conjugado - pronombre>. El estado monoléxico se analizará en detalle en la siguiente sección.

3.3. Verbo no conjugado + pronombre (*tener-lo*)

Observación

En la lengua española moderna, la relación posicional entre el verbo no conjugado del verbo (infinitivo, gerundio, participio pasado) y el pronombre se invierte y se convierte en una palabra constituida por <verbo no conjugado - pronombre> (*tener-lo*). De hecho, desde una perspectiva histórica, este orden de palabras rara vez se vio en la Edad Media y se estableció en los tiempos modernos. En esta sección, observaremos estos antecedentes históricos y luego exploraremos las razones por las cuales la distribución de frecuencia histórica es diferente del caso del <pronombre + verbo conjugado> (*lo-tiene*).

Formas correspondientes (en orden de frecuencia): darle (76), averse (63), hallarse (43), habiéndose (40), haberse (39), hallándose (38), aviéndose (35), llamarse (35), darme (33), mandarme (33), dándole (29), hacerlo (29), diciéndola (25), siéndolo (25), serbirte (24), haberse (23), dándoles (22), averle (21), darla (21), siéndole (21)... [Total: 5,842].

Análisis equilibrado-1: variables objetivas = región:año, referencia=sociedad.

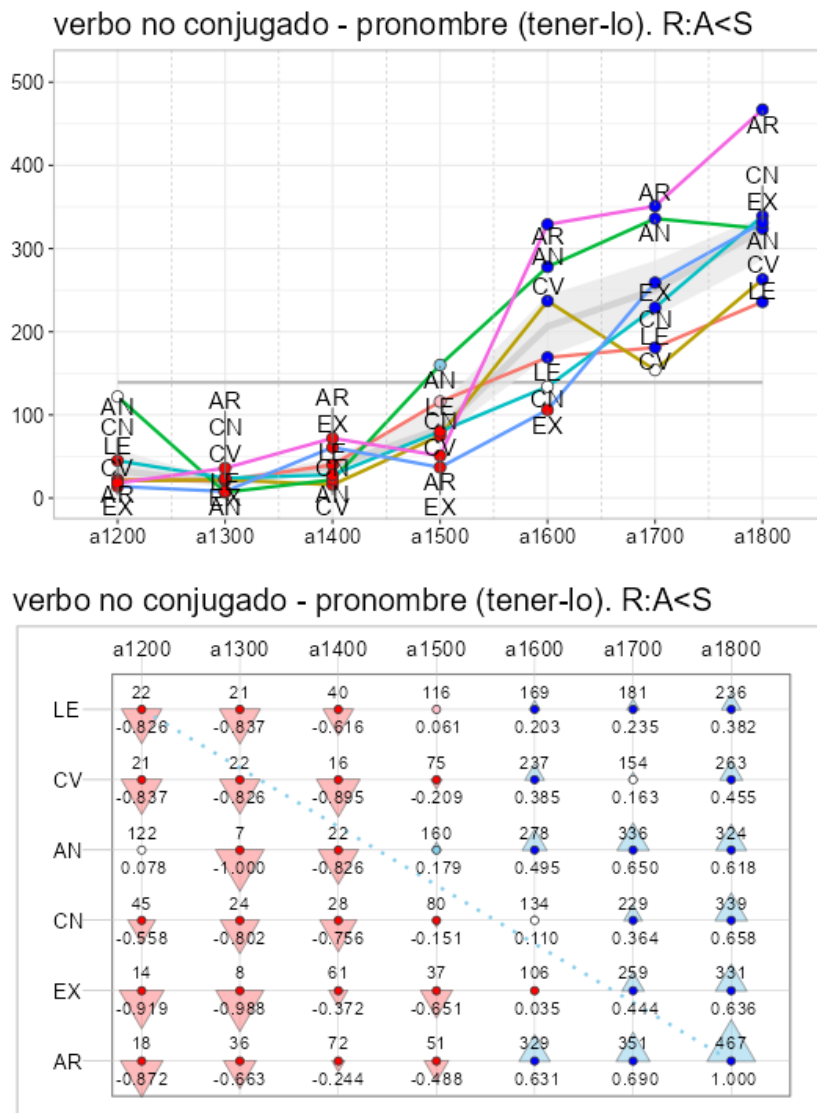


Fig. 10a, b. Verbo no conjugado+pronombre. Región: Año<Sociedad (Media 146).

(AN: Andalucía, AR: Aragón, CN: Castilla la Nueva, CV: Castilla la Vieja, EX: Extremadura, LE: León)

El <verbo no conjugado - pronombre> (*tener-lo*) no se usó mucho en a1200-a1500, pero aumentó significativamente en a1600. En particular, es notable el aumento en las regiones innovadoras: AR, CN y AN. Por tanto, en el caso de <verbo no conjugado - pronombre> (*tener-lo*), no hubo cambio en Edad Media: enclisis → Edad Moderna: proclisis y, por el contrario, se puede decir que la enclisis se estableció en la época moderna.

Análisis equilibrado-2: variables objetivas=sociedad:año, referencia=región.

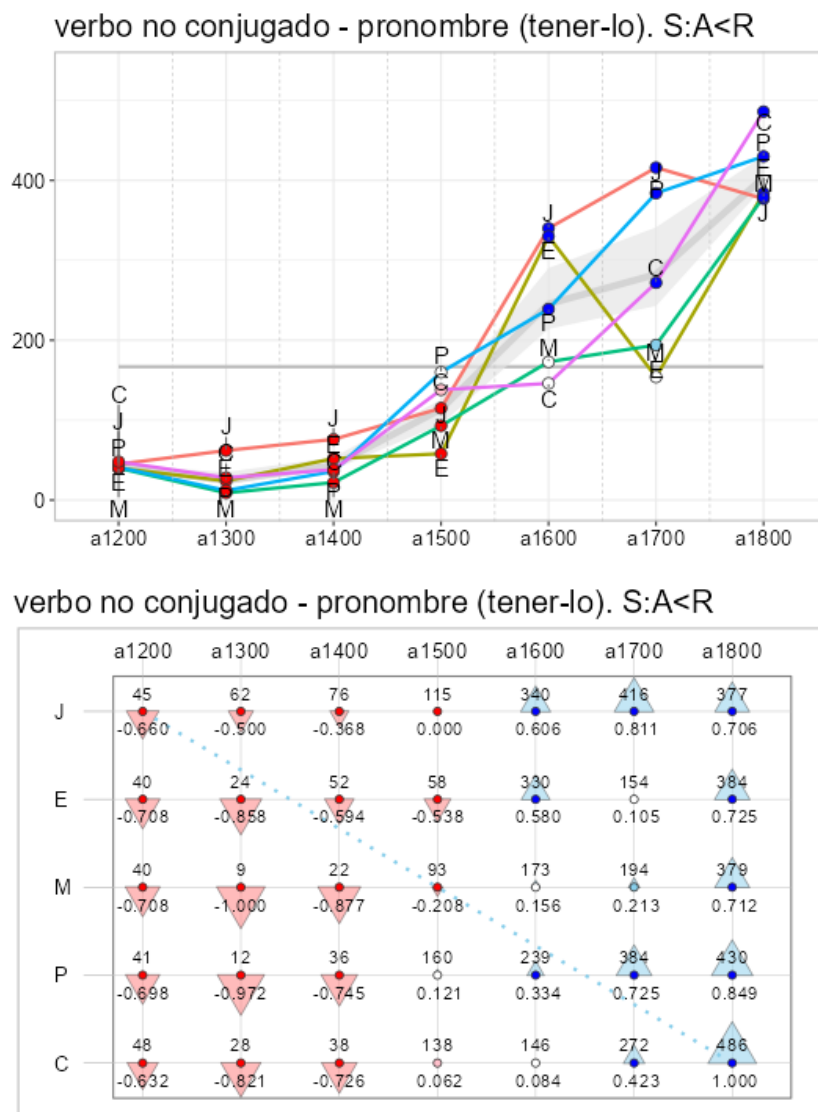


Fig. 11a,b. Verbo no conjugado+pronombre. Sociedad: Año<Región. (Media 178)

(C: documento canclleresco, E: d. eclesiástico, J: d. judicial, M: d. municipal, P: d. particular)

En general, la forma de <verbo no conjugado - pronombre> se usó menos en a1200-1500, pero aumentó significativamente en a1600.

Consideración

Monolexicalización de <verbo no conjugado - pronombre>

En el orden de frecuencia, después de <pronombre + verbo conjugado> (*lo+tiene*) se encuentra <verbo no conjugado - pronombre> (*tener-lo*). Esta forma <*tener-lo*> era rara en la Edad Media, pero aumentó rápidamente en el período

posterior. Este aumento no es similar a la predominancia estable de <pronombre + verbo conjugado> (p+vc: *lo+tiene*) que hemos visto en 3.1. Ambos tienen diferentes propiedades: aumento dinámico en <vn-p> (*tener-lo*) y estabilidad predominante estática en <p+vc> (*lo tiene*). La razón lingüística estriba en que el pronombre (débil) del <verbo conjugado - pronombre> (vn-p: *tener-lo*) se coloca al final, y el verbo no conjugado (*tener*) permanece sin cambios morfológicos, totalmente diferente de la estructura nominal (*la casa, mi casa*) que apoyaría al sintagma verbal <pronombre + verbo conjugado> (*lo+tiene*) (→3.1).

La estructura sintáctica aparentemente excepcional de <palabra acentuada + palabra débil> de <verbo no conjugado - pronombre> (vn-p: *tener-lo*) no constituye una excepción sintáctica si se considera como una sola palabra constituida de <palabra acentuada - clítico>. A continuación, examinaremos el hecho de que <palabra acentuada - clítico> (*tener-lo*) es una sola palabra desde los parámetros del sistema lingüístico (fonología, morfología, sintaxis) y el uso del lenguaje (escritura no normativa).

En primer lugar, veamos el aspecto fonológico. El pronombre no conjugado tiene un acento único colocado cerca del final de la palabra: *tenerlo, teniéndolo*. Esto es lo mismo para todas las palabras acentuadas (palabras individuales), como hemos visto en 3.1.

Por otro lado, el principio español de 'acento en posición final' en la estructura de <palabra débil + palabra acentuada> mencionado anteriormente (3.1) contradice la continuación de las dos palabras <verbo no conjugado> + <pronombre>. Esto se debe a que este pronombre es débil. Creemos que por esta razón <pronombre no conjugado> pasó a ser una sola palabra.

De hecho, la siguiente fusión frecuente de los infinitivos y clíticos átonos (*tener-lo > tenello*) indica su estado monoléxico:

Fusión de los verbos no conjugados y clíticos átonos: dalle (11), dallo (3), hazello (3), vella (3), abelle (2), avelles (2), avellos (2), cobrallos (2), dalla (2), dalles (2), dejalle (2), dezille (2), dichavilla (2), hazella (2), hazelle (2), negallas (2), pagallo (2), parcelle (2), recibillos (2), ..., [Total: 200].

En términos de morfología, observamos que el primer elemento (*tener*) del <verbo no conjugado - pronombre> permanece sin cambios, y el último elemento (*lo*) indica cambio paradigmático: *tenerme, tenerte, tenerlo*, etc. Esta forma de <parte invariante + parte cambiante> es idéntica a la creada por nombres, adjetivos

y verbos generales (todos son unidades monoléxicas).

A continuación, veamos el aspecto sintáctico. En la Edad Media, los pronombres *non* o pronombres de sujeto a menudo se interpolaban entre la forma <pronombre + verbo conjugado> (*nos non vala, les yo fago, vos yo vendo*) (Menéndez Pidal 1908, 1976: 409-411, Castillo Lluch 1996: 293-302). La siguiente es una lista de <verbo + interpolación + pronombre> encontrada en los documentos notariales utilizados en este estudio.

<verbo + interpolación + pronombre>: *nos non vala* (frecuencia: 76), *me non vala* (45), *les yo fago* (18), *vos yo vendo* (17), *los yo aya* (14), *nos non valan* (13), *vos yo fago* (13), *vos así vendo* (12), *lo non consientan* (12), *me non valan* (12), *les nós fazemos* (12), *me non pueda* (10), *les non mengüe* (10), *vos nós vendemos* (10), *les non vayan* (9), *nos non podamos* (9), *lo nós avemos* (8), *la yo tenía* (7), *gelo non consientan* (7), *las non podades* (6), ..., [Total: 930].

Sin embargo, esta interpolación no se advierte entre el verbo y el pronombre en este orden: **de tenernonlo, *para fazeryovos*. Esto también se debe al hecho de que el <verbo no conjugado - pronombre> era una sola palabra.

El hecho de que <pronombre + verbo conjugado> (*lo+tiene*) sea sintácticamente dos palabras y <verbo no conjugado - pronombre" (vn-p: *tener-lo*) sea una palabra, también se confirma sintácticamente en los siguientes ejemplos de RAE y AALE.

RAE y AALE (2009: 1208) explica la diferencia en la estructura de coordinadas entre proclisis y enclisis con la siguiente oración de ejemplo:

Lo leyó y resumió / **para leer y resumirlo*.

Fukushima (2021: 141) explica (traducción nuestra):

Cuando dos o más objetos forman una estructura coordinada y están precedidos por un verbo, se pueden omitir todos los pronombres excepto el primero. En cambio, si los pronombres van después de un verbo, no se puede omitir ningún pronombre.

El mismo autor agrega el siguiente ejemplo imperativo y explica que el segundo *lo* en (1) se puede omitir, pero ninguno de los *lo* en (2) se puede omitir. Para mostrar la diferencia entre las estructuras coordinadas de proclisis y enclisis, debemos considerar la posibilidad de omitir el siguiente pronombre, como en

Fukushima (*ibid.*):

(1) Paco lo leyó y (lo) resumió en un santiamén.

(2) Léelo y resúmelo cuanto antes.

Este fenómeno ocurre de manera similar cuando el verbo es infinitivo (3) y gerundio (4) (ejemplos creados del autor).

(3) Lo voy a leer y resumir / Voy a leerlo y [resumirlo / *resumir].

(4) Lo estoy leyendo y resumiendo / Estoy leyéndo y [resumiéndo / *resumiendo].

Esto muestra que <verbo no conjugado + pronombre> se combina con un verbo para formar una sola palabra, por lo que el elemento *lo* que contiene no se puede compartir con otras palabras, lo mismo que, por ejemplo, la siguiente oración de ejemplo no está permitida (ejemplo creado por nosotros).

(5) No lo vamos a incluir sino [excluir / *ex-].

Otro hecho histórico apoya nuestra teoría monoléxica. Castillo Lluch (1999), que investigó ampliamente la posición de los pronombres débiles en la literatura medieval, la historia y la literatura judicial, encontró muchas formas combinadas de *tener-lo* en su extensa colección de ejemplos¹³: *pedirle, colgarla, ganarlas, ponerlas, etc.; otorgandolis, dandovos, pidiendol, etc.* Por otro lado, casi todos los pronombres en <pronombre + verbo conjugado> (*lo+tiene*) están separados del infinitivo:¹⁴ *las fazer, gelos dar, se quejar, se mantener, delos menazar, nol amar, se defendiendo, nol conociendo, les dando, etc.* Dado que en esa época no existían la gramática escolar ni la ortografía estándar, prescriptivas, la notación continua como esta probablemente se formaría por la intuición del hablante local. Creemos que esto indica que el pronombre conjugado (*tener-lo*) se consideraba como una sola palabra, y el <pronombre + verbo no conjugado> (*lo+tener*) se consideraba como dos palabras.

La *Ortografía* moderna de la Real Academia (RAE y AALE 2010: 530) afirma ([...] el subrayado es del autor):

¹³ Se encuentran también las formas separadas: *sacar les, uestir les, llamando la.*

¹⁴ La única excepción es *l'auia dito Aristotus (Liber Regum, 9, 29)* (Castillo Lluch. *ibid.*). No se pueden encontrar otras formas fusionadas o continuas de <pronombre + verbo conjugado>.

[...] si [los pronombres átonos] van pospuestos se llaman *enclíticos* y forman con el verbo una sola palabra gráfica: *Dilo tú; Decilo vos; Dádmelas; Voy a pedírsela; Se marchó diciédoos adios; Haberlo hecho antes.*

La expresión mencionada aquí "una sola palabra gráfica" no es solo una cuestión de notación, sino que también se considera "una sola palabra" dentro del sistema lingüístico.

El siguiente gráfico muestra los cambios históricos en el <verbo no conjugado - pronombre>:

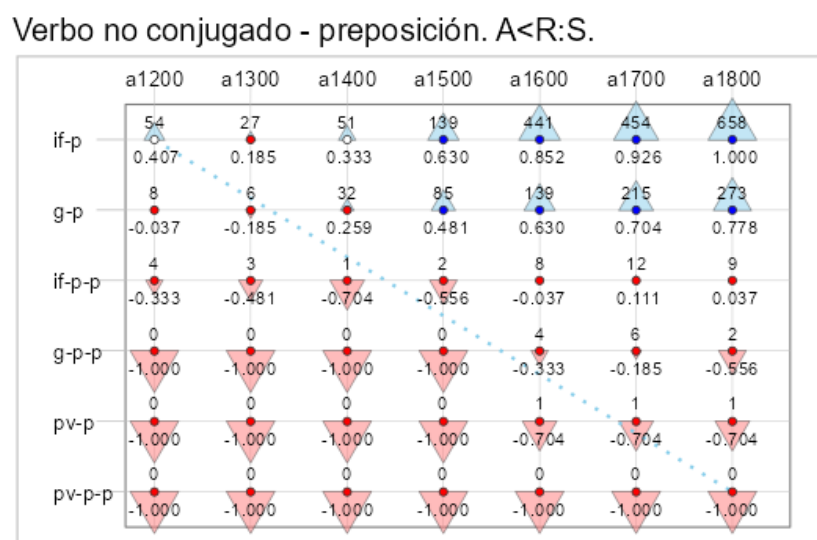
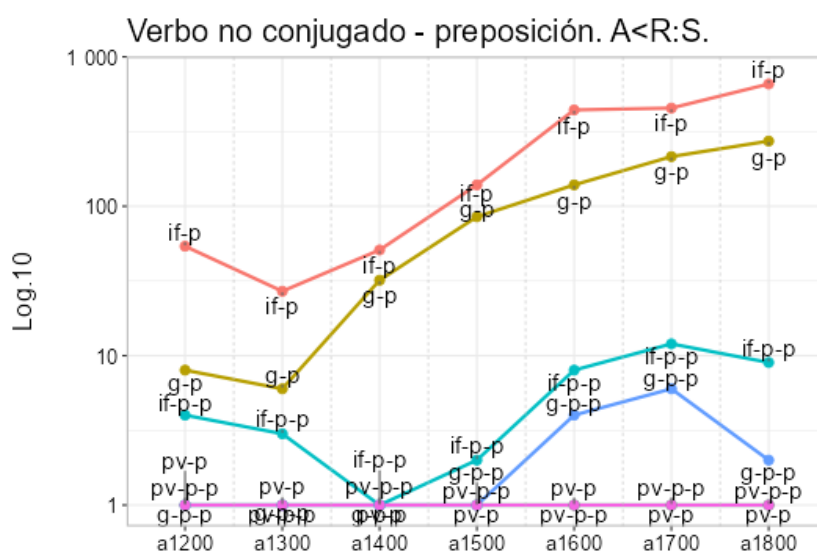


Fig. 12a, b. *Verbo no conjugado - pronombre.*
(if:infinitivo, g:gerundio, pv:participio verbal).

El <infinitivo - pronombre> (*tener-lo*) ha aumentado significativamente. En cambio, el <gerundio - pronombre> no se usa tanto como <infinitivo + pronombre>. Una posible razón para esto es el patrón de acentuación marcado (acentuación en la tercera sílaba contando desde el final de la palabra: *teniéndolo*). Por otro lado, el patrón de acentuación de <infinitivo + pronombre> no está marcado, porque el acento está en la segunda sílaba contando desde el final de la palabra, lo que es normal en español (*comprar-lo*) (Ueda 2024).

Además, hay que tener en cuenta que la frecuencia total de los infinitivos es muy superior a la frecuencia total de los gerundios, como muestra la siguiente tabla.

*	~	~.p	~p.p.	Total
if	33363	3900	167	37430
g	6100	1599	182	7881
Total	39463	5499	349	45311

Tabla 6. *Frecuencia total. Infinitivo (if) y gerundio (g) (p: pronombre)*

De esta manera, el <verbo no conjugado - pronombre> (*tener-lo*) no recibe una acción analógica de <palabra átona + palabra tónica> asumida en el análisis de <pronombre + verbo conjugado> (*lo + tiene*), en particular el <artículo definido / posesivo - nombre>. Parece que la analogía del sintagma nominal no funciona eficazmente debido a diferencias estructurales y se ha desarrollado de forma independiente. Se intuye, por tanto, que lo que contribuyó a su desarrollo fue la monolexicalización de <verbo conjugado - pronombre>. La razón del uso particularmente frecuente de <infinitivo - pronombre> es que su patrón de acentuación no está marcado (el acento está en la segunda sílaba contando desde el final de la palabra) y que el infinitivo en sí se usaba con frecuencia.

Incluso otras formas más reducidas son aquellas en las que se combinan dos pronombres (if-p-p, g-p-p, pv-p-p, ip-p-p). Las ubicaciones de acento en estas formas tampoco coinciden con el <...'CVC(V[s,n])#> no marcado. Por tanto, consideramos que la conversión en una sola palabra fue incompleta.

A continuación, exponemos las listas de las formas y frecuencias correspondientes:

if-p: darle (76), averse (63), hallarse (43), haberse (39), llamarse (35), darme (33), mandarme (33), hacerlo (29), serbirte (24), haberse (23)... [Total: 3,953].

g-p: haviéndose (40), hallándose (38), aviéndose (35), dándole (29), diciéndola (25), siéndolo (25), dándoles (22), siéndole (21), aviéndola (17), aviéndole (17)... [Total: 1,599].

pv-p: curádola (2), dádole (2), dádose (2), portádose (2), acavádose (1), acordádose (1), adornádaslas (1), anunciádole (1), aprendídola (1), atrasádose (1)... [Total: 42].

if-p-p: avérselo (6), dársela (5), dárselo (5), havérsele (5), avérsele (3), concertélo (3), fazérvoslo (3), avérseles (2), contrallárgela (2), dárselo (2)... [Total: 171].

g-p-p: aviéndosele (11), aviéndosele (3), haviéndosele (3), abiéndosele (2), aperziviéndosele (2), diciéndosele (2), dándosele (2), dándosele (2), habiéndosele (2), haviéndosele (2)... [Total: 74].

pv-p-p: dádosele (2), pagádosele (1) [Total: 3].

3.4. Pronombre + verbo no conjugado (*lo+tener*)

Observación

El caso final de la combinación del pronombre y verbo es <pronombre + verbo no conjugado> (*lo + tener*). Al igual que el <verbo conjugado - pronombre> (*tiene-lo*) que vimos anteriormente, este es un residuo de la forma latina (→4.2.) y no se encuentra en las lenguas modernas. Si solo miramos las formas de las palabras en ambas secuencias, no podemos encontrar puntos en común, pero ambas se parecen en que compiten con las formas no marcadas, es decir, <verbo no conjugado - pronombre> (*tener-lo*) y <pronombre + verbo conjugado> (*lo + tiene*) respectivamente y, por consecuencia, han quedado marcadas. Se cree que esta marcación es la causa de la disminución en la frecuencia de uso. Entre las cuatro combinaciones, este orden de palabras <pronombre + verbo no conjugado> (*lo + tener*) ha sido el menos frecuente a lo largo de la historia.

Formas correspondientes (en orden de frecuencia): lo fazer (53), vos fazer (53), les fazer (44), vos dar (37), les ir (30), lo vender (22), los tener (21), la fazer (20), vos defender (18), lo dar (16), los dar (16), lo aver (15), nos ayuntar (15), gela quebrantar (13), les passar (12), lo tener (12), le dar (11), les dar (11), lo cumplir (11), vos redrar (11)... [Total: 1,536].

Análisis equilibrado-1: variables objetivas = región:año, referencia=sociedad.

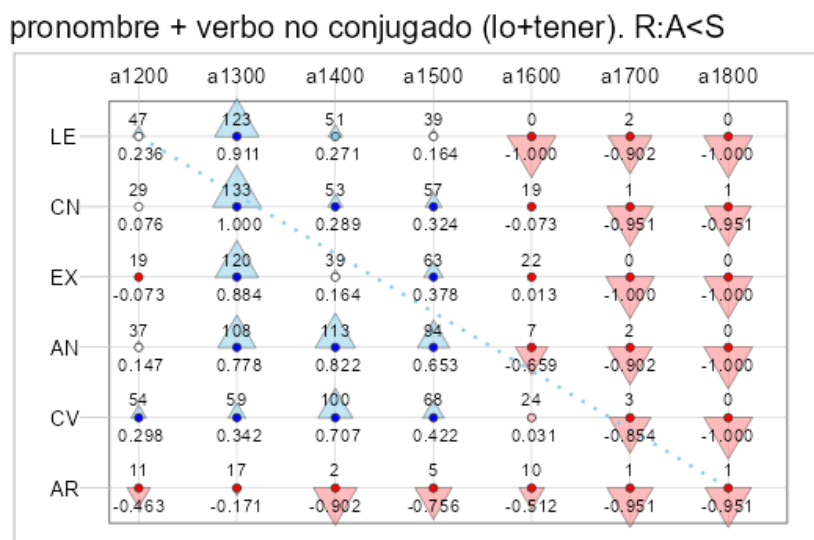
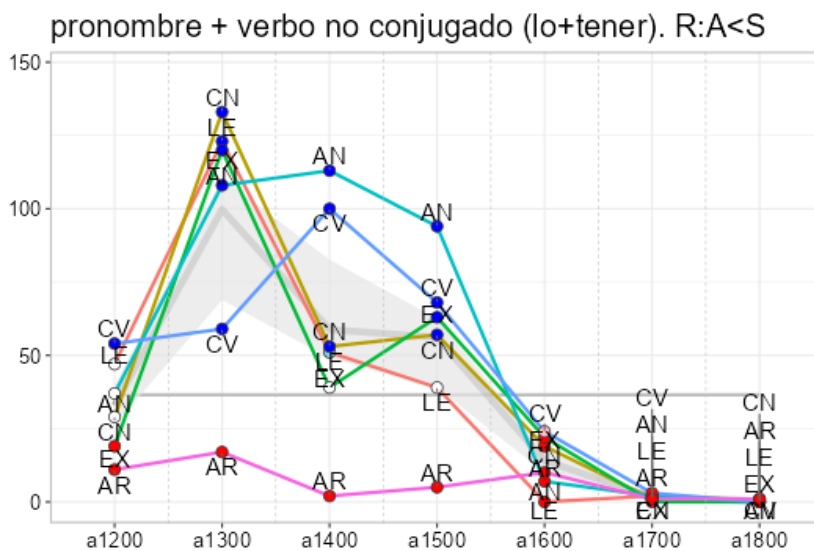


Fig. 13a, b. *Pronombre+verbo no conjugado. Región: Año<Sociedad.(Media: 37)*
 (AN: Andalucía, AR: Aragón, CN: Castilla la Nueva, CV: Castilla la Vieja, EX: Extremadura, LE: León)

<Pronombre + forma flexionada> (*lo+tener*) fue menos frecuente en general y disminuyó gradualmente. Desde a 1600, fue casi nulo. En a1300 tuvo una mayoría significativa en todas las regiones excepto AR (la región más innovadora). En general, el AR fue pequeño. Por lo tanto, se cree que esta forma de palabra se usó con relativa frecuencia en años y regiones que preservan la tradición y relativamente menos en años y regiones que exhiben innovación.

Análisis equilibrado-2: variables objetivas=sociedad:año, referencia=región.

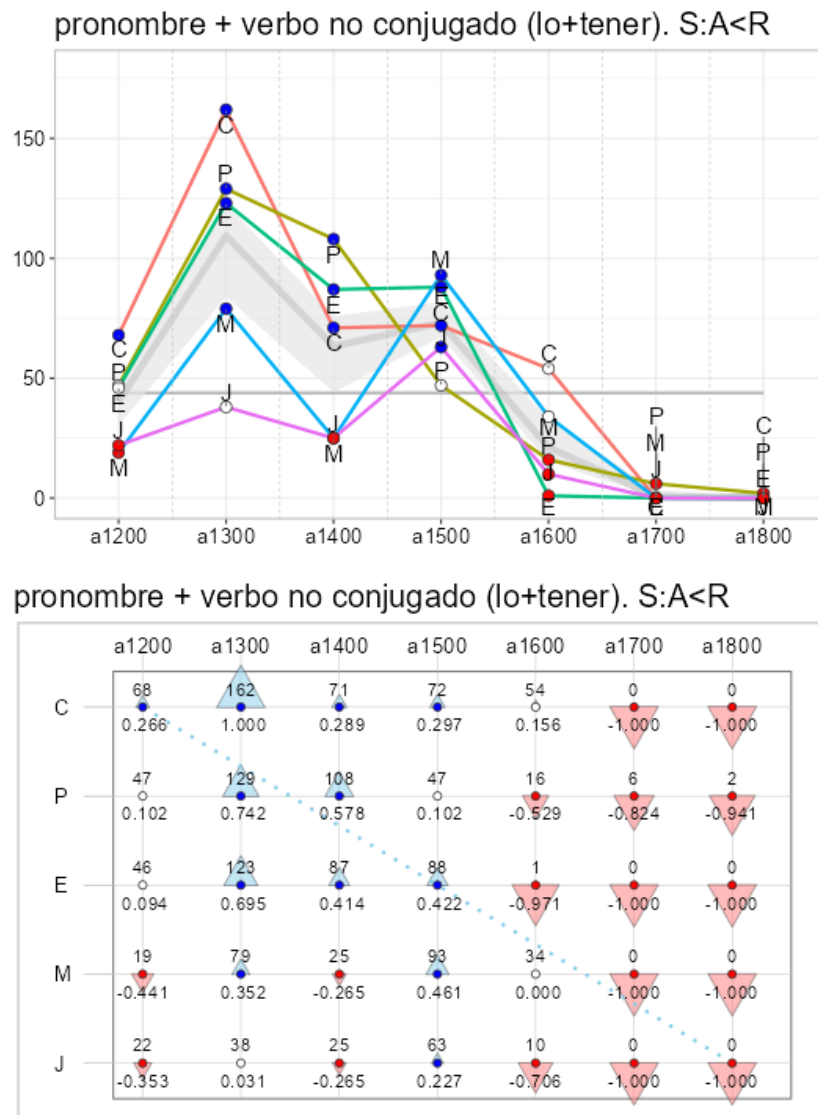


Fig. 14a, b. *Pronombre+verbo no conjugado. Sociedad: Año<Región.*

(C: documento canclleresco, E: d. eclesiástico, J: d. judicial, M: d. municipal, P: d. particular)

En a1300, C, P y E fueron más numerosos. En todo el conjunto, M y J tuvieron muy poca cantidad. Por lo tanto, <pronombre + verbo no conjugado> (*lo+tener*) puede considerarse un rasgo estilístico tradicional.

Consideración

Razones de la disminución de <pronombre + verbo no conjugado>

El <pronombre + verbo no conjugado> (*lo+tener*) se ha convertido en un término no marcado debido al abrumador aumento de la secuencia competidora <verbo no conjugado - pronombre> (*tener-lo*) (→3.3).

En comparación con <artículo definido / posesivo + nombre> (*la / mi casa*), la estructura de <pronombre + verbo no conjugado> (*lo + tener*) es parcialmente similar en que es <palabra átona + palabra acentuada>, por lo que su uso debería aumentar. Sin embargo, el verbo es un verbo no conjugado y por ello, es morfológicamente diferente de <artículo / posesivo + nombre (con flexión)> (*las / mis casas*), por lo que la acción analógica de <artículo / posesivo + nombre > no funcionaría.

La siguiente es una lista de combinaciones de categorías gramaticales de <palabra débil + palabra acentuada>.

p+n (60,604), ar_df+n (56,889), cl+v_cj (35,731), psa+n (34,872), c+v_cj (28,264), pt+n (27,783), c+n (27,397), p+t (22,063), ar_df+a_pa (21,411), pn_r+v_cj (18,411), pt+a_nm (14,634), p+pn_ps (14,318), p +np (13,979), pt+v_cj (13,486), p_ar+n (13,028), c+av (12,351), p+v_if (11,473), p+a_id (10,749), c+a_nm (10,618), pt+ av (10,199))... [Total: 643,601].

La mayoría de los elementos que aparecen en segunda posición en esta lista son nombres (n), verbos (v) y adjetivos (a). Estas son categorías gramaticales con flexiones morfológicas. Por otro lado, el segundo elemento de p+t, p+pn_ps, p+np, c+av, c+a_nm, pt+av no está flexionado, por lo que es similar a <pronombre + verbo no conjugado> (*lo+ tener*). Al parecer, el primero de estos elementos es una preposición (p) y una conjunción (c). Como se muestra a continuación, los miembros de estas dos categorías gramaticales no tienen formas similares a los pronombres:

Conjunciones: aunque, ca, como, conforme, conformes, confuso, conque, desde, e, empero, maguer, maguera, mas, mientras, ni, nin, o, pero, porque, pues, que, si, sino, ni siquiera, tú, y.

Preposiciones: a, ante, bajo, bajos, cabe, con, conlos, contra, de, desde, durante, en, enta, entre, entrada, excepto, hacia, hasta, mediante, para, por,

salvo, salvos, sin, entonces, sobre, tras, troa.

Pronombres: la, las, le, les, lo, los, me, mi, nos, os, se, se la, se las, se les, se lo, se los, te.

Por lo tanto, entre las combinaciones de categorías gramaticales de <palabra débil + palabra acentuada>, no hubo ninguna forma de alta frecuencia que actuara de manera analógica a <pronombre + verbo no conjugado> (*lo + tener*). Por ello, consideramos que el <pronombre + verbo no conjugado> (*lo + tener*) quedó marcado y desapareció, sin ningún apoyo analógico.

3.5. Verbo imperativo + pronombre (*ten-lo*)

Esta sección trata de los pronombres con formas imperativas del verbo como (*ten-lo*). La forma imperativa se conjuga dependiendo de la persona, pero a diferencia de la construcción verbo conjugado + pronombre (*tiene-lo*), aquella no ha disminuido a lo largo de la historia¹⁵.

Formas correspondientes : dátgela (19), quédate (13), dále (12), sírvase (12), dádgela (11), díme (7), sírbase (7), déle (6), póngase (6), dálos (5), dásele (5), llámese (4), quédese (4), véase (4), diémosgelo (3), dáme (3), dénos (3), dígame (3), díle (3), llámenle (3)... [Total: 303].

Dado que el número de apariciones de formas de palabras es pequeño, no es posible observar la distribución de frecuencia por región, edad o sociedad. Por ello, nos dedicamos solo a la siguiente consideración.

Consideración

Razones de la conservación de <verbo imperativo - pronombre>

La existencia de la conjugación del verbo en el <verbo imperativo - pronombre> según la persona es la misma que en el <verbo conjugado + pronombre>. Exploraremos a continuación la razón por la que el <verbo conjugado + pronombre> llegó a su fin y la "forma imperativa - pronombre" logró un

¹⁵ En raras ocasiones se encuentra la secuencia <pronombres + imperativos>: es tan escasa que no será tratada aquí como excepciones:: la embiad (3), lo cumplido (2), la entregad (1), lo dad (1) , lo embiad (1), lo enwiad (1), lo hazed (1), los executad (1), los rematad (1), vos curiat (1) [Total: 13].

desarrollo estable, a pesar de la similitud mencionada,

En primer lugar, es necesario confirmar las características sintácticas de la forma imperativa en latín. Cuando nos referimos a las secciones relevantes de los estudios de sintaxis latina clásica (Bassols de Climent 1992, Calero Calero 2003), además de ejemplos con verbos imperativos al comienzo de las oraciones, hay muchos otros ejemplos de oraciones donde los verbos imperativos no están al comienzo de las oraciones como los siguientes:¹⁶

de palla memento (Plauto) "Recuerda el manto." [Bassols de Climent 1992: 212] / hic hodie cenato (id.) "Cena aquí hoy." [id] / bonum animum habe (id.) "Ten buen ánimo" [id.] / Cras petito, dabitur. Nunc abi. (id.) "Pídelo mañana, se te dará. Ahora vete." [Calero Calero 2003: 117] / Laudate ingentia rura, exiguum colito (Virgilio) "Alabad los campos enormes, cultivad uno pequeño." [id.]

Por otro lado, todos los verbos imperativos en las oraciones de ejemplo enumeradas en libros especializados que tratan de la traducción latina bíblica de la Biblia (*Vulgata*, finales del siglo IV) se colocan al comienzo de la oración (Adachi 2014: 244-251, 17 ejemplos; Tabuchi 2006: 32 ejemplos). Aquí, la posición después de la coma (,) y la conjunción *et* en la oración también se incluye en el "principio de la oración" (La traducción moderna es de Reina y Valera 1960):¹⁷

Surge, et ascende Bethel, et habita ibi, facque altare Deo qui apparuit tibi quando fugiebas Esau fratrum tuum. "Levántate y sube a Bet-el y quédate allí; y haz un altar al Dios que se te apareció cuando huías de tu hermano Esaú." (Génesis: 35:1). [Adachi, *ibid.*] / Antecede populum, et sume tecum de senioribus Israel: et virgam, qua percussisti fluvium, tolle in manu tua, et vade. "Pasa delante del pueblo; y toma contigo de los ancianos de Israel; y toma también en tu manotú vara con que golpeaste el río, y ve." (Éxodo: 17, 5). [*id.*]

¹⁶ Cf. Higuchi y Fuzii (1963: 48-49): peccavi, veniam da, obsecro. "He pecado. Dame perdón, te lo ruego." / aqua gelida, inquit medicus, puero pedes lavato. "Con agua fría, dijo el médico, lava los pies del niño."

¹⁷ Por ejemplo, en *Medice cura te ipsum* "Médico, cúrate a ti mismo." (S. Lucas: 4. 23) [Tabuchi: 131], *Medice* está en caso vocativo, por lo que añadiendo una coma (,), *cura* se considera en el comienzo de la frase. De hecho, se añade una coma en *Domine, doce nos orare* "Señor, enséñanos a orar". (S. Lucas: 11, 1) [Tabuchi: 54]. El único imperativo en Tabuchi (*ibid.*) que no es del comienzo de una oración es el siguiente: *Si mihi non vultis credere, operibus credite*. "aunque no me creáis a mí, creed a las obras": 10, 38).

/ Benefacite his qui oderunt vos. "Benedicid a los que os maldicen." (S. Mateo: 5, 44) [Tabuchi: 37] / Considerate lilia agri quomodo crescunt, non laborant neque nent. "Considerad los lirios del campo, cómo crecen: no trabajan ni hilan." (S. Mateo: 6, 28). [*id.*:48]

Al buscar la forma imperativa (*dic* 'di', *dicite* 'decid')¹⁸ del verbo extremadamente frecuente *dicere* 'decir' en toda la *Vulgata*, encontramos 129 ejemplos. El desglose es: (a) 28 casos de nombres (en dativo), (b) 21 casos de pronombres (en dativo) y (c) 13 casos de nombres propios (en dativo), y (d) ad+nombre/pronombre/nombre propio (caso acusativo) son 14 ejemplos, todos los cuales aparecen después del verbo imperativo. Casi no hay ejemplos de que estos elementos aparezcan antes del imperativo¹⁹. Por lo tanto, aunque el material se limita a la forma imperativa de *dicere* en la *Vulgata*, creemos que la forma imperativa es básicamente seguida por todos los objetos indirectos, incluidos los pronombres. Los siguientes ejemplos son de Weber und Gryson (1969). Su traducción moderna es de Reina y Valera (1960):

- (a) 1. dicite filiae Sion ecce rex tuus venit tibi [...] "Decid a la hija de Sion: He aquí, tu Rey viene a ti, [...]" (S. Mateo: 21, 5). / 2. dic regi et dominatrici humiliamini sedete [...] "Di al rey y a la reina: Humillaos, sentaos en tierra; [...]" (Jeremías: 13, 18).
- (b) 1. dic nobis quando ista fient [...] "Dinos, ¿cuándo serán estas cosas?" (S. Marcos: 13-4). / 2. vade et dic eis revertimini in tentoria vestra. "Ve y diles: Volveos a vuestras tiendas." (Deuteronomio: 5, 30).
- (c) 1. dicite Deo quam terribile opus tuum [...] "Decid a Dios: ¡Cuán asombrosas son tus obras! [...]" (Salmos: 65, 3) / 2. et dicite Archippo vide ministerium quod accepisti in Domino ut illud impleas. "Decid a Arquipo: Mira que cumplas el ministerio que recibiste en el Señor." (A los Colosenses: 4, 17)
- (d) 1. ideo dic ad eos haec dicit Dominus Deus [...] "Diles, por tanto: Así ha dicho Jehová el Señor: [...]" (Ezequiel: 12, 23). / 2. tu itaque fili hominis dic ad filios populi tui iustitia iusti non liberabit eum in quacumque die peccaverit [...] "Y tú, hijo de hombre, di a los hijos de tu pueblo: La justicia del justo no

¹⁸ No hemos encontrado las formas futuras imperativas del latín clásico *dicito* y *dicitote*.

¹⁹ La única excepción es la siguiente: et Amasae dicite nonne os meum es et caro mea "Asimismo diréis a Amasa: ¿No eres tú también hueso mío y carne mía? [...]" (1º de Samuel: 19, 13).

lo libraré el día que se rebelare; [...]" (Ezequiel: 33, 12).

Estos imperativos suelen aparecer al principio de una frase (a-1, a-2, b-1, c-1: 20 casos en total) o después de la conjunción *et* (b-2, c-2: 31 casos). Incluso en otros casos, como en d-2, cuando el sujeto del imperativo se antepone, podemos suponer una pequeña pausa. De hecho, en las traducciones modernas se utiliza una coma (,). De este modo, se puede decir que la forma imperativa de *dicere* en la *Vulgata* aparece básicamente al principio de una oración o en una posición similar. Por otro lado, como vimos anteriormente, en el latín clásico hay muchos ejemplos en los que el imperativo va precedido de un objeto u otros elementos. Por lo tanto, en latín clásico, la forma imperativa del verbo se podía colocar libremente, pero en la versión latina bíblica de la *Biblia* parece haberse limitado al comienzo de las oraciones. Así, la aparición del imperativo al comienzo de una oración en latín bíblico parece indicar un "cambio desde abajo" en la lengua del pueblo. Se puede imaginar que cuando las personas del pueblo vulgo hacían una petición o daban una orden, expresaban primero el verbo y luego añadían el objeto o adverbio.

En lo siguiente, extraemos sólo el <verbo imperativo - pronombre> de los documentos notariales españoles, donde observamos su posición, distribución de frecuencia y tendencias históricas. Al mismo tiempo, indagamos en la razón histórica por la cual los pronombres se colocaron de manera única después del imperativo.

Como observamos anteriormente, la forma imperativa en latín bíblico parece haber aparecido básicamente al comienzo de una oración. De manera coherente, en los documentos notariales españoles, si nos fijamos en las partes gramaticales de las palabras que preceden al pronombre imperativo, la mayoría son signos de puntuación (pt) y conjunciones (c)²⁰. Y si nos fijamos en el desglose de las conjunciones, todas son conjunciones coordinantes, en su mayoría y y *e*:

Categoría gramatical de la palabra antecedente: pt (179), c (65), a_pa (26), n (12), pn_d (4), v_cj (4), np (3), v_if (3), av (2), a (1), p (1), pn_r (1), t (1), v_pv (1) [Total: 303] // Formas de puntuación inicial: [.] (124), [,] (43),] (4), “ (4), ” (2), ((1), - (1) [Total: 179] // Formas de conjunciones precedentes: y (43), e (20), mas (1), o (1) [Total: 65].

²⁰ La categoría gramatical más común de la palabra que aparece después de las conjunciones son los participios adjetivales (pa). La mayoría de ellos están en la construcción del participio pasado, *La carta leída dádgela*, por lo que, si se supone una pausa después de leída, se puede tratar como el comienzo de una oración.

La conjunción coordinante antes del imperativo al comienzo de una oración es, simplemente, un elemento añadido y es una palabra indicativa sin significado esencial. Por lo tanto, si nos situamos tanto después de un signo de puntuación como después de una conjunción coordinante como "el comienzo de una oración" (#), podemos pensar que la forma imperativa se usó principalmente al comienzo de una oración. De esta forma, la posición del imperativo que aparece en todo el documento notarial es al inicio de la frase, tal como ocurre en el latín bíblico.

El siguiente es un gráfico que muestra los cambios históricos en el uso de pronombres construidos con formas imperativas al comienzo de oraciones (#: después de los signos de puntuación y conjunciones coordinantes) y en otras posiciones (=: en medio de oraciones).

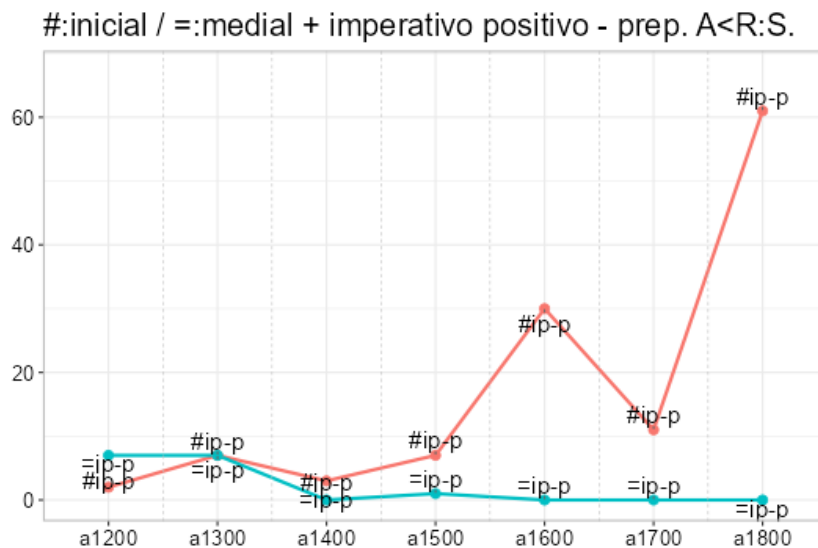


Fig. 15. *Tendencias cronológicas en <verbo imperativo - pronombre>*

Al observar este gráfico, parece que el <verbo imperativo - pronombre> se usaba al principio de una oración (#) y en medio de una oración (=) en el período medieval (A1200-11400), al igual que <verbo conjugado -pronombre> en general. Dado que el número de apariciones es pequeño, no se puede determinar con certeza. Por ello, dado que solo hay unos pocos casos, los hemos buscado todos y examinado sus contextos antecedentes, y hemos descubierto que son construcciones participiales, oraciones dependientes o temas, como se muestra a continuación:

Construcción de participio: e nós vista la carta diémosgelo (id-45, a1200) /

La carta leída dádgela (id-76, a1200) // **Oración subordinada:** E si assí non lo quisieren fazer tomadles (id-78, a1200) / ante que metades foz en las miesses fazétnoslo (id-150, a1200) // **Tema:** E lo que fiziéredes en esto embiátnoslo (id-136, a1300) / E los dichos maravedís dátgelos (id-537, a1400).

Dado que se supone una breve pausa después de estos precontextos, se puede decir que el imperativo inmediatamente siguiente estaba en una posición casi inicial de oración.

El siguiente gráfico muestra las tendencias en la frecuencia de los pronombres débiles que aparecen al principio de las palabras (después de una pausa) a lo largo del tiempo. Según este gráfico, los pronombres apenas aparecen al principio de las palabras entre a1200 y a1600. Por este motivo, pensamos que la forma imperativa que aparecía al principio de una palabra iba solo seguida de un pronombre regido.

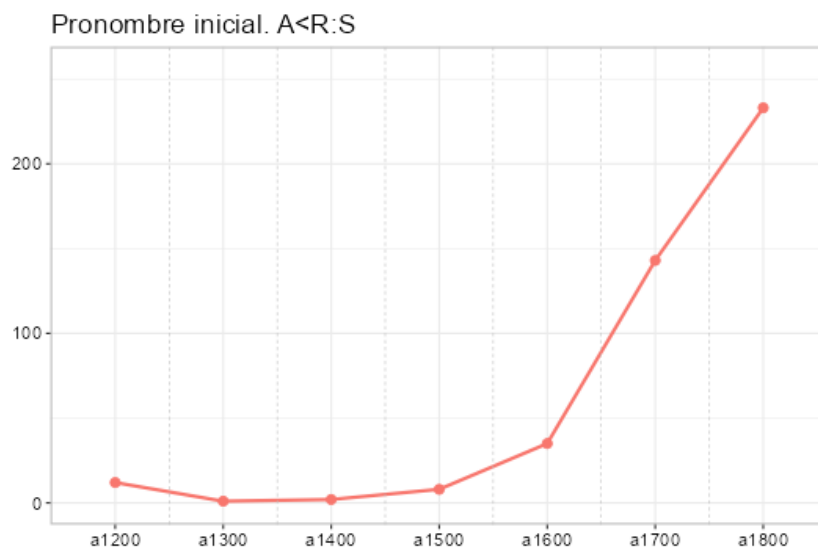


Fig. 16. *Pronombre débil que aparece al principio de una palabra (después de una pausa)*

Por otro lado, desde principios de la Edad Moderna, el uso del <verbo imperativo - pronombre> ha aumentado y se ha establecido, aunque solo sea al comienzo de una oración (#). Incluso en las lenguas modernas, la forma imperativa rara vez se usa en medio de oraciones, excepto marcadores del discurso como *mira* y *miré*, y generalmente se usa al comienzo de una oración después de una pausa, o al comienzo de una oración.

Por lo tanto, la razón por la cual un pronombre sigue al imperativo es que, en el latín bíblico, la forma imperativa se usa casi exclusivamente, y desde el latín

bíblico hasta el español medieval y el español moderno temprano y moderno, la forma imperativa aparece siempre en la posición inicial, o casi inicial, de la oración. Por lo tanto, pensamos que los pronombres débiles no podían ubicarse al comienzo de las oraciones sin apoyo de la palabra tónica. Con el tiempo, la estructura <verbo imperativo - pronombre> se convirtió en casi la única secuencia (→3.2) y se mantuvo estable, porque <pronombre + verbo imperativo> era casi inexistente.

De esta manera, la principal diferencia entre el <verbo imperativo - pronombre> (*ten-lo*) y el <verbo conjugado - pronombre> (*tiene-lo*) es que la primera construcción solo puede aparecer al principio de una oración, mientras que la segunda puede aparecer en cualquier lugar. Por consiguiente, el <verbo imperativo - pronombre> (*ten-lo*) se conserva en a1200-a1500, cuando el pronombre no aparece al principio de la oración, mientras que el <verbo conjugado - pronombre> (*tiene-lo*), como el pronombre va en la segunda posición de manera similar, se mantuvo igualmente en 1200-1500, pero se perdió en la posteridad (a1600-a1800) cuando el pronombre apareció al comienzo de la oración, ante la abrumadora mayoría del <pronombre - verbo conjugado> en competencia que se encontraba en las mismas posiciones libres (→3.2)²¹.

Dado que el imperativo negativo está precedido por *non* o *no*, el pronombre siempre se coloca antes del verbo con apoyo de estas formas negativas, lo que coincide con el patrón general de sintagmas nominales (→3.1).

3.6. Totalidad

En las secciones anteriores, hemos analizado cada una de las combinaciones de pronombre y verbo individualmente. Por lo tanto, en esta sección resumiremos y compararemos los resultados generales (p: pronombre, vc: verbo conjugado, vn: verbo no conjugado).

Cambios cronológicos

La siguiente figura muestra la frecuencia relativa vertical de cada caso ajustada por el valor promedio de las mediciones reales absolutas:

²¹ RAE y AALE (2009: 1207) tratan las siguientes fusiones de las formas imperativas y clíticas como una "integración de los pronombres enclíticos en la base a la que se adjuntan": *pongámos + nos > pongamonos*, *digamos + selo > digámoselo*, *callad + os > callaos*. Nosotros los trataremos como estructuras monolíticas (→3.5).

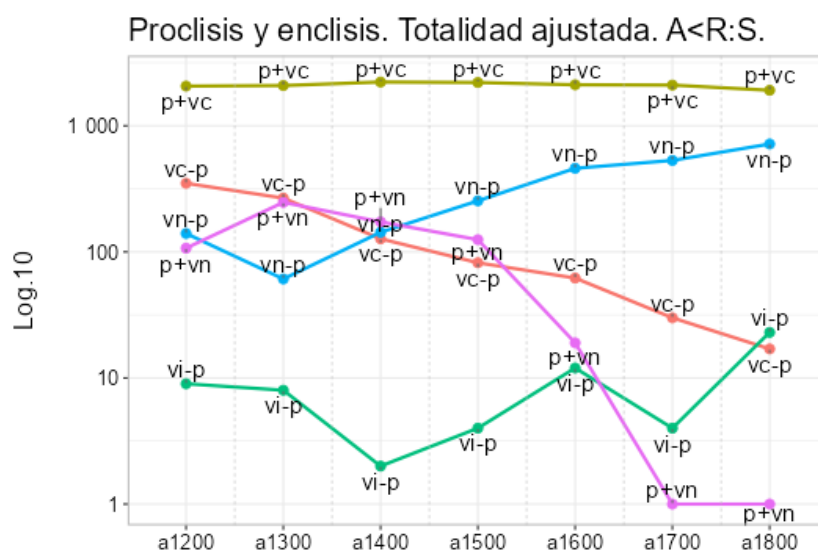


Fig. 17. Frecuencia relativa ajustada por media. Año<Sociedad: Región.
(p: pronombre, vc: verbo conjugado, vn: verbo no conjugado, vi: verbo imperativo)

Al observar la figura anterior con respecto al verbo conjugado (vc), <pronombre + verbo conjugado> (p+vc: *lo+tiene*) muestra consistentemente la frecuencia más alta, y <verbo conjugado - pronombre> (vc-p: *tiene-lo*) se encuentra en disminución. Por lo tanto, el cambio histórico general no es precisamente un cambio de la enclisis a la proclisis, sino más bien una disminución exclusiva de la enclisis, ya que la situación de la proclisis no ha cambiado, pues mantiene la supremacía absoluta.

El <verbo conjugado - pronombre> (vc-p: *tiene-lo*) es una forma antigua relativamente común en la Edad Media. Por ser una forma antigua, está ampliamente distribuida en la parte occidental peninsular (CV, CN, LE), donde se mantienen formas arcaizantes (Andrés Díaz 2013: 651) y se utiliza, principalmente, en documentos cancillerescos que usan un estilo de escritura ampuloso. Se usó con frecuencia en documentos reales (C), documentos judiciales (J) y documentos de la iglesia (E) (→2.2).

Si volvemos la mirada atrás en el tiempo, en el siglo XIX, apreciamos que Pérez Galdós, en *Marianela* (1878), también opta en raras ocasiones por la secuencia <verbo conjugado - pronombre> (vc-p) en oraciones conversacionales (lengua hablada). Esta estructura, por el contrario, era mayoritaria en las narraciones (lengua escrita). Se encuentra la misma secuencia (vc-p) ocasionalmente en el español moderno (después del siglo XX), pero nuevamente, se limita al lenguaje escrito arcaico (Keniston 1937: 68, Eberenz 2006: 174). Por

tanto, la disminución de las posposiciones de pronombres se considera un "cambio desde abajo" (Trask 2000, s.v. "change from below"), en el que se marcó una forma arcaica que permaneció en el español medieval.

Respecto a las formas no conjugadas (vn), en la época medieval (a1300-a1400), <pronombre + verbo no conjugado> (*lo+tener*) era más común que <verbo no conjugado - pronombre> (*tener-lo*), pero en la Edad Moderna (a1500-), las primeras disminuyeron y las segundas aumentaron por la monolexicalización, con lo cual se revirtió la situación. En este caso se reconoce un cambio histórico de la proclisis a la enclisis.

El <verbo imperativo - pronombre> se usa con poca frecuencia y es tan constante que casi no hay cambios cronológicos. Finalmente, en a1800, debido a la disminución y desaparición de <verbo conjugado - pronombre> (vc-p), <verbo imperativo - pronombre> (vi-p) ocupó el tercer lugar después del <pronombre + verbo conjugado> y <verbo no conjugado - pronombre>. Estas tres primeras estructuras han determinado el estado actual de la lengua moderna.

La siguiente tabla resume todo lo anterior:

Forma		Ejemplo	Posición
1.<pronombre + verbo conjugado>	p+vc	<i>lo+tiene</i>	1 (mayoría absoluta)
2.<verbo conjugado - pronombre>	vc-p	<i>tiene-lo</i>	2→4 (descenso lento)
3.<verbo no conjugado - pronombre>	vn-p	<i>tener-lo</i>	3→2 (ascendente)
4.<pronombre + verbo no conjugado>	p+vn	<i>lo+tener</i>	2→4 (minoría, descendente)
5.<verbo imperativo - pronombre>	vi-p	<i>ten-lo</i>	5→3 (frecuencia constante)

Tabla 7. *Cambios de la totalidad*

En general, 1.<p+vc> (*lo+tiene*) siempre ocupó una abrumadora mayoría, y 3.<vn-p> (*tener-lo*) casi siempre ascendió para quedar muy cerca de <p+vc> (*lo+tiene*); 2.<vc-p> (*tiene-lo*) se usó de forma relativamente estable con un descenso lento y 4. <p+vn> (*lo+tener*) se usó hasta a1500, luego se redujo en a1600 y terminó desapareciendo en a1700-1800. Por lo tanto, si lo miramos en su conjunto, las tendencias de 1.<p+vc> (*lo+tiene*), que ocupó siempre el primer lugar, y 3.<vn-p> (*tener-lo*), que ha aumentado constantemente, pueden reconocerse como las dos tendencias principales.

Según los dos gráficos anteriores, el conjunto se puede clasificar en los tres siguientes grupos:

- (1) <pronombre + verbo conjugado> p+vc (*lo+tiene*). Estable en gran medida. No marcada.
- (2) <verbo no conjugado - pronombre> vn-p (*tener-lo*). Ascendente. No marcada. Razones de “innovación lingüística”
- (3) <verbo conjugado - pronombre> vc-p (*tiene-lo*), <pronombre + verbo no conjugado> p+vn (*lo+tener*). Descendente. Marcada. Razones de "tradición lingüística"
- (4) <verbo imperativo - pronombre> vi-p (*ten-lo*) Constante. Sin marcación.

El siguiente es un resumen de la frecuencia (frecuencia absoluta), el patrón de sintagma nominal, la monolexicalización y la marca lingüística de las cuatro formas tratadas.

Forma	Ejemplo	Frecuencia	Patrón de sintagma nominal	Monolexicalización	Marca
1. p+vc	<i>lo+tiene</i>	35,986	V	-	no marcada
2. vc-p	<i>tiene-lo</i>	2,135	-	V (no completa)	marcada
3. vn-p	<i>tener-lo</i>	5,842	-	V (completa)	no marcada
4. p+vn	<i>lo+tener</i>	1,536	-	-	marcada
5. vi-p	<i>Ten-lo</i>	303	-	V (no completa)	sin marca

Tabla 8. *Frecuencia absoluta de las cinco combinaciones.*

(p: pronombre, vc: verbo conjugado, vn: verbo no conjugado, vi: verbo imperativo)

Entre las cinco formas de la tabla anterior, las dos formas más frecuentes no marcadas son 1.<pronombre + verbo conjugado> (*lo+tiene*) y 3.<forma no conjugada - pronombre> (*tener-lo*). Entre ellas, dado que el orden de las palabras de <pronombre + verbo conjugado> era <cambio parabólico + cambio flexivo>, coincidía con el patrón de <artículo definido / posesivo + nombre>, por lo que era estable y abrumadoramente abundante. Después de esto, 3. el <verbo no conjugado - pronombre> (*tener-lo*) ganó gradualmente poder a través de una monolexicalización completa. Por otro lado, 3. <verbo conjugado - pronombre> (*tiene-lo*) es un remanente de la forma sintáctica latina. Esto se debió a que no se convirtió completamente en una sola palabra, por lo que se marcó y la frecuencia disminuyó. El 4. <pronombre + verbo conjugado> (*lo+tener*) también es un resto de la forma sintáctica latina. Esto no coincidía con el patrón de las frases nominales en español y, como estaba precedido por un pronombre, era imposible convertirlo en una sola palabra, por lo que desapareció en a.1600. El 5. <verbo imperativo - pronombre> (*ten-lo*) aparecía sólo al principio de una oración y se mantenía

constantemente, aunque con reducida frecuencia.

Variación regional y social

A continuación, realizamos un análisis concentrado con los resultados de un análisis equilibrado doble utilizando variables regionales y sociales como variables objetivas y edad como variable de referencia. Dado que las variables regionales y sociales no asumen un orden específico, las concentramos en filas (regionales) y columnas (sociales).

(1)	*	C	E	J	M	P
EX	278	28	64	33	80	
CV	154	66	98	32	53	
LE	106	72	104	23	91	
AR	11	16	120	4	23	
AN	64	44	128	15	76	
CN	74	42	73	49	113	

verbo conjugado - pronombre (tiene-lo). R:S<A

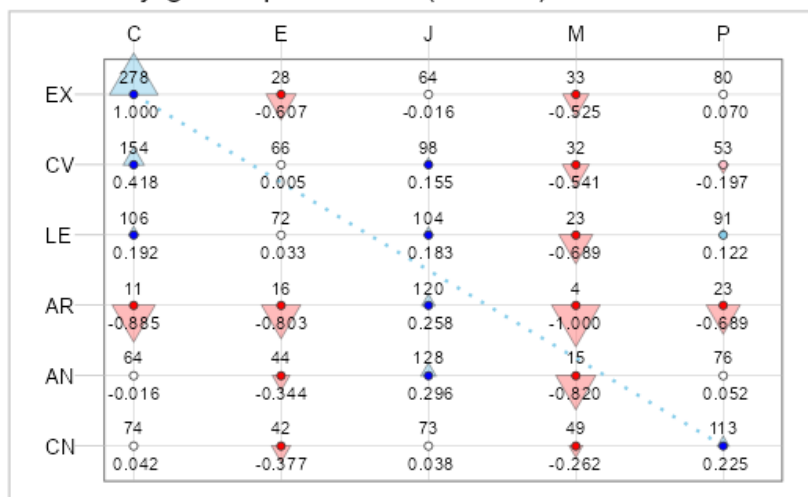


Fig. 18. *Región:sociedad < año, en distribución concentrada*

(AN: Andalucía, AR: Aragón, CN: Castilla la Nueva, CV: Castilla la Vieja, EX: Extremadura, LE: León)

(C: documento canclleresco, E: d. eclesiástico, J: d. judicial, M: d. municipal, P: d. particular)

En la zona superior izquierda de este mapa de distribución de concentración, [Región CV-EX: Sociedad C] forma un grupo grande y utiliza enclisis con frecuencia. Por otro lado, [AR-AN-CN] forma otro grupo. Creemos que el primero representa un antiguo estilo de escritura tradicional y el segundo representa un nuevo estilo de escritura popular.

La siguiente figura es el mismo diagrama de grupos de intervalos que el anterior "Región: sociedad < año, en distribución concentrada".

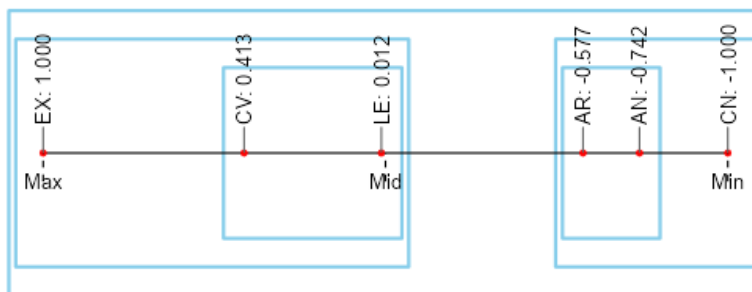


Fig. 19. *Región:sociedad < año, en distribución concentrada. Región*
 (AN: Andalucía, AR: Aragón, CN: Castilla la Nueva, CV: Castilla la Vieja, EX: Extremadura, LE: León)

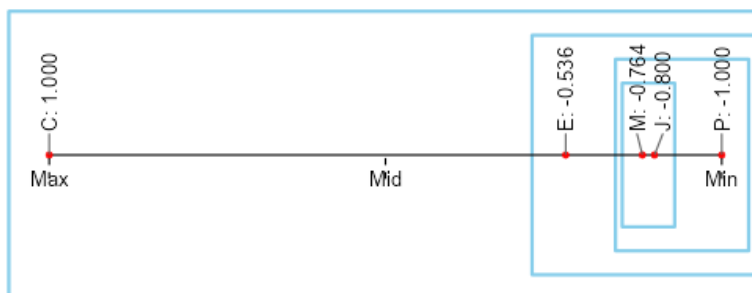


Fig. 20. *Región:sociedad < año, en distribución concentrada. Sociedad.*
 (C: documento cancilleresco, E: d. eclesiástico, J: d. judicial, M: d. municipal, P: d. particular)

La estructura de la variable regional es:

$$[[EX>[CV>LE]]>[[AR>AN]>CN]]$$

La estructura de la variable social es:

$$[C>[E>[[M>J]>P]]]$$

Ambas muestran una estructura endocéntrica continua. Esta sucesión de variables sociales parece indicar una transición estilística de la tradición a la innovación. Si esto es correcto, la misma gradación existe tanto en las regiones como en las sociedades. Concretamente, $[EX>[CV>LE]]$ (regiones noroestes) + C forma un grupo de región y sociedad tradicionales, por una parte, y $[[AR>AN]>CN]$ (regiones surestes) + $[E>[[M>J]>P]]$, otro grupo innovador.

4. Conclusión

Con lo anterior, hemos confirmado que la siguiente hipótesis presentada (2. Método) es válida.

Hipótesis: desde la Edad Media hasta los primeros tiempos modernos y hasta la actualidad, <pronombre + verbo conjugado> era el orden no marcado de las palabras. El orden de <verbo conjugado - pronombre>, heredado del latín, es un orden marcado de palabras y ha sobrevivido hasta hace poco en determinadas condiciones. Por otro lado, el orden correspondiente a <verbo no conjugado - pronombre> se ha desarrollado a manera diferente.

A continuación, se enumeran nuevamente las cuestiones de investigación planteados en 2. Método y las soluciones obtenidas en el texto principal se indican con "→".

Pregunta-1: ¿En las lenguas medievales (fuentes textuales: 1200-1500), la enclisis (*tiene-lo*) era más común y la proclisis (*lo+tiene*) se limitaba a ciertas condiciones?

→ Proclisis (*lo+tiene*) se usó tan frecuentemente en lenguas medievales como en lenguas modernas. Por otro lado, la enclisis (*tiene-lo*) aparecía bajo ciertas condiciones, como al principio de una palabra.

Pregunta-2: ¿Ha habido un cambio en la conjugación de los verbos de enclisis (*tiene-lo*) a proclisis (*lo+tiene*) a lo largo de la historia (en los documentos de 1200-1800)?

→ Desde el principio (1200), la proclisis (*lo+tiene*) fue constante y abrumadoramente popular. Por lo tanto, no hubo cambio en la transición de enclisis (*tiene-lo*) a proclisis (*lo+tiene*), sino más bien una disminución de la enclisis (*tiene-lo*). La razón por la cual la proclisis (*lo+tiene*) era abrumadoramente común es que la forma de <pronombre + verbo conjugado> es similar al patrón de <palabra débil + palabra acentuada>, especialmente el patrón de <artículo definido / posesivo + nombre>, ambos extremadamente frecuentes. Por otro lado, los factores que contribuyen a la disminución de la enclisis incluyen la singularidad de su forma y lo distintivo de su frecuencia de uso.

Pregunta 3: Por el contrario, ¿ha habido un cambio en la estructura del verbo

no conjugado de los verbos a lo largo de la historia, de proclisis (*lo+tener*) a enclisis (*tener-lo*)?

→ La enclisis (*tener-lo*) en la secuencia con el verbo no conjugado ha ido aumentando paulatinamente a lo largo de la historia. Una posible razón para esto es que <pronombre no conjugado> (*tener-lo*) forma una sola palabra, es decir, forma un patrón diferente del <pronombre + verbo no conjugado> de dos palabras (*lo+tener*).

Pregunta-4: ¿Cuál es la razón histórica del imperativo positivo enclítico (*ten-lo*), a diferencia del imperativo negativo?

→ La forma imperativa se usa al comienzo de una oración, después de una conjunción coordinante, o en una posición equivalente al comienzo de una oración, como después de una construcción participial, una oración subordinada o un sintagma nominal tematizado. En la posición inicial de oración o cuasi-inicio de oración, el pronombre débil no disponía de una palabra acentuada que lo respaldara, por lo que permanecía después del verbo (palabra acentuada), dando lugar al imperativo positivo enclítico. Dado que el imperativo negativo está precedido por *non* o *no*, el pronombre se coloca delante del verbo con apoyo del adverbio negativo.

La literatura anterior señala que el orden <verbo conjugado - pronombre> (*tiene-lo*) se usó normalmente en el período medieval y que era más común que en el período moderno temprano, pero la frecuencia de <pronombre + verbo conjugado> (*lo + tiene*) no fue considerada (excepto Nakaoka 1993: 40). En este estudio, comparamos el <verbo conjugado - pronombre> medieval (*tiene-lo*) con el <pronombre + verbo conjugado> (*lo+tiene*), que prevalecía en ese momento. Además, hemos llegado a la conclusión de que el <pronombre + verbo conjugado> (*lo + tiene*) no tiene una relación significativa con la posición dentro la oración, sino que dicho orden puede verse como parte de la supremacía general en el número de <palabras débiles + palabras acentuadas>, especialmente del sintagma nominal <artículo definido / posesivo + nombre>.

Por otro lado, si consideramos el <verbo no conjugado - pronombre> (*tener-lo*) no como una serie de dos palabras, sino como una sola palabra fusionada, podemos ver que no constituye un contraejemplo del patrón general de <palabra débil>. + <palabra acentuada>. La base para considerar el <verbo - pronombre> como una sola palabra es la fusión de verbo y pronombre, oposiciones sintácticas

como *Lo voy a leer y resumir / Voy a leerlo y [resumirlo / *resumir]*, y la interpolación medieval como *les yo fago* ocurrida en la proclisis, pero no en la enclisis, y la escritura medieval continua de la enclisis y la separada de la proclisis.

El orden de las palabras del <verbo imperativo - pronombre> (*ten-lo*) era una característica sintáctica en la posición inicial de la oración en el latín bíblico y fue heredada por el español medieval. Esto se debe a que en la Edad Media era difícil colocar el pronombre débil al comienzo de una oración. Sin embargo, en la Edad Moderna, se hizo posible colocar pronombres débiles al comienzo de las oraciones, por lo que el <pronombre + verbo conjugado> (*lo + tiene*), que anteriormente había aparecido principalmente en medio de oración, comenzó a aparecer al inicio de las oraciones también. Por otro lado, se mantuvo el orden de las palabras del pronombre imperativo (*ten-lo*). La razón de esto estriba en que los pronombres rara vez aparecen al comienzo de las oraciones en el español medieval, por lo que casi no había <pronombres + formas imperativas> en la misma posición.

En general, el cambio de idioma no implica una alternancia completa de los términos binarios [A]→[B], sino más bien en tal secuencia como [A]→[AB] (coexistencia: A es dominante)→[BA] (coexistencia: B es dominante)→[B]. Por lo tanto, el intercambio completo de <verbo imperativo - pronombre> → <pronombre + verbo imperativo> era imposible. Tampoco tuvo lugar el proceso de [A]→[AB] (coexistencia: donde A es dominante), debido a que se le dio prioridad al verbo (imperativo) sobre el objeto (pronombre) en la expresión directa del imperativo en el lenguaje hablado, y junto con la influencia analógica de la monolexicalización del <verbo no conjugado - pronombre>, el <verbo imperativo + pronombre> se había establecido, por lo que no había lugar para una forma invertida, el <pronombre + verbo imperativo> (**lo+ten*).

Como resultado de investigar todos los materiales, descubrimos que, a lo largo de la historia, la estructura <pronombre + verbo conjugado> (*lo + tiene*) fue abrumadoramente común, y que el <verbo conjugado - pronombre> (*tener-lo*) gradualmente se volvió más frecuente. Se confirmaban ambas posibilidades como las dos tendencias principales. Estas dos formas quedaron no marcadas y determinaron la distribución exclusiva de las palabras modernas. Por otro lado, el <verbo conjugado - pronombre> (*tiene-lo*), marcado en oposición con el término no marcado de <pronombre + verbo conjugado> (*lo + tiene*), y el <pronombre + verbo conjugado> (*lo+tener*), marcado en oposición con el término no marcado de <verbo conjugado - pronombre> (*tener-lo*), disminuyeron gradualmente y casi

desaparecieron en el siglo XIX (a1800). Al observar las variables regionales, advertimos que la supervivencia de estas marcadas formas arcaicas es cronológicamente característica de la Edad Media, geográficamente de las partes occidental y norte de la península, y socialmente característica de los documentos reales.

Al analizar cada formato, tomamos en consideración cada tipo de la marca (no marcado / marcado). Esto se debe a que la frecuencia de cada formato difiere de manera extrema. Si se aceptan dichas marcas, la descripción del sistema lingüístico se limita a mostrar las características de los términos marcados y no es necesario describir específicamente los términos no marcados. Por tanto, el término marcado del verbo conjugado es el <verbo conjugado - pronombre> (*tiene-lo*), y sus características más importantes son los hechos de que es forma antigua y que aparece preferentemente en posición inicial en la oración; también, que se concentran en los documentos de CV, LE y EX; C (documento cancilleresco) en la Edad Media.

Por otro lado, así mismo es término marcado el <pronombre + verbo no conjugado> (*lo+tener*), y su mayor característica es que es una <forma antigua> que se remonta a principios del periodo medieval (a1300, a1400) y que se concentran en LE, CV, CN en documentos C, E, P. Las características de estos términos marcados no eran las de formas de palabras recién creadas, sino que funcionaban como condiciones para la supervivencia de las estructuras antiguas.

Las conclusiones alcanzadas en este estudio difieren notablemente de las opiniones de estudios anteriores. Esto puede deberse a que estudios anteriores se basaron mayoritariamente en obras literarias de estilo tradicional, mientras que la presente investigación se ocupó de documentos notariales que reflejan la sociedad lingüística de la época, del lugar y de la sociedad. Aunque las obras literarias son importantes en el estudio de la historia de la lengua, debemos prestar especial atención en las lenguas utilizadas en los contextos de la realidad lingüística.

En este estudio, analizamos los problemas de la forma y estructura del lenguaje mientras observamos la relación con variables históricas, regionales y sociales. Las características lingüísticas relacionadas con la historia, la región y la sociedad se entendieron en la transición importante de la tradición (Edad Media, parte occidental de la península, documentos de estilo elevado) a la innovación (Edad Moderna, parte oriental de la península, documentos de estilo popular). Esta transición es un cambio unificado que es común entre diferentes áreas como la

lengua, la historia, la región y la sociedad.

Estos conocimientos históricos son útiles para comprender el estado actual del español moderno. Por ejemplo, proporcionará una explicación histórica de razones por las que los pronombres españoles se colocan delante y después de los verbos conjugados, infinitivos, gerundios e imperativos positivos. Además, al rastrear la historia del lenguaje, queda clara la razón por la que los pronombres aparecieron detrás de verbos conjugados, solo al comienzo de las oraciones hasta hace relativamente poco tiempo. La historia de la lengua en esa época no será una línea recta unidimensional que simplemente mide la progresión del tiempo, sino que tendrá que ser multidimensional, incluyendo las diferencias regionales y sociales.

Saussure (1916, 1972) hizo una clara distinción entre sincronía y diacronía para evitar confusión entre las dos. Sin embargo, la forma actual del lenguaje está formada por cambios y variaciones en una larga y vasta historia, región y sociedad. Por tanto, es necesario no solo un análisis estrictamente sincrónico y unidimensional sino también diacrónico y multidimensional. Esto se debe a que resulta sugerente observar aspectos diacrónicos y multifacéticos incluso en la investigación del lenguaje sincrónico de determinadas regiones y sociedades.

5. Referencias

- Alvar, Manuel / Pottier, Bernard. 1983. *Morfología histórica del español*. Madrid. Gredos.
- Andrés Días, Ramón. 2013. *Gramática comparada de las lenguas ibéricas*. Gijón. Ediciones Trea.
- Bassols de Climent, Mariano. 1992. *Sintaxis latina*. 10a ed. Madrid. Consejo Superior de Investigaciones Científicas.
- Calero Calero, Francisco. 2003. *Sintaxis latina*. Madrid. Universidad Nacional de Educación a Distancia.
- Castillo Lluch, Mónica. 1996. *La posición del pronombre átono en la prosa hispánica medieval*. Tesis doctoral. Universidad Autónoma de Madrid.
- Corominas, Joan. 1976. *Breve diccionario etimológico de la lengua castellana*. Madrid. Gredos.
- Eberenz, Rolf. 2000. *El español en el otoño de la Edad Media. Sobre el artículo y*

- los pronombres*. Madrid. Gredos.
- Ernout, Alfred / Meillet, Antoine. 2001. *Dictionnaire étymologique de la langue latine*. Paris. Klincksieck.
- Fukushima, Notitaka. 福嶋教隆. 2021. 『詳説スペイン語文法』東京. 白水社. *Gramática detallada de español*. (en japonés) Tokio. Hakusuisha.
- García de Diego, Vicente. 1970. *Gramática histórica española*. 3a ed. Madrid. Gredos.
- García Yebra, V. e H. Escolar. 1980. *César, Guerra de las Galias I*. Madrid. Gredos.
- Hanssen, Federico. 1913. *Gramática histórica de la lengua castellana*. Halle. Max Niemeyer.
- Higuchi Katsuhiko y Noboru Fuzii. 1963. 樋口勝彦・藤井昇. 1963. 『詳解ラテン文法』東京. 研究社. *Gramática latina. Explicación detallada*. (en japonés) Tokio. Kenkyusha.
- Keniston, Hayward. 1937a. *The Syntax of Castilian Prose. The Sixteenth Century*. The University of Chicago Press.
- Keniston, Hayward. 1937b. *Spanish Syntax List. A Statistical Study of Grammatical Usage in Contemporary Spanish Prose on the Basis of Range and Frequency*. Henry Holt and Company.
- Lapesa, Rafael. 1982. *Historia de la lengua española*. Madrid. Gredos.
- Lausberg, Heinrich. 1973. *Lingüística románica. Morfología*. Madrid. Gredos.
- Luraghi, Silvia. 2016. "The position of Weak Pronouns in Latin prose", *Latinitatis Rationes: Descriptive and Historical Accounts for the Latin Language / Paolo Poccetti (ed. lit.)*, 2016, 246-261.
https://www.academia.edu/5774218/_The_position_of_weak_pronouns_in_Classical_Latin_prose_Submitted
- Menéndez Pidal, Ramón. 1976. *Cantar de Mio Cid. Texto, gramática y vocabulario. Primera parte. Crítica del texto - gramática*. Madrid. Espasa-Calpe.
- Navarro Tomás, Tomás. 1972. *Manual de pronunciación española*. 19a ed. Madrid. Consejo Superior de Investigaciones Científicas.
- Nakaoka, Shoji. 中岡省治. 1993. 『中世スペイン語入門』東京: 大学書林. *Introducción al español medieval*. Tokio. (en japonés) Daigakusgorin.
- Ordóñez, Francisco. 2012. "Clitics in Spanish", in José Ignacio Hualde, Antxon Olarrea and Erin O'Rourke (ed.). *The Handbook of Hispanic Linguistics*. West Sussex, UK. Wiley-Blackwell.
- Palmer, L. R. 1974. *Introducción al latín*. Barcelona. Editorial Ariel.
- Penny, Ralph. 2006. *Gramática histórica del español*. Barcelona. Ariel.
- Pinkster, Harm. 1995. *Sintaxis y semántica del latín*. trad. M. Esperanza Torrego,

- Jesús de la Villa. Madrid. Ediciones Clásicas.
- Quilis, Antonio. 1981. *Fonética acústica de la lengua española*. Madrid. Gredos.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
<<https://www.R-project.org/>>.
- RAE y AALE. Real Academia Española y Asociación de Academias de la Lengua Española. 2009. *Nueva gramática de la lengua española. Morfología y sintaxis I*. Madrid. Espasa Libros.
- RAE y AALE. Real Academia Española y Asociación de Academias de la Lengua Española. 2010. *Ortografía de la lengua española*. Madrid. Espasa Libros.
- Reina, Casiodoro de y Cipriano de Valera. 1569, 1602, 1960. *La santa biblia. Antiguo y nuevo testamento*. Buenos Aires. Sociedades bíblicas en América Latina.
- Saussure, Ferdinand de. 1916: 1971. *Curso de lingüística general*. (trad.) Amado Alonso. Buenos Aires: Editorial Losada.
- Trask, R. I. *The Dictionary of Historical and Comparative Linguistics*. Edinburgh University Press.
- Ueda, Hiroto. 上田博人. 2021. 「スペイン語の語尾と強勢：辞書・大規模コーパス・文体・社会地理・歴史地理資料の分析」 『地理言語学研究』 1, pp. 51-105. "Terminaciones de palabras y acentuación en español. Análisis de diccionarios, corpus grande, materiales estilísticos, sociogeográficos, históricos y geográficos" *Investigación geolingüística*. 1 (en japonés), pp. 51-105.
<https://zenodo.org/records/5529282>
- Ueda, Hiroto. 2024. "Posición del pronombre átono español. Observaciones del corpus CODEA en busca de razonamientos lingüísticos y sociales", *Jornadas CHARTA 2024. Herramientas y recursos digitales para el desarrollo de corpus históricos*, 2024/2/29-3/1, Universidad de Alcalá.
<https://h-ueda.sakura.ne.jp/kenkyu/rekisi/clitico/clitico.pdf>
- Weber, Robert und Roger Gryson. 1969. *Biblia sacra. Iuxta vulgatam verdionem*. Stuttgart. Deutche Bibelgesellschaft.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://doi.org/10.1007/978-3-319-24277-4_9
- Yamamoto, Keiko. 1997. "Placement of Object Personal Pronouns in the New Testament Greek, Latin and Old English. research".
<https://core.ac.uk/outputs/143631193/?source=oai>

6. Apéndice. Métodos estadísticos

6.1. Desviación desde mediana

Al explorar las características de una matriz de datos numéricos, debemos prestar atención no solo a los números grandes, sino también a los números pequeños. Esto se debe a que es necesario observar, por un lado, que "una forma de palabra se usa con frecuencia" en el conjunto y, por otro, que "una forma de palabra no se usa" o que "la tasa de uso es baja". En este caso, es más fácil de entender si se relaciona la "pequeñez" de un número pequeño con el "tamaño" de un número grande (en la misma escala) y se indica con un número negativo. Y si el rango se establece en $[-1, 1]$, se puede realizar una evaluación dentro de una escala estandarizada.

La mediana indica el punto central que divide los datos ordenados, por lo que, si se resta la mediana de los datos desde los mismos datos, los números mayores que la mediana serán positivos y los números menores que la mediana serán negativos. La mediana, así, se convierte en cero y la cantidad de números positivos y la de los negativos será igual. Para cada valor de datos así transformados, los números positivos divididos por el valor máximo están en el rango $[0, 1]$; por otro lado, al dividir los números negativos por el valor absoluto del valor mínimo el resultado es un rango de $[-1, 0]$. Este método lo denominamos "Desviación desde mediana".

Las siguientes líneas muestran los procesos para obtener la desviación desde mediana de un ejemplo simple:

```
(2) V=c(1,3,5,8,9)
     W=V-Median(V); W # -4 -2  0  3  4
     ifelse(W>0,W/Max(W),W/abs(Min(W)))
     # -1.00 -0.50  0.00  0.75  1.00
```

La mediana del vector numérico $V = c(1,3,5,8,9)$ es 5. Restar 5 de V da W $(-4, -2, 0, 3, 4)$. En este momento, el valor mediano de V , 5, se vuelve 0 en W . Para convertir el vector entero para que el valor máximo de W sea 1, dividimos los números $(3, 4)$ que exceden 0 por el valor máximo de W , 4, por lo que $(3, 4)$ se convierte en $(0,75, 1)$. Los números debajo de la mediana $(-4, -2)$ se dividen por el valor absoluto 4 del valor mínimo de W (-4) , lo que da como resultado $(-1, -0,5)$. En este caso, si dividimos por el valor mínimo de W , -4 , el resultado será $(1, 0,5,$

0), por lo que dividiremos por el valor absoluto de W, 4, en lugar del valor mínimo de W, -4.

Aplicación

El siguiente gráfico es el resultado del análisis equilibrado (→6.7.4) de <pronombre + forma conjugada> (lo + tiene). Cada celda contiene la frecuencia normalizada (→6.7.1) y la desviación desde mediana para la región y la edad. El color de los puntos indica el grado de significatividad (→6.6) y el tamaño de los triángulos obedece a la desviación desde la mediana.

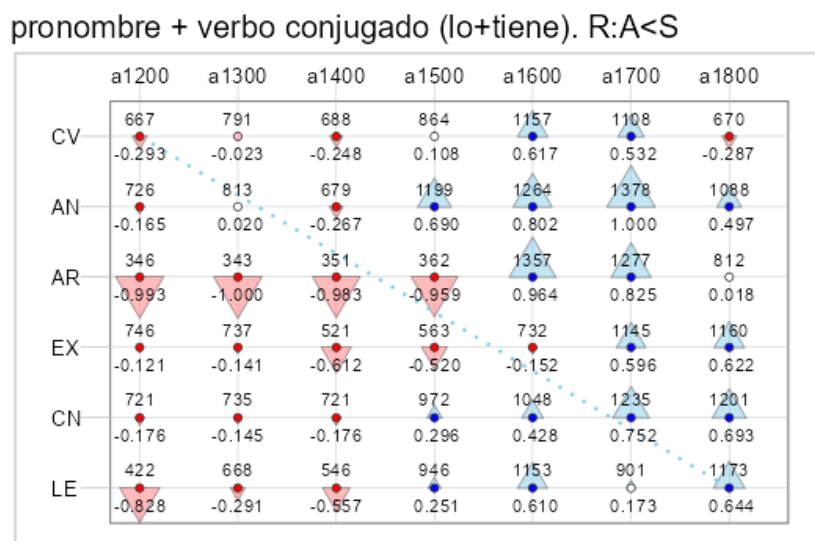


Fig. 21. <Pronombre + verbo conjugado> Región:Año<Sociedad

6.2. Frecuencia ajustada por media

Al calcular la frecuencia relativa o la frecuencia normalizada (→6.7.1), se utiliza un cierto multiplicador para expandir el valor numérico. Por ejemplo, la frecuencia relativa es un número pequeño, $30/100 = 0,3$, pero para convertirlo a porcentaje, se multiplica por 100, por ejemplo, $0,3 * 100 = 30$ por ciento. Para frecuencias normalizadas, se seleccionan multiplicadores como 1 millón (10^6) o 100.000 (10^5). En este momento, es deseable utilizar un multiplicador que haga que el resultado sea un número cercano a la escala del valor absoluto de la medición real original.

Por lo tanto, en lugar de usar un multiplicador como 10^2 , 10^5 o 10^6 , podemos usar un valor que haga que el valor promedio de la frecuencia relativa o la frecuencia normalizada sea igual al valor promedio de los valores absolutos

observados. Entonces, la escala de la matriz numérica resultante va a ser igual a la escala del valor de medición real. Tal multiplicador (m) se obtiene de la siguiente manera:

$m = \text{media de las frecuencias absolutas} / \text{media de las frecuencias relativas (normalizadas)}$

La frecuencia ajustada de esta manera la denominamos "Frecuencia ajustada por media", que incluye "Frecuencia relativa ajustada por media" y "Frecuencia normalizada ajustada por media".

A continuación, presentamos la tabla cruzada por región y año de frecuencia observada absoluta de <forma conjugada - pronombre>, seguida de la media (mo):

(3)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
AN		61	12	15	29	23	67	15
AR		4	20	28	6	27	21	1
CN		58	80	51	216	139	74	16
CV		314	121	111	56	18	11	5
EX		4	36	20	146	55	11	1
LE		60	99	42	14	6	38	4
# mean (mo)		50.8333						

La siguiente es la misma tabla con media de frecuencia normalizada:

(4)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
AN		272	93	76	77	136	87	50
AR		29	50	46	33	164	89	10
CN		331	216	100	141	91	98	42
CV		316	191	130	81	100	54	20
EX		278	152	127	159	35	106	9
LE		242	326	84	69	69	164	28
# mean (ms)		118.3571						

A partir de estas dos tablas y medias, calculamos la tabla cruzada de frecuencia normalizada ajustada por media:

(5)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
AN		117	40	33	33	58	37	21
AR		12	21	20	14	70	38	4
CN		142	93	43	61	39	42	18
CV		136	82	56	35	43	23	9

EX	119	65	55	68	15	46	4
LE	104	140	36	30	30	70	12
# mean (ma)	50.80952						

Por ejemplo, la frecuencia ajustada por media de AN:a1200 (=117) se obtiene de la siguiente forma: 272* mo / ms. De esta manera, la escala de frecuencia normalizada ajustada por media coincide con la de los valores observados (frecuencia absoluta), por lo que la frecuencia ajustada por media indica una situación realista. Por lo tanto, las frecuencias normalizadas que no realizan un ajuste por media no permiten evaluar la magnitud absoluta de los valores numéricos, pero el uso de la frecuencia ajustada por media lo hace posible. Esto es, por así decirlo, una "absolutización realista de los valores relativos".

Aplicación

Vamos a sumar y comparar todas las combinaciones de pronombres y verbos. La siguiente figura muestra la frecuencia normalizada de cada caso ajustada por la media de las frecuencias reales.

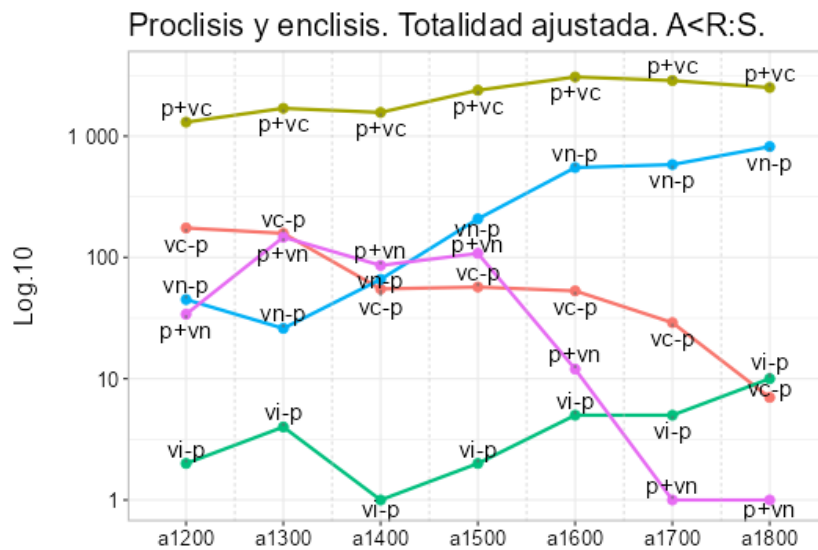


Figura 22. Cambio general (frecuencia normalizada ajustada por media).
(p: pronombre, vc: verbo conjugado, vn: verbo no conjugado, vi: verbo imperativo)

6.3. Puntuación contrastiva

Cuando aparece una forma lingüística, dadas las condiciones posibles (x) y la forma que aparece (y), cambiando cada condición (x) y buscando <pronombre + forma conjugada> (y), calculamos los siguientes valores: a, b, c, d.

a (x+, y+): número de casos en los que se busca (y) cuando existe la condición (x)

b (x+, y-): número de casos en los que (y) no se encuentra bajo la condición (x)

c (x-, y+): número de casos en los que se busca (y) sin satisfacer la condición (x)

d (x-, y-): número de casos en los que no se encuentra (y), cuando la condición (x) no está presente

Finalmente, cuantificamos la efectividad de la condición (x) utilizando el siguiente coeficiente (r):

$$r = (a - b) / (a + b)$$

Un valor como este r, donde el numerador es la diferencia entre los dos términos y el denominador es la suma de los dos términos, lo denominamos "Contraste". El rango del contraste es [-1, 1]. Cuando a=0, r presenta el valor mínimo (-1), cuando b=0, el valor máximo (1), y cuando a=b, presenta el valor intermedio (0). Cuanto mayor sea el valor de contraste r, más fuerte será la relación entre la condición y el resultado. Cuanto más cerca esté el valor de contraste r de 0, menor será la relación. Además, cuanto más cerca esté el valor de contraste r de -1, más fuerte será la relación inversa.

En este r, hemos utilizado solo a y b, pero también se han ideado otros coeficientes (Romesburg 1989, 1992: 177-202). Estos se denominan "Puntuación asociativa".

$(a+d) / (a+b+c+d)$ # Coincidencia simple

$a^2 / (a^2+b+c)$ # Dados

$a / (a+b+c)$ # Jaccard

como # Russel y Rao

$a^m / (a^m+b+c+d)$ # Russel & Rao.m

$((a+d)-(b+c)) / ((a+d)+(b+c))$ # Hamann

$(a*d-b*c) / (a*d+b*c)$ # Navidad

$(a*d-b*c) / \sqrt{(a+b)*(c+d)*(a+c)*(b+d)}$ # Phi

$a / \sqrt{(a+b)*(a+c)}$ # Ochiai

$(a^2-b-c) / (a^2+b+c)$ # Ueda (nuestra propuesta)

$\log_2(a*s / (a+b) / (a+c))$ # Información mutua

$1 / \sqrt{a*(a-(a+b)*(a+c) / s)}$ # puntuación t

Aplicación

El siguiente es el resultado de calcular la puntuación contrastiva (r) de <pronombre + forma conjugada> (*lo + tiene*) (y) comparando las condiciones (x) del comienzo de la oración (#) y de la posición intermedia de la oración (=). El número total de palabras es la suma de las combinaciones de pronombres y verbos relacionados (los datos se extraen de los dos órdenes de palabras: (*lo+tiene*) y (*tiene-lo*) de a1200-a1400):

(6)	s. x	s. y	x	y	a	b	c	d	s	r
#	c1, v_cj		36	9,819	19	17	9,800	1,119	10,955	0.056
=	c1, v_cj	10,919	9,819	9,800	1,119	19	17	10,955	0.795	

Tabla 9. Relación entre el comienzo de una oración (#), la posición medial de una oración (=) y <pronombre + verbo conjugado> (*lo+tiene*)

Este resultado muestra que es relativamente menos probable que <pronombre + verbo> (*lo+tiene*) aparezca cuando la condición está al principio de una oración (#), pero es más probable que aparezca cuando la condición está en el medio de una oración (=). Sin embargo, el orden de las palabras <pronombre + verbo> (*lo+tiene*) era aún más común al comienzo de las oraciones (=19) que el orden inverso de las palabras (=17).

Además, hay casos en los que la condición de "al principio de la oración" (#) aparece, aunque el resultado (y) no se cumpla (c). También hay muchos casos en los que no se cumple la condición y el elemento no aparece (d). Sin embargo, en este caso, hay una gran cantidad de (c) y (d), y si se utiliza un coeficiente que los incorpora, el resultado estará muy influenciado por (c) y (d), y la relación original entre (a) y (b) se volverá invisible. Más bien, es mejor cambiar las condiciones y observar los valores simétricos de (a) y (b) en la posición medial de oración (=).

6.4. Media mayoritaria

El siguiente es un ejemplo de datos (Data) con 4 filas y 5 columnas, y la salida (Stat) del valor promedio (*mean*) y mediana (*median*: el valor en el centro de los datos ordenados)²² de las filas.

²² DLE (s.v. *mediana*): “f. Mat. Elemento de una serie ordenada de valores crecientes de forma que la divide en dos partes iguales, superiores e inferiores a él.”

(7)	[Data]	c1	c2	c3	c4	c5	[Stat]	mean	median
1	r1	1	2	3	4	10	r1	4	3
2	r2	1	2	3	4	100	r2	22	3
3	r3	1	2	3	4	200	r3	42	3
4	r4	1	2	3	4	300	r4	62	3

Cada fila de estos datos (*Data*) contiene valores atípicos cada vez más grandes (columna: c5: 10, 100, 200, 300). Es bien sabido que la media se ve afectada por los valores atípicos. Es cierto que la media de los datos con grandes valores atípicos está significativamente sesgada hacia los valores atípicos (r2, r3, r4). No se puede decir que la media de dichos datos sesgados sea representativa de los datos. Por ejemplo, la media (=62) de datos como en r4 se desvía mucho de la mayoría de los datos (1, 2, 3, 4).

Por otro lado, la mediana no se ve afectada por valores atípicos, por lo que para datos con valores atípicos se recomienda utilizar la mediana en lugar de la media como valor representativo para indicar el centro (por ejemplo, Morales Vallejo 2008: 38, Vickers 2010: 4-5). Sin embargo, el valor de la mediana permanece constante, independientemente de cómo cambien los números a su izquierda y derecha, por ejemplo, (3, 5, 30, 32, 35), (10, 12, 30, 42, 45), (10, 20, 30, 420, 450), etc. Por ello, no se puede decir que dicha mediana indique adecuadamente el centro de todos los datos. Además, se selecciona uno de los dos valores centrales, la media y la mediana, dependiendo de la presencia o ausencia de valores atípicos, pero como la magnitud de los valores atípicos es continua, el establecimiento del umbral entre el uso de la media y el de la mediana se vuelve arbitrario.

La "Media recortada" (*trimmed mean*) se ideó como un método para reducir la influencia de los datos atípicos. Dado que la media recortada elimina los valores mínimos y máximos, se pierde información sobre los elementos eliminados, puesto que no se perciben los valores recortados. Al igual que la mediana, el uso de la media recortada es arbitrario.

Por lo tanto, en lugar de recortar los valores mínimos y máximos, recortamos gradualmente los grupos ordenados que contienen más de la mitad de los datos de la izquierda, encontramos el valor promedio en cada etapa y luego calculamos el valor promedio desde la suma. La media obtenida de esta manera la denominamos "Media mayoritaria" (*major mean; M.mean*). A continuación, se muestra su proceso utilizando datos (1, 2, 4, 5, 55) ordenados en orden ascendente (NA: *not available*, no disponible: no incluido en el cálculo). Como el número de

datos es 5, la mayoría es 3 (1, 2, 3; 2, 4, 5; 4, 5, 55):

(8)		c1	c2	c3	c4	c5	mean
	r1	1	2	4	NA	NA	2.3
	r2	NA	2	4	5	NA	3.7
	r3	NA	NA	4	5	55	21.3

Procesos

$$(r1) \quad (1 + 2 + 4) / 3 = 2.3$$

$$(r2) \quad (2 + 4 + 5) / 3 = 3.7$$

$$(r3) \quad (4 + 5 + 55) / 3 = 21.3$$

$$(*) \quad (2.3 + 3.7 + 21.3) / 3 = 9.1 \text{ \#M.mean}$$

En primer lugar, encontramos la media (=2,3) de los números (1, 2, 4) de los elementos c1, c2 y c3, más a la izquierda (r1). A continuación, encontramos la media (=3.7) de los números (2, 4, 5) de los tres elementos c2, c3 y c4, que siguen a la derecha (r2). Finalmente, calculamos el valor promedio (21.3) de los números (4, 5, 55) de c3, c4 y c5 (r3). En este momento, como esta es la última vez que los tres elementos consecutivos se presentan, se finaliza el cálculo y se calcula la media de los promedios determinados en r1, r2 y r3 (=9.1), y esto se usa como la media mayoritaria.

Esta media mayoritaria se obtiene extrayendo las tres muestras consecutivas, que son la mayoría, de todos los datos, por lo que no hay "sesgo" en el método de extracción de datos. Por lo tanto, la muestra que contiene la mediana en el centro de los datos (=4) es la que más se incluye en los procesos sucesivos, las muestras que contienen los valores circundantes disminuyen gradualmente y el número de muestras que contienen el mínimo y el máximo se limita a uno. De este modo, se garantiza la centralidad requerida para el valor medio.

Observamos a continuación la salida de la media (*mean*), la mediana (*median*) y la media mayoritaria (*M.mean*) de las filas de la matriz de datos D:

(9)	[Data]	c1	c2	c3	c4	c5	[Stat]	mean	median	M.mean	
	1	r1	1	2	3	4	10	r1	4	3	4
	2	r2	1	2	3	4	100	r2	22	3	14
	3	r3	1	2	3	4	200	r3	42	3	25
	4	r4	1	2	3	4	300	r4	62	3	36

El siguiente es el gráfico lineal que compara la media (*mean*), la mediana (*median*) y la media mayoritaria (*M.mean*) de los mismos datos.

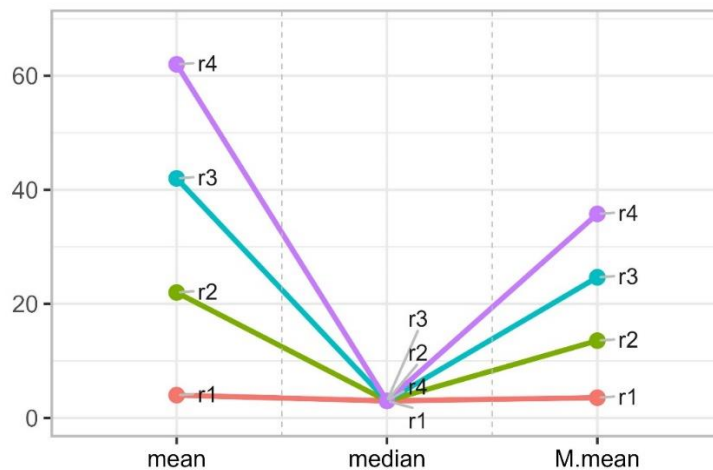


Fig.23. Comparación de media, mediana y media mayoritaria

Este gráfico muestra que la media está fuertemente influenciada por los valores atípicos mientras que la mediana no se ve afectada por estos mismos valores en absoluto. Por otro lado, la media mayoritaria se ve menos afectada por los valores atípicos y es, además, sensible a la magnitud de los valores atípicos. Cuando el sesgo de la distribución de datos es grande (r3, r4), la media mayoritaria se modifica significativamente; cuando el sesgo es pequeño, se aproxima a la media (r2) y cuando el sesgo es mínimo, coincide con la media. Por lo tanto, en general, todos los datos se pueden comparar utilizando la media mayoritaria.

Aplicación

En el siguiente gráfico de líneas de <pronombre + forma conjugada> (*lo+tiene*), la línea horizontal gris muestra el valor promedio general y la línea poligonal gris muestra la media mayoritaria de las frecuencias verticales. La superficie gris que la rodea indica un área donde el límite superior es la mitad de los valores que exceden la media mayoritaria y el límite inferior es la mitad de los valores que son menores que la media mayoritaria. De esta manera, al observar las líneas y superficies grises se puede determinar la tendencia general.

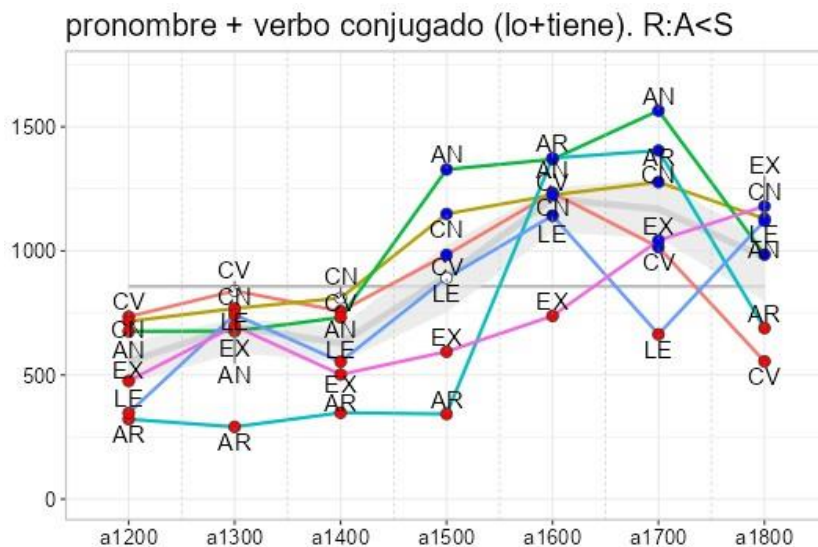


Fig. 24. Pronombre + verbo conjugado. Región:Año<Sociedad

6.5. Análisis de concentración

Denominamos "Análisis de concentración" al método de cambiar el orden de las filas y el de las columnas de la matriz de datos; por ejemplo, D1 en la siguiente tabla, para que se concentren las cifras mayores en la zona cercana a la línea diagonal trazada desde la parte superior izquierda hasta la inferior derecha, en D2.

(10)	[D1]	c1	c2	c3	c4	[D2]	c1	c3	c4	c2
	r1	1	8	5	6	r3	9	4	1	2
	r2	2	8	2	4	r1	1	5	6	8
	r3	9	2	4	1	r2	2	2	4	8

D1 es la matriz de datos de entrada y D2 es la matriz concentrada de salida. En D1 no se puede ver la tendencia general de los datos, pero en D2 los valores altos se concentran cerca de la línea diagonal, por lo que es posible leer secuencias de las filas r3-r1-r2 y las columnas c1-c3-c4-c2 con la misma tendencia (por ejemplo, fila: forma de palabra y columna: región, fila: región y columna: año, etc.)²³.

A continuación, utilizaremos los siguientes datos de muestra (D) para demostrar el procedimiento de centralización.

²³ Una vez que se especifica el año en una columna para analizar el orden cronológico, no se puede cambiar el orden de esa columna. En ese caso, cambiamos solo el orden de las líneas (forma de palabra, región, etc.).

(11)	[D]	c1	c2	c3	c4
	r1	1	8	5	6
	r2	2	8	2	4
	r3	9	2	4	1
A=D # Conservar el estado inicial D en A.					

Preparamos un vector W que indique la distancia desde el origen dado a cada fila de datos (D) (rango [-1, 1]).

(12)	-1.000	-0.333	0.333	1.000	...	W
------	--------	--------	-------	-------	-----	---

Ponderamos cada fila de datos (D) multiplicándola por el vector de distancia al origen (W), equivalente a la "distancia desde el origen (0)" de cada valor numérico.

(13)	*	c1	c2	c3	c4	...	D * W
	r1	-1.000	-2.667	1.667	6.000		
	r2	-2.000	-2.667	0.667	4.000		
	r3	-9.000	-0.667	1.333	1.000		

Elevamos toda la matriz de salida S a la tercera potencia.

(14)	*	c1	c2	c3	c4	...	S3
	r1	-1.000	-18.963	4.630	216.000		
	r2	-8.000	-18.963	0.296	64.000		
	r3	-729.000	-0.296	2.370	1.000		

Calculamos la suma de filas de esta matriz.

(15)	r1	r2	r3	...	RS(S3)
	200.667	37.333	-725.926		

Calculamos la suma de filas de la matriz de datos D.

16)	r1	r2	r3	...	RS(D)
	20	16	16		

Dividimos la suma de filas de la matriz de potencia (RS(S3)) por la suma de filas de la matriz de datos D (RS(D)) para obtener el valor promedio.

(17)	r1	r2	r3	...	RS(S3) / RS(D)
	10.033	2.333	-45.370		

Calculamos las raíces cúbicas de estos valores promedios y los devolvemos a la unidad original. Esta es la distancia de cada origen (distancia de Minkowsky).

(18)	r1	r2	r3
	2.157	1.326	-3.567

Encontramos el orden ascendente de la distancia de origen (O). En dicho orden r3 ocupa el primer lugar y r1, el tercer lugar.

(19)	r1	r2	r3	...	0
	3	2	1		

Cambiamos el orden de las filas de la matriz de datos según el orden ascendente de la distancia del origen.

(20)	*	c1	c2	c3	c4
	r3	9	2	4	1
	r2	2	8	2	4
	r1	1	8	5	6

Esta es la primera ordenación de la fila. Si se trata de un "Análisis de concentración de filas", termina aquí. En el caso del "Análisis de concentración de filas y columnas", se hace lo mismo en las columnas:

(21)		c1	c3	c4	c2	#	Reordenar las columnas.
	r3	9	4	1	2		
	r2	2	2	4	8		
	r1	1	5	6	8		

Después de reorganizar las filas y columnas de esta manera, si la matriz de datos [D] es diferente de la matriz de inicio [A] que hemos guardado antes de la reorganización, entonces las filas y columnas se reorganizan nuevamente. En este momento, si [D] se vuelve lo mismo que [A], el proceso finaliza.

(22)	[D]	c1	c2	c3	c4	[C]	c1	c3	c4	c2
	r1	1	8	5	6	r3	9	4	1	2
	r2	2	8	2	4	r1	1	5	6	8
	r3	9	2	4	1	r2	2	2	4	8

La matriz de datos (D) está centralizada. Se encuentran números grandes cerca de la diagonal.

Aplicación

El programa para análisis equilibrado doble (2.6) genera la siguiente matriz bidimensional de <verbo conjugado - pronombre> (*tiene-lo*) (R:A<S, ajustado por media):

(23)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
	AN	54	83	33	35	83	29	12
	AR	16	33	11	13	49	24	1

CN	145	123	48	54	37	16	18
CV	152	108	36	40	39	14	9
EX	126	126	39	37	5	34	3
LE	115	159	52	32	47	13	31

Las columnas de esta matriz (a1200, 1300, ..., a1800) son años, por lo que su orden debe ser fijo para mantener la secuencia cronológica, pero las filas están en orden alfabético, cambiante, por lo que al modificar esto, los números en la matriz comienzan desde arriba a la izquierda. Se puede concentrar cerca de la línea diagonal dirigida a la parte inferior derecha. Al hacerlo, se pueden observar en la matriz los movimientos regionales según la época.

(24)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
CV		152	108	36	40	39	14	9
CN		145	123	48	54	37	16	18
EX		126	126	39	37	5	34	3
LE		115	159	52	32	47	13	31
AN		54	83	33	35	83	29	12
AR		16	33	11	13	49	24	1

En cada celda del gráfico siguiente, hay una frecuencia normalizada y debajo de ella una desviación desde la mediana en el rango [-1, 1] ($\rightarrow 2.1$).

verbo conjugado - pronombre (tiene-lo). R:A<S

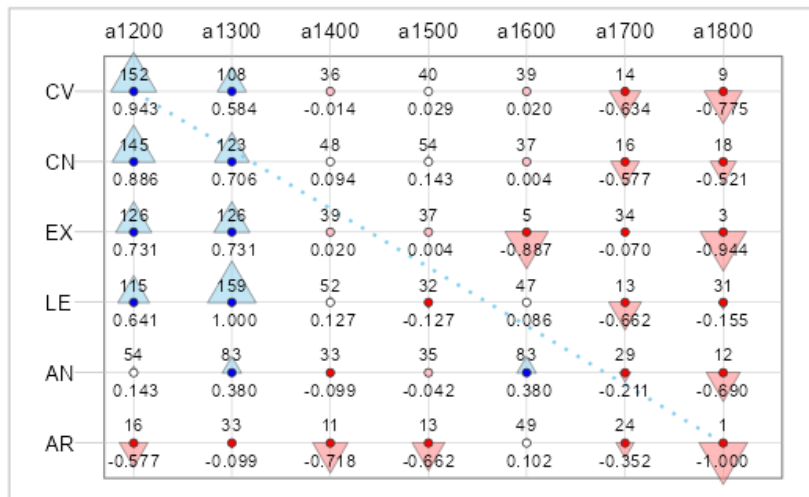


Fig.25. Distribución concentrada. Región : Año < Sociedad.

El siguiente es el resultado de la función de agrupamiento, que organiza los coeficientes de distancia en el eje vertical obtenidos mediante el análisis de concentración en una línea recta que los agrupa midiendo los intervalos entre coeficientes individuales:

$$[[CV>[CN>[EX>LE]]]>[AN>AR]]$$

De la misma manera, hemos creado una función que genera el siguiente gráfico.

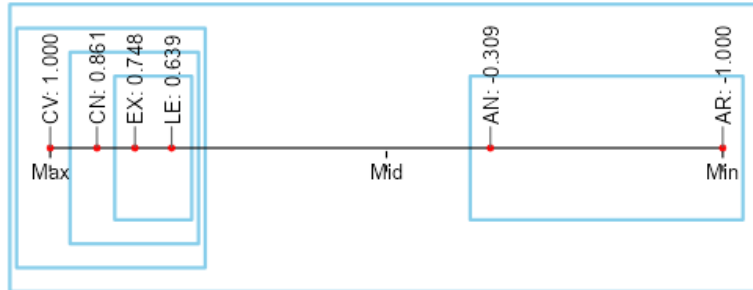


Fig.26. *Análisis de grupo por distancia. Región: Año < Sociedad.*

Esta figura muestra la estructura de grupos múltiples basada en la distancia (intervalo) histórica y lingüística de las regiones según el análisis de concentración. Hicimos posible visualizar claramente los grupos reduciendo las áreas donde los intervalos numéricos son grandes y ampliando las áreas donde los números son pequeños. Este tipo de agrupación generalmente se realiza mediante varios tipos de "Análisis de conglomerados" (*cluster analysis*), pero aquí intentamos un método que simplemente considera intervalos numéricos. Este método lo denominamos "Análisis de grupo por intervalo". Los gráficos en el análisis de conglomerados en general se visualizan en un dendrograma, pero en este análisis de grupo por distancia se visualizan en una estructura endocéntrica, tal y como se muestra en el diagrama anterior. Se puede comprender así, intuitiva y visualmente, cómo la estructura multigrupo se forma en un grupo superior uno tras otro.

El siguiente es el resultado de realizar un análisis de concentración de las soluciones de un análisis equilibrado doble con la sociedad y el año como variables objetivas y la región como variable de referencia (S:A<R).

verbo conjugado - pronombre (tiene-lo). S:A<R

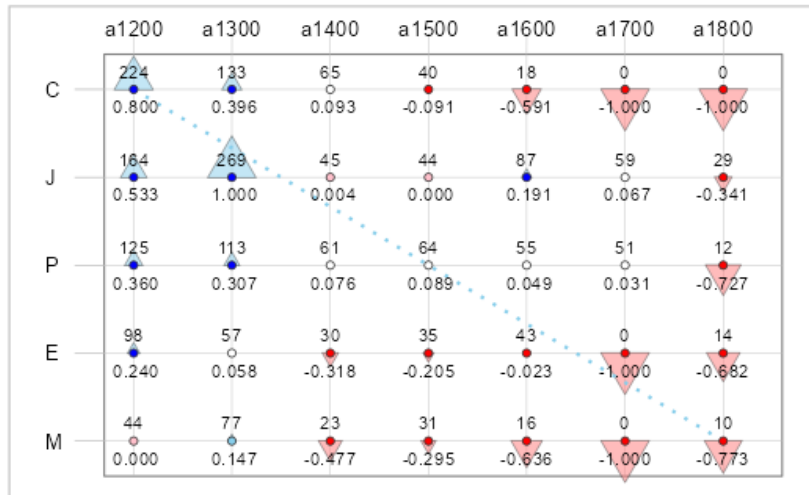


Fig.27. Distribución concentrada. Sociedad: Año < Región

La siguiente figura es un grupo de intervalos que se crea agrupando el coeficiente de distancia social (escala a la derecha de la figura anterior, Fig.-11), que es la variable objetivo en el diagrama de distribución de concentración anterior, según su intervalo.

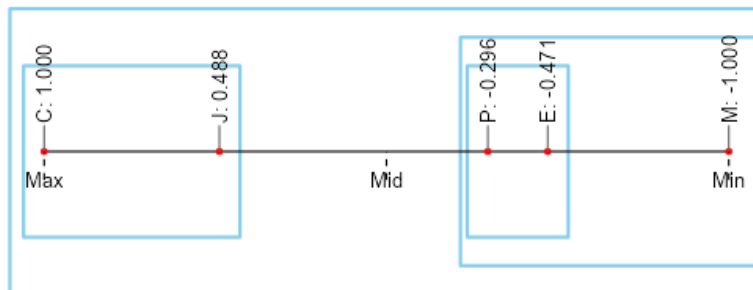


Fig.28. Análisis de grupo por distancia. Sociedad: Año < Región.

A continuación, se realizará un análisis de concentración en los datos obtenidos a partir del análisis equilibrado doble (\rightarrow 6.7.4), con las variables regionales y sociales como variables objetivas y el año como variable de referencia. Esta vez, como no existe un orden específico para las variables regionales y sociales, las concentramos tanto en filas como en columnas (R:S<A).

verbo conjugado - pronombre (tiene-lo). R:S<A

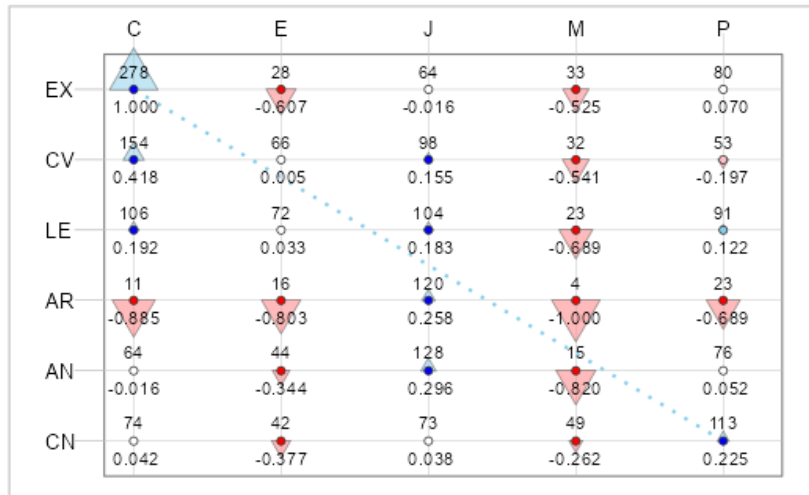


Fig.29. Distribución concentrada. Región: Sociedad < Año.

La siguiente figura se ha elaborado desde el mismo diagrama de grupos de intervalos: Región: Sociedad < Año.

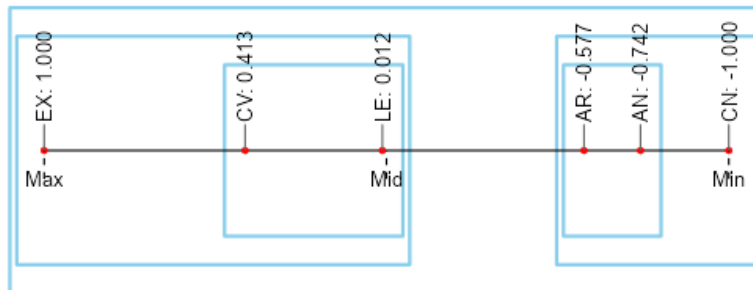


Fig.30. Análisis de grupo por distancia. Región.

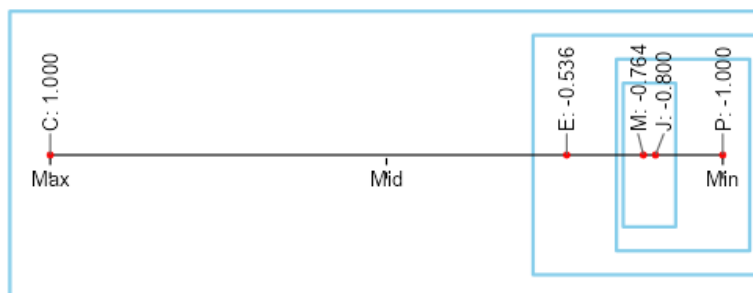


Fig.31. Análisis de grupo por distancia. Sociedad

La variable regional tiene una estructura continua:

$$[[EX>[CV>LE]]>[[AR>AN]>CN]]$$

Estructura continua de la variable social (documento):

```
[C>[E>[[M>J]>P]]]
```

6.6. Grado de significatividad

Para abordar el concepto del grado de significatividad, comencemos con un ejemplo sencillo. Si 8 personas en un grupo de 10 respondieron que "usan una determinada forma de palabra", la pregunta es si $8 / 10 = 0,8$ es "estadísticamente significativo". En tales casos, generalmente se utiliza una "Prueba binomial" para realizar dicha prueba estadística. El siguiente es el resultado de la función R `binom.test` (prueba superior: 'greater').

```
(25) binom.test(8, 10, p=.5, 'greater')
Exact binomial test
data: 8 and 10
number of successes = 8, number of trials = 10, p-value = 0.05469
alternative hypothesis: true probability of success is greater than
0.5
95 percent confidence interval:
 0.4930987 1.0000000
sample estimates:
probability of success 0.8
```

El valor-p (*p-value*) de este resultado es 0.05469, que indica la suma de las probabilidades de que "8 o más personas utilicen la forma de la palabra". Suponiendo que la probabilidad de utilizar una forma de palabra o no es 0,5 (que denominamos "Probabilidad esperada"), la probabilidad de que haya 8 o más personas es superior al 5%. Por ejemplo, es lo mismo que cuando lanzamos 10 monedas y salen cara en 8 o más monedas. Se espera que ocurra con una probabilidad de 0,05469 (5,5%), incluso por casualidad. Si primero establecemos el 5% como nivel de significatividad, se considerará "no significativo", porque excede el 5%. Si hay 9 (o más) personas que usan la forma de la palabra, entonces $p\text{-valor} = 0.01074$ en `binom.test(9, 10, p=0.5, 'greater')`, y la probabilidad de que tal cosa suceda por casualidad es del 1%, por lo que 9 de cada 10 personas pueden considerarse un "número significativamente grande". Esto se debe a que la probabilidad de que el juicio de "significatividad" sea incorrecto es extremadamente pequeña (1%). El siguiente resultado muestra el valor-p para cada resultado de la prueba binomial.

```
(26) binom.test(8, 10, p=.5, 'greater')$p.value # 0.0546875
      binom.test(9, 10, p=.5, 'greater')$p.value # 0.01074219
```

Estas pruebas estadísticas se basan en la probabilidad de distribución binomial. Por ejemplo, cuando lanzamos una moneda, la probabilidad de que salga cara es de 0,5 (50%). Para los siguientes números #1 y #2, la probabilidad de que una moneda (tamaño de muestra = número de monedas, s (size) = 1) sea "cruz" ($x = 0$) o "cara" ($x = 1$) es 0,5. Incluso sin llevar a cabo este experimento, podemos entender inmediatamente que dado que una moneda tiene dos opciones: cruz y cara, la probabilidad de cara es 0,5:

```
(27) #dbinom(x, size, prob) # binomial probability density
      dbinom(x=0, s=1, p=.5) #1. 0.5
      dbinom(x=1, s=1, p=.5) #2. 0.5
```

Sin embargo, debemos realizar cierta consideración para conocer la probabilidad de obtener 0, 1 o 2 caras cuando se lanzan dos monedas. En este momento, si enumeramos todos los casos, serán "cruz / cruz", "cruz / cara", "cara / cruz", "cara / cara", por lo que, cuando se lanzan dos monedas, las probabilidades de que haya 0, 1 y 2 caras son 0,25 ("cruz / cruz"), 0,5 ("cruz / cara", "cara / cruz") y 0,25 ("cara / cara), respectivamente (#3, #4, #5). Sin embargo, es difícil determinar la probabilidad de que 9 monedas presenten la cara cuando se lanzan 10 monedas, por lo que se necesita la función `dbinom` de R. Según el punto #6 que abordamos en estos momentos, la probabilidad es 0,009765625:

```
(28) dbinom(x=0, s=2, p=.5) #3. 0.25
      dbinom(x=1, s=2, p=.5) #4. 0.50
      dbinom(x=2, s=2, p=.5) #5. 0.25
      dbinom(x=9, s=10, p=.5) #6. 0.009765625
```

A continuación, para encontrar la probabilidad acumulada (cp: *cumulative probability*) en lugar de la probabilidad individual (*probability density*), podemos utilizar `pbinom(q,s,p)`. Por ejemplo, cuando lanzamos dos monedas, la probabilidad de que el número de cara sea 0, 1 o menos (la suma de las probabilidades de 0 y 1) y 2 o menos (la suma de las probabilidades de 0, 1 y 2) podemos averiguarlo mediante las siguientes ejecuciones (#7, 8, 9), por las que obtenemos 0,25, 0,75 y 1, respectivamente.

```
(29) pbinom(0, 2, .5) #7. 0.25
      pbinom(1, 2, .5) #8. 0.75
      pbinom(2, 2, .5) #9. 1.00
```

El siguiente resultado es la probabilidad individual (*density*: d) obtenida usando `dbinom` y la probabilidad acumulada obtenida con `pbinom`, con frecuencia

(f) = 0:2 (es decir: 0, 1, 2), suma (s) = 2 y probabilidad esperada (e) = 0.5. La columna cp (*cumulative probability*) es de la probabilidad acumulada y la columna up (*upper probability*) es la suma de la probabilidad complementaria superior (1 - cp).

(30)	f	t	e	d	cp	up
	0	2	0.5	0.25	0.25	0.75
	1	2	0.5	0.50	0.75	0.25
	2	2	0.5	0.25	1.00	0.00

A continuación, con frecuencias (f) = 0:5 (0, 1, 2, ..., 5), suma (s) = 5 y probabilidad esperada (e) = 0,5, están calculadas la densidad de probabilidad (d) y la probabilidad acumulada (cp) y la suma de la probabilidad superior (up=1-cp). Aquí nuevamente, la densidad de probabilidad (d) es la mayor (d=0,31) donde f está en el centro (f=2, 3) y la suma de d es 1, lo que confirma que el punto de llegada de cp es 1:

(31)	f	t	e	d	cp	up
	0	5	0.5	0.03125	0.03125	0.96875
	1	5	0.5	0.15625	0.18750	0.81250
	2	5	0.5	0.31250	0.50000	0.50000
	3	5	0.5	0.31250	0.81250	0.18750
	4	5	0.5	0.15625	0.96875	0.03125
	5	5	0.5	0.03125	1.00000	0.00000

Si el nivel de significatividad se establece en 5% (0,05), un valor f=0 con una probabilidad acumulada (cp) menor que 0,05 se considera "significativamente pequeño". Esta probabilidad acumulada se convierte en el valor-p de la "prueba binomial" de la siguiente manera.

```
round(binom.test(0,5,p=0.5,'less')$p.value, 2) # 0.03
```

Se debe extremar el cuidado a la hora de encontrar los valores 'significativamente grandes'. Según la tabla anterior, cuando f=4, cp=0,97, su complemento (superior) es 0,03. Sin embargo, f=4 no puede considerarse como un valor 'significativamente grande'. Esto se debe a que el complemento de la probabilidad acumulada cuando f=4 (0,03) no es la suma de las densidades de probabilidad de f=4 y 5 (0,16+0,03=0,19), sino la probabilidad individual de solo f=5 (0,03). Esto es porque la suma de las densidades de probabilidad de f=4, 5 (0,19) es hasta f=3. La razón es que la probabilidad acumulada (cp) en f=4 es la suma de las densidades de probabilidad de f=(0,1,...,4), donde está incluida la probabilidad individual de f=4. En este momento, si no se incluye la probabilidad

individual de $f=4$ ($f=0, 1, 2, 3$), el complemento será la suma correcta de las densidades de probabilidad de $f=4, 5$. Por lo tanto, la probabilidad acumulada al encontrar un número "significativamente grande" debe ser la probabilidad acumulada calculada con $f-1$ en $pbinom(f-1,s,p)$. De hecho, el valor-p de la "prueba binomial" es el siguiente.

```
round(binom.test(4,5,p=0.5,'mayor')$p.value,2) # 0.19
```

En este estudio, en lugar de determinar la importancia de las frecuencias individuales utilizando el valor-p de una prueba binomial dicotómica ("significativa" o "no significativa"), evaluamos el grado de significatividad de todas las frecuencias continuamente. Por lo tanto, en lugar de usar el complemento de la probabilidad acumulada, correspondiente al valor-p, usamos la probabilidad acumulada misma ($pbinom(f,t,e)$) para el lado inferior del todo ($p=pbinom(f,t,e) < 0.5$), y $pbinom(f - 1,t,e)$ para el lado superior ($p \geq 0.5$). Esta probabilidad la llamamos "Probabilidad acumulada ajustada" (ap). La siguiente aplicación es su resultado:

(32)	f	t	e	d	cp	ap
	0	5	0.5	0.03125	0.03125	0.03125
	1	5	0.5	0.15625	0.18750	0.18750
	2	5	0.5	0.31250	0.50000	0.50000
	3	5	0.5	0.31250	0.81250	0.50000
	4	5	0.5	0.15625	0.96875	0.81250
	5	5	0.5	0.03125	1.00000	0.96875

Según el resultado anterior, la probabilidad acumulada simple (cp) coincide con la probabilidad acumulada ajustada (ap) hasta $f=0:2$, y en $f=3:5$ ($3,4,5$) coincide con la probabilidad acumulada (cp) de $f=2:4$ ($2,3,4$). Los valores de ap en $f=2:3$ en el centro y desciende y asciende con la misma pendiente antes y después del centro. Como resultado, la suma de ap en $f=0$ (0.03) y ap en $f=5$ (0.97) es igual a 1, la suma de ap en $f=1$ (0.19) y ap en $f=4$ (0.81) es igual a 1, y la suma de ap en $f=2$ (0.50) y ap en $f=3$ (0.50) es igual a 1, lo cual es razonable y convincente. Cuando p es menor que 0,05, "f es significativamente pequeña", y cuando ap es menor que 0,01, "f es significativamente muy pequeña". Cuando ap es mayor que 0,95, "f es significativamente grande", y cuando ap es mayor que 0,99, "f es significativamente muy grande".

Aplicación

En este estudio, se utiliza la probabilidad acumulada ajustada (de

distribución binomial) para calcular la significatividad de cada número en un conjunto de números. El siguiente es el resultado (O: objetivo) del número de apariciones de Región: Año del <verbo conjugado – pronombre> (*tiene-lo*).

(33)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
	CV	152	108	36	40	39	14	9
	CN	145	123	48	54	37	16	18
	EX	126	126	39	37	5	34	3
	LE	115	159	52	32	47	13	31
	AN	54	83	33	35	83	29	12
	AR	16	33	11	13	49	24	1

A continuación, se muestra una matriz (P) que indica la significatividad (probabilidad acumulada ajustada) de cada valor numérico en la matriz Cc:

(34)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
	CV	1.000	1.000	0.017	0.068	0.050	0.000	0.000
	CN	1.000	1.000	0.379	0.656	0.025	0.000	0.000
	EX	1.000	1.000	0.050	0.025	0.000	0.007	0.000
	LE	1.000	1.000	0.548	0.003	0.326	0.000	0.002
	AN	0.656	1.000	0.005	0.012	1.000	0.001	0.000
	AR	0.000	0.005	0.000	0.000	0.435	0.000	0.000

Por ejemplo, el grado de significatividad de 0,545 para CV:a1400 se calcula de la siguiente manera.

```
(35) BinP=function(f, t, e=0.5) {
      p=pbinom(f, t, e); ifelse(p<=.5, p, pbinom(f-1, t, e))
    }
E=L$cc
s=sum(E); s # 2135
nr=nrow(E); nr # 6
nc=ncol(E); nc # 7
e=1/(nr*nc); e # 0.02380952
BinP(E[1, 3], s, e) # 0.5465587
```

Aquí, preparamos una función BinP que calcula el grado de significatividad; luego calculamos la suma de la matriz E=Cc (s=2136) así como el número de filas de E (nr=6) y el número de columnas de E (nc=7). La probabilidad esperada (e) es e=1/(nr*nc)=0,02380952. Esta probabilidad esperada (e) indica la probabilidad distribuida a cada celda. La función BinP recibe la frecuencia (f), la suma (s) y la probabilidad esperada (e) y devuelve el grado de significatividad (0,5452151).

La gradación cromática de puntos en la siguiente figura (de azul → azul

claro → blanco → rojo claro → rojo) refleja el grado de significatividad (s) de la siguiente manera:

1. azul: $s > 0.99$ (significativamente muy grande)
2. azul claro: $0.95 < s \leq 0.99$ (significativamente grande)
3. blanco: $0.05 \leq s \leq 0.95$ (no significativo)
4. rojo claro: $0.01 \leq s < 0.05$ (significativamente pequeño)
5. rojo: $s < 0.01$ (significativamente muy pequeño)

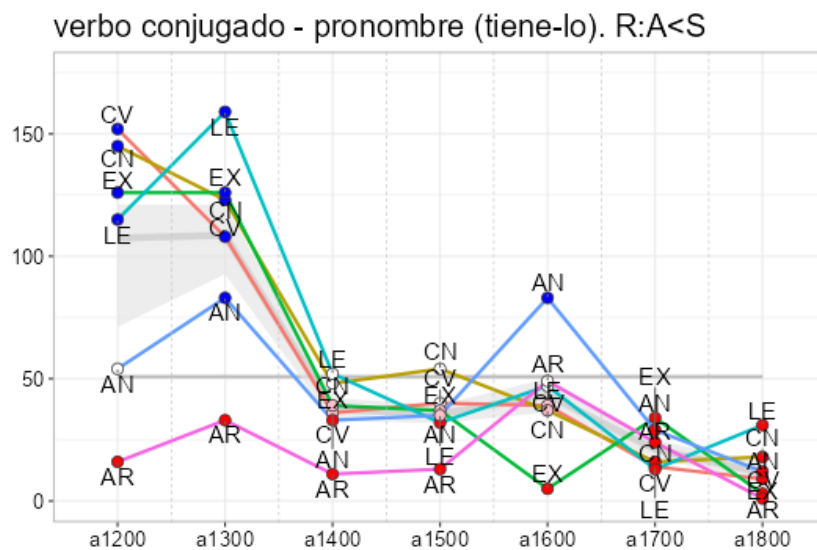


Fig.32. Grado de significatividad. Región:Año<Sociedad. Cambio cronológico.

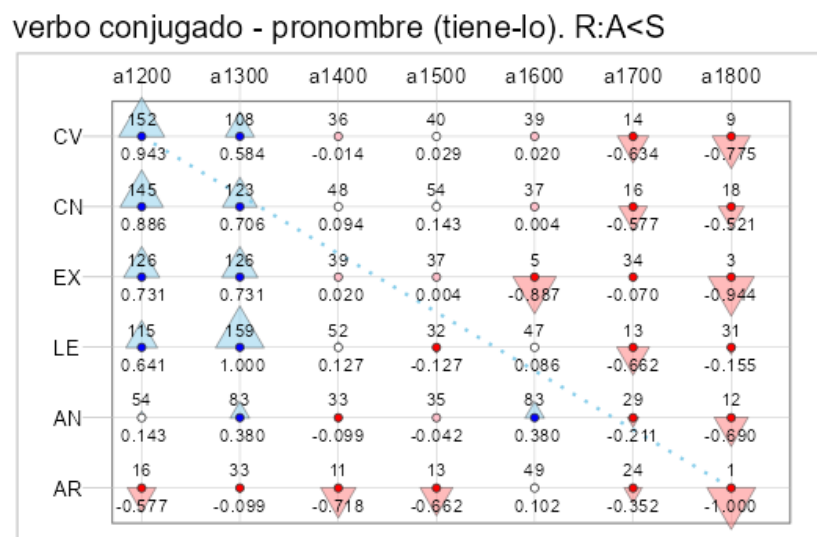


Fig.33. Grado de significatividad. Región:Año<Sociedad. Concentración.

6.7. Análisis equilibrado

El método de calcular el valor promedio general utilizando la "Media de ratios" para eliminar la influencia de algunos grupos grandes, lo denominamos "Análisis equilibrado". Primero, explicaremos la "Frecuencia normalizada" y "Media de ratios", a continuación explicaremos el "Análisis de equilibrio simple" y, finalmente, nos ocuparemos del "Análisis equilibrado doble", que analiza dos variables, ambas con la frecuencia normalizada y la media de ratios.

Frecuencia normalizada

A continuación, mostramos la frecuencia absoluta observada (O), el número total de palabras (S) y la frecuencia normalizada (N) de enclisis (*tiene-lo*) para cada grupo de edad:

36)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
	O	501	368	267	467	268	222	42
	S	179,393	207,189	282,943	390,394	369,095	230,332	128,871
	N	279	178	94	120	73	96	33

Por ejemplo, $N:a1200 = 279$ (frecuencia normalizada) se calcula de la siguiente manera:

$$N = O / S * 10^5 = 501 / 179393 * 10^5 = 279,2751$$

Las frecuencias absolutas observadas: (O), tal como están, no se pueden comparar, porque sus bases (número total de palabras: S) son diferentes, pero las frecuencias normalizadas (N) se dividen por el número total respectivo de palabras (S) y se multiplican. por 10^5 . Por lo tanto, la comparación es posible. Esto es similar al hecho de que, aunque no se pueden comparar frecuencias absolutas con diferentes bases, se pueden comparar los porcentajes (divididos por cada base y multiplicados por 100).

A continuación, hacemos lo mismo con las tablas cruzadas. La siguiente es una tabla cruzada de las mediciones reales de enclisis (*tiene-lo*) por región y año (O):

(37)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800	... 0
	AN	61	12	15	29	23	67	15	
	AR	4	20	28	6	27	21	1	
	CN	58	80	51	216	139	74	16	
	CV	314	121	111	56	18	11	5	
	EX	4	36	20	146	55	11	1	

LE	60	99	42	14	6	38	4
----	----	----	----	----	---	----	---

La que sigue es una tabla cruzada del número total de palabras por región y año (S):

(38)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800	... S
	AN	22,429	12,920	19,786	37,715	16,953	76,955	30,078	
	AR	13,992	39,778	60,740	18,189	16,451	23,680	10,346	
	CN	17,544	37,115	50,829	153,661	152,357	75,776	38,308	
	CV	99,219	63,278	85,497	68,742	18,029	20,422	24,676	
	EX	1,439	23,723	15,805	91,872	156,564	10,367	11,130	
	LE	24,770	30,375	50,286	20,215	8,741	23,132	14,333	

Más adelante, observamos una tabla cruzada de las frecuencias normalizadas (N) calculadas a partir de las dos tablas anteriores:

(39)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800	... N
	AN	272	93	76	77	136	87	50	
	AR	29	50	46	33	164	89	10	
	CN	331	216	100	141	91	98	42	
	CV	316	191	130	81	100	54	20	
	EX	278	152	127	159	35	106	9	
	LE	242	326	84	69	69	164	28	

Por ejemplo, la frecuencia normalizada (=272) de AN:a1200 se puede obtener de la siguiente manera (N: frecuencia normalizada, O: frecuencia absoluta observada, S: número total de palabras, m: multiplicador):

$$N = O / S * m = 61 / 22429 * 10^5 = 271,9693$$

Media de ratios

Suponemos, por ejemplo, que en un grupo de 100 personas (X), 30 respondieron que usan una determinada forma de palabra, y en otro grupo de 200 personas (Y), 40 respondieron que usan la misma palabra. Veamos dos maneras de calcular la media de frecuencia de uso de la misma forma de palabra:

(40)	
x=30; xb=100; rx=x/xb; rx #a1 0.30 (X)	
y=40; yb=200; ry=y/yb; ry #a2 0.20 (Y)	
(x+y)/(xb+yb) #a3 0.233: biased to Y	
(rx+ry)/2 #a4 0.250: better (equilibrated, representative)	

Aquí, cuando es necesario encontrar la tasa de uso general, al igual que en #a3, se divide el número total de usuarios (x+y=70) por la población base (xb+yb=300):

$$(x+y) / (xb+yb) = 0,233 \text{ (\#a3)}$$

A esto, lo llamaremos "Ratio total".

Por otro lado, se dice que no se debe utilizar la siguiente "Media de ratios":

$$(rx+ry) / 2 = 0,250 \text{ (\#a4)}$$

Esto se debe a que, si se calcula para todo el conjunto combinado, será 0,233 como #a3 y no 0,250 como #a4.

Sin embargo, la ratio total (#a3) de 0,233 se encuentra cerca de la ratio del grupo Y (0,20) y lejos de la ratio del grupo X (0,30).

Es cierto que el método de la ratio total (#a3) es correcto en términos de la proporción total, pero en el análisis de datos reales, se puede decir que la media de ratios (#a4) es una proporción más representativa de la distribución general. Por ejemplo, si X (número total: 100) es un "grupo de hombres" y Y (número total: 200) es un "grupo de mujeres", entonces #a4 (ratio de medias) es una forma razonable de tratar a hombres y mujeres de manera igualitaria. Al calcular la tasa de compra de un producto específico en un mercado total constituido de distintos lugares, se debe utilizar el método #a3 (ratio total). Pero al intentar calcular la tasa de uso de una forma de palabra que no esté sesgada por género, en este caso conviene utilizar el método de #a4 (media de ratios). Esto se debe a que el resultado del método #a3 (proporción total) está fuertemente influenciado por la tasa de uso de las mujeres, por lo que no se puede decir que las tasas de uso de hombres y mujeres reciban el mismo trato. En este caso, la proporción total indica un valor que está sesgado hacia <la tasa de uso de formas de palabras en Y (femenino)" en lugar de "la tasa equitativa de uso de formas de palabras>.

Se debe tener cuidado al analizar ratios cuando las poblaciones son diferentes. Por otro lado, cuando las poblaciones de cada grupo son iguales, como se muestra a continuación, la ratio total y la media de ratios son iguales, por lo que no ocurre ningún problema²⁴:

(41) $x=30; xb=100; rx=x/xb; rx \#c1 0.30 (X)$ $y=20; yb=100; ry=y/yb; ry \#c2 0.20 (Y)$

²⁴ Por ejemplo, los diccionarios de frecuencia de vocabulario de García Hoz (1953) y Juilland y Chang-Rodríguez (1964) fijan el número de palabras en cada campo en 100.000, por lo que la tasa de uso del vocabulario en cada campo se puede calcular directamente y sin problemas.

$$\begin{aligned} (x+y)/(xb+yb) & \#c3 \ 0.25 \\ (rx+ry)/2 & \#c4 \ 0.25 \end{aligned}$$

Análisis equilibrado simple

A continuación, mostramos la frecuencia absoluta observada de la enclisis (*tiene-lo*) (O), la frecuencia total de palabras (S) y la frecuencia normalizada (N) para cada edad (\rightarrow 6.7.1):

(42)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
	O	501	368	267	467	268	222	42
	S	179,393	207,189	282,943	390,394	369,095	230,332	128,871
	N	279	178	94	120	73	96	33

La siguiente es una tabla cruzada de la frecuencia normalizada, seguida de la media mayoritaria (*M.mean* \rightarrow 6.4) :

(43)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800	... N
	AN	242	66	58	67	109	84	40	
	AR	15	43	42	21	134	72	2	
	CN	290	200	93	141	91	95	35	
	CV	316	185	128	78	78	40	13	
	EX	111	130	99	158	35	72	2	
	LE	216	302	77	53	39	142	15	
	M. mean	209	152	83	84	81	82	17	

En la fila final de la tabla anterior, está inserta la media mayoritaria (*M.mean*) como referencia; por ejemplo, la media mayoritaria en el a.1200 tiene en cuenta las diferencias de frecuencia entre regiones y, por ello, representa adecuadamente la situación en el a1200. Por otro lado, si calculamos la frecuencia normalizada basándonos únicamente en el mismo año sin considerar estas diferencias regionales (41), estaremos ante el valor que indica principalmente las características de la región con alta frecuencia más que las características generales del año. De hecho, al comparar la frecuencia normalizada para cada grupo de año ignorando las diferencias regionales (41) y la frecuencia normalizada que toma en cuenta las diferencias regionales por igual (media mayoritaria en la tabla 42), los resultados son bastante diferentes. De esta manera, en lugar de la frecuencia normalizada basada en la variable independiente, utilizamos la media mayoritaria de la distribución de frecuencia normalizada obtenida al cruzar la variable en cuestión, que llamamos "variable objeto", con otra variable, que llamamos "variable de referencia". Este método lo denominamos "Análisis equilibrado".

A continuación, explicaremos el método de Análisis equilibrado utilizando múltiples variables de referencia. La siguiente es una tabla cruzada con la columna como el año (por ejemplo: a1200) y las filas como la región + sociedad (por ejemplo, AN:C)²⁵.

(44)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
AN:C		298	54	54	74	*	*	*
AN:E		80	78	51	79	114	*	0
AN:J		*	978	*	75	240	76	0
AN:M		0	0	*	66	39	0	52
AN:P		*	136	120	83	309	96	52
AR:C		*	0	*	*	34	*	*
AR:E		25	31	47	47	0	*	0
AR:J		*	416	0	0	412	158	*
AR:M		64	0	0	*	*	0	0
AR:P		22	64	47	30	75	25	11
CN:C		481	279	137	87	32	0	0
CN:E		221	56	32	18	144	0	*
CN:J		0	269	95	192	102	132	56
CN:M		*	234	132	91	0	0	53
CN:P		467	473	99	189	99	85	34
CV:C		470	283	184	70	*	*	*
CV:E		226	167	81	130	124	0	0
CV:J		207	713	*	50	104	79	76
CV:M		578	0	0	118	0	0	0
CV:P		244	141	56	72	82	53	17
EX:C		466	418	*	*	*	*	*
EX:E		*	99	76	79	16	0	0
EX:J		*	169	85	69	0	183	*
EX:M		0	784	0	0	0	0	23
EX:P		304	88	178	181	37	124	0
LE:C		219	331	93	56	*	*	*
LE:E		282	234	44	35	53	0	224
LE:J		769	*	123	117	*	0	62
LE:M		0	*	121	0	189	0	0
LE:P		102	455	162	136	80	171	22
M. mean		221	206	79	76	78	38	21

Los nombres de las filas en este resultado son una combinación de variables regionales y variables sociales. Por ejemplo, AN:C indica AN=Andalucía + C=documento canclleresco. La media mayoritaria (*M.mean*) en la última fila es

²⁵ Los asteriscos (*) indican valores faltantes. Por ejemplo, AN:C:a1600 no tiene datos. Los valores faltantes se ven en todas las columnas, pero al calcular la media mayoritaria (*M.media*), el cálculo se realiza utilizando un vector que excluye los valores faltantes.

el resultado de un análisis equilibrado con el año como variable objetiva, y la región y la sociedad como variables de referencia.

A continuación, presentamos una tabla con las siguientes cuatro filas: (1) frecuencia absoluta observada (Obs), (2) frecuencia normalizada con el año como variable independiente (No eq.), (3) frecuencia normalizada con el año como variable objetiva y región como variable de referencia (Eq. A<R), (4) frecuencia normalizada con el año como variable objetiva, y la región y la sociedad como variables de referencia (Eq. A<R:S). En la última columna se encuentra la media horizontal:

(45)	* a1200	a1300	a1400	a1500	a1600	a1700	a1800	mean
Obs	501	368	267	467	268	222	42	305
No eq. A	279	178	94	120	73	96	33	125
Eq. A<R	261	169	95	91	99	97	26	120
Eq. A<R:S	221	206	79	76	78	38	21	103

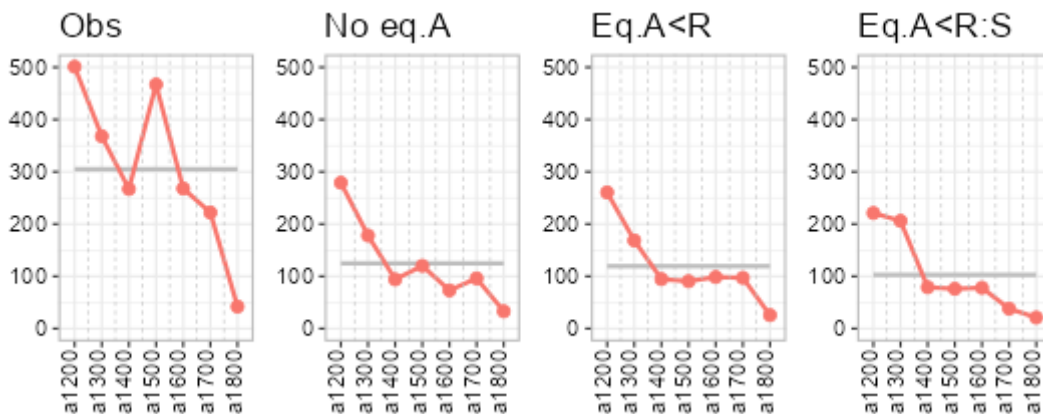


Fig.34a,b,c,d. Frecuencia absoluta observada / No equilibrado / Equilibrado A<R / Equilibrado A<R:S

De esta manera, la frecuencia normalizada de análisis no equilibrado (No eq. A) presenta un valor numérico relativamente alto. Sin embargo, cuando la variable de referencia es R (Eq. A<R), la frecuencia normalizada, al usar la correspondiente a cada región, muestra unos índices relativamente bajos. Finalmente, si la variable de referencia es R:S (Eq. A<R:S), el valor será aún menor, puesto que se utiliza la frecuencia normalizada de la combinación de región y sociedad.

Por lo tanto, en lugar de presentar los resultados del análisis de equilibrio utilizando las frecuencias normalizadas tal como son, si las ajustamos por media

de frecuencias absolutas observadas para que su media sea igual a la de las mediciones reales (frecuencia absoluta observada), obtenemos lo siguiente (véase la columna de *mean*, donde las cifras son todas iguales):

(46)	* a1200	a1300	a1400	a1500	a1600	a1700	a1800	mean
Obs	501	368	267	467	268	222	42	305
No eq.	682	435	230	293	179	235	81	305
Eq. A<R	666	431	242	231	252	248	65	305
Eq. A<R:S	656	613	235	225	232	112	62	305

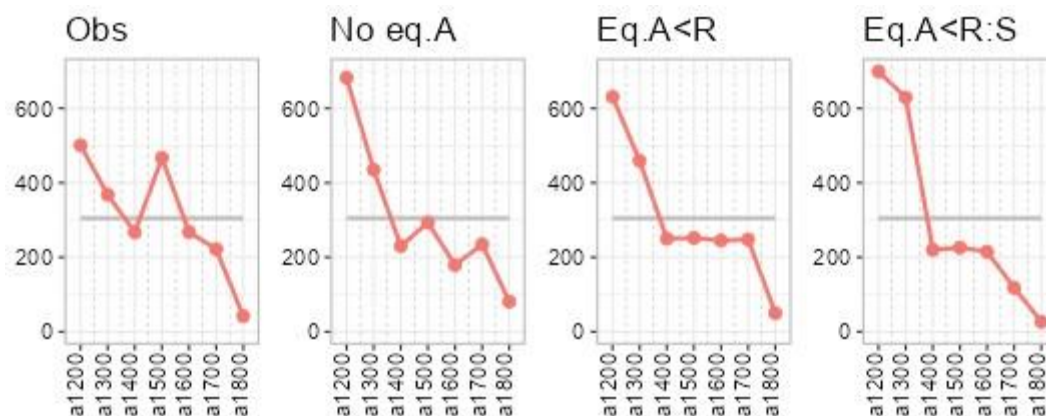


Fig.35a,b,c,d. *Frecuencia absoluta observada / No equilibrado / Equilibrado A<R / Equilibrado A<R:S, ajustada por media*

Las tres tablas y gráficos anteriores, al lado de los de frecuencia absoluta observada, muestran una tendencia a la baja en general, pero los detalles son bastante diferentes. En particular, la forma y escala del segundo gráfico sin análisis equilibrado (A) son significativamente diferentes de las del segundo y tercer gráfico (A<R, A<R:S), ambos con análisis equilibrado. Finalmente, la media mayoritaria de la frecuencia normalizada con múltiples variables de referencia se mantiene baja (A<R:S), pero esto se debe a que la población se vuelve más pequeña en el caso de múltiples variables de referencia (región y sociedad). Por ello, conviene aplicar el ajuste por media para obtener la distribución con la media igualada (Tabla-(45), Fig.15d).

La curva de A<R muestra una gran disminución entre a1200 y a1300, pero en el caso de múltiples variables de referencia (A<R:S), el cambio entre los dos períodos es pequeño y muestra que la tendencia de mayor equilibrio se encuentra en el caso de la tercera variable de referencia múltiple (A<R:S), por lo que el último análisis (A<R:S) proporciona una conclusión válida.

Análisis equilibrado doble

A continuación, realizamos un "Análisis de equilibrio doble" (R:A<S) de la estructura enclítica <verbo conjugado - pronombre> (*tiene-lo*). Para ello, en primer lugar y a modo de comparación, creamos una tabla cruzada simple (frecuencia absoluta) de región (R) y año (A), utilizando el método convencional sin equilibrar (R:A).

(47)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800	...	0
AN		61	12	15	29	23	67	15		
AR		4	20	28	6	27	21	1		
CN		58	80	51	216	139	74	16		
CV		314	121	111	56	18	11	5		
EX		4	36	20	146	55	11	1		
LE		60	99	42	14	6	38	4		

La siguiente es la tabla cruzada simple de frecuencia normalizada de región (R) y año (A):

(48)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800	...	N
AN		272	93	76	77	136	87	50		
AR		29	50	46	33	164	89	10		
CN		331	216	100	141	91	98	42		
CV		316	191	130	81	100	54	20		
EX		278	152	127	159	35	106	9		
LE		242	326	84	69	69	164	28		

Por otro lado, en el análisis equilibrado doble que se aborda aquí, hay que preparar la siguiente tabla cruzada de frecuencias normalizadas con la región y año en distintas columnas (C, E, J, M, P), seguida de la media mayoritaria (*M.mean*) de manera siguiente (R:A<S):

(49)	*	C	E	J	M	P	M. mean
AN:a1200		298	80	*	0	*	115
AN:a1300		54	78	978	0	136	177
AN:a1400		54	51	*	*	120	70
AN:a1500		74	79	75	66	83	76
AN:a1600		*	114	240	39	309	176
AN:a1700		*	*	76	0	96	62
AN:a1800		*	0	0	52	52	26
AR:a1200		*	25	*	64	22	34
AR:a1300		0	31	416	0	64	71
AR:a1400		*	47	0	0	47	24
AR:a1500		*	47	0	*	30	27
AR:a1600		34	0	412	*	75	105

AR:a1700	*	*	158	0	25	52
AR:a1800	*	0	*	0	11	3
CN:a1200	481	221	0	*	467	310
CN:a1300	279	56	269	234	473	262
CN:a1400	137	32	95	132	99	102
CN:a1500	87	18	192	91	189	115
CN:a1600	32	144	102	0	99	79
CN:a1700	0	0	132	0	85	34
CN:a1800	0	*	56	53	34	38
CV:a1200	470	226	207	578	244	323
CV:a1300	283	167	713	0	141	229
CV:a1400	184	81	*	0	56	76
CV:a1500	70	130	50	118	72	86
CV:a1600	*	124	104	0	82	83
CV:a1700	*	0	79	0	53	31
CV:a1800	*	0	76	0	17	18
EX:a1200	466	*	*	0	304	269
EX:a1300	418	99	169	784	88	268
EX:a1400	*	76	85	0	178	83
EX:a1500	*	79	69	0	181	80
EX:a1600	*	16	0	0	37	12
EX:a1700	*	0	183	0	124	72
EX:a1800	*	0	*	23	0	6
LE:a1200	219	282	769	0	102	244
LE:a1300	331	234	*	*	455	338
LE:a1400	93	44	123	121	162	111
LE:a1500	56	35	117	0	136	68
LE:a1600	*	53	*	189	80	101
LE:a1700	*	0	0	0	171	29
LE:a1800	*	224	62	0	22	65

La media mayoritaria (*M.mean*) en cada fila de esta tabla es el valor medio equilibrado por la variable de referencia social (S). En este momento, cada fila tiene la región y el año como variable objetiva compuesta. Por ejemplo, la primera línea, AN:a1200, consta de la región (AN: Andalucía) y el año (a1200). A continuación, creamos la siguiente tabla cruzada que desglosa esta variable objetiva doble en región y sociedad en fila y columna, respectivamente.

(50)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
AN	115	177	70	76	176	62	26	
AR	34	71	24	27	105	52	3	
CN	310	262	102	115	79	34	38	
CV	323	229	76	86	83	31	18	
EX	269	268	83	80	12	72	6	
LE	244	338	111	68	101	29	65	

Por ejemplo, AN:a1200 (=115) indica la media mayoritaria en la primera fila de la tabla anterior (48). Se muestra la tabla (49) que descompone R y A como resultado del análisis equilibrado doble.

Sin embargo, dado que la escala numérica en la tabla anterior no coincide con el valor observado real (frecuencia absoluta) de los datos, a continuación, la ajustamos a la escala del valor medido real. La siguiente tabla muestra las frecuencias normalizadas ajustadas por media:

(51)	*	a1200	a1300	a1400	a1500	a1600	a1700	a1800
AN		54	83	33	35	83	29	12
AR		16	33	11	13	49	24	1
CN		145	123	48	54	37	16	18
CV		152	108	36	40	39	14	9
EX		126	126	39	37	5	34	3
LE		115	159	52	32	47	13	31

Los siguientes dos gráficos se crearon a partir de una tabla cruzada no equilibrada y una tabla cruzada equilibrada, respectivamente. La línea gris oscura, dibujada en el centro del área gris clara, indica la ubicación de la media mayoritaria de la frecuencia normalizada regional en cada año. Por su parte, el área gris clara indica el área rodeada por las desviaciones medias superior e inferior, que nos permiten comprender la tendencia de los datos en la mitad media central²⁶. La línea horizontal gris en el gráfico indica la media mayoritaria de todos los datos.

²⁶ Cuando se indica una línea de tendencia general, generalmente se utilizan la mediana y los cuartiles, pero como la mediana no indica adecuadamente el centro del todo, utilizamos la línea central con la media mayoritaria. Cuando no se usa la mediana, el uso de cuartiles como el ancho del centro es inconsistente, por lo que usamos la desviación media para distinguir entre las partes superior e inferior. Si en este caso se utiliza la desviación estándar, es inexacta porque no distingue entre las dos partes.

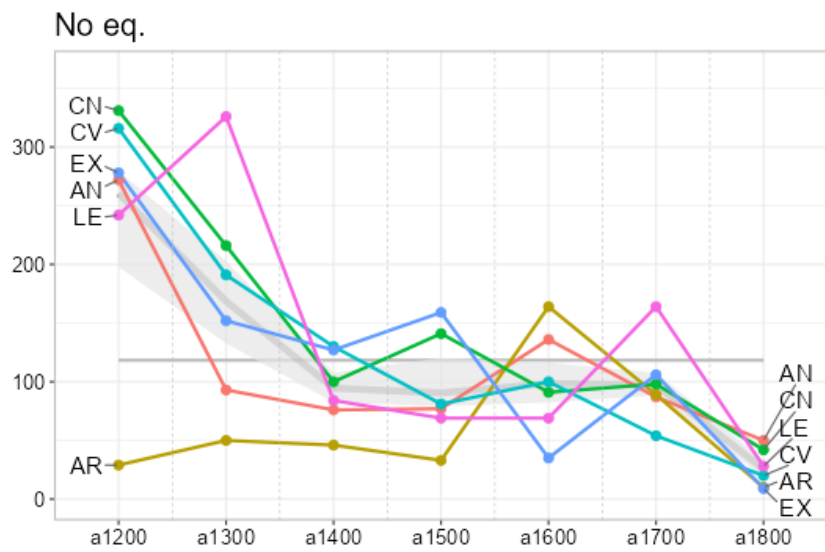


Fig.36. *Análisis no equilibrado. Región : Año (R:A)*

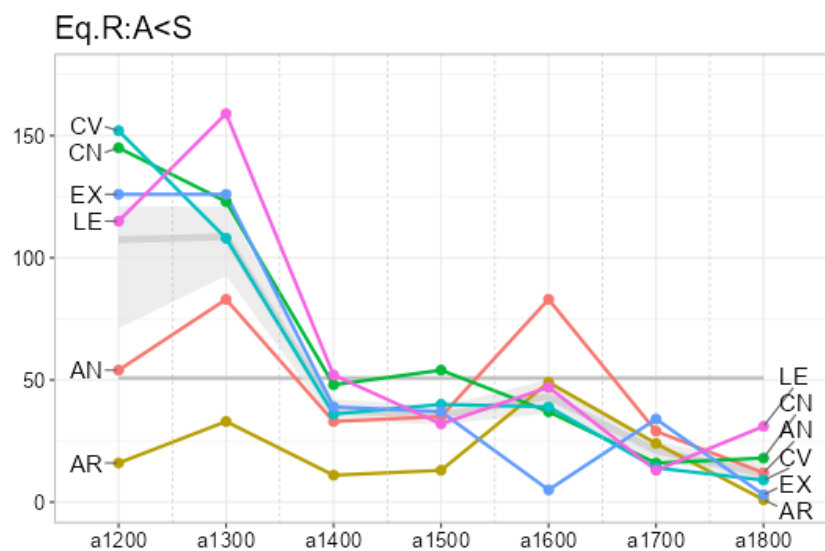


Fig.37. *Análisis equilibrado. Región : Año < Sociedad (R:A<S).*

Los dos gráficos son similares, pero en detalle, por ejemplo, AN:a1200 es muy diferente. En la fila AN:a1200 en la tabla anterior (R:A), se encuentra un valor alto de C (documento cancilleresco) (Tabla-48): 298. Sin embargo, la media mayoritaria de esta fila AN:a1200 es solo 115. Por tanto, consideramos que la frecuencia normalizada de datos no equilibrados (47) no indica el estado de a1200, sino más bien las características de C (cancilleresco). De esta manera, los datos no equilibrados son inexactos debido al sesgo en las variables sociales, por lo que los análisis deben realizarse utilizando datos equilibrados con referencia a la sociedad.

6.8. Final

El método estadístico utilizado en este estudio aborda problemas comúnmente observados en la tabulación cruzada que muestra variación cuantitativa. Exploramos las características de las distribuciones de datos con variaciones cuantitativas y sus especificidades, pero el problema es que la distribución de datos con variaciones tan específicas tiene distorsiones y sesgos. En ese momento, la distribución de todos los datos depende de algunos valores numéricos destacados y, por ello, la distribución de todos los valores numéricos no se muestra correctamente. Sería una gran contradicción (esencial) decir que cuando exploramos las características de una distribución de frecuencia buscando su especificidad, el hecho de que la distribución de frecuencia sea singular es un obstáculo. En este estudio, hemos propuesto una solución para resolver este problema utilizando varios métodos estadísticos.

En la "Encuesta del Carácter Nacional Japonés", que el Instituto de Matemática Estadística ha venido realizando cada cinco años desde 1953, se utiliza un método llamado "muestreo multietápico estratificado" para seleccionar a los entrevistados. Citamos el procedimiento a continuación (la traducción es nuestra.)²⁷.

Las ciudades, distritos, pueblos y aldeas de todo el país se han estratificado en seis áreas: distritos, ciudades con una población de 200.000 o más, ciudades con una población de 100.000 o más, ciudades y distritos con una población de menos de 100.000, y prefectura de Okinawa. A continuación, se seleccionó un total de 400 ciudades de distintas poblaciones de cada estrato mediante muestreo probabilístico proporcional. Finalmente, se tomó una muestra del número de personas asignadas a cada ubicación (promedio 16) a intervalos iguales del registro básico de residentes en los pueblos extraídos.

Observemos que los tamaños de población de los primeros seis estratos son diferentes. Las unidades más pequeñas, como pueblos y aldeas, se seleccionan entre seis estratos de diferentes tamaños. Si se adopta tal método, debería ser posible evitar el sesgo hacia las grandes ciudades, que tiende a ocurrir con el muestreo aleatorio general.

Sin embargo, con nuestros materiales históricos, solo podemos utilizar los datos existentes y es difícil controlar la muestra como en una encuesta por

²⁷ <https://www.ism.ac.jp/kokuminsei/page9/page10/index.html>

cuestionario. Por lo tanto, el sesgo entre distintas regiones y distintos estilos de escritura de sociedad se puede evitar promediando las frecuencias normalizadas calculadas para cada región y estilo de escritura de la sociedad de diferente tamaño, lo mismo que en el método de muestreo estratificado de múltiples etapas que toma en consideración el tamaño de la población y elimina el sesgo causado por grandes ciudades.

De esta manera, la "Media mayoritaria" de la frecuencia normalizada calculada a partir de todas las frecuencias y poblaciones no es un valor influenciado por una región o una sociedad en particular, sino un valor equilibrado debido a que todas las regiones y las sociedades se tratan en equilibrio utilizando los mismos estándares y métodos. Por ello, la transición de año se muestra correctamente sin sesgos. En este caso, en lugar de utilizar la media, que está influida por los valores atípicos, utilizamos la "Media mayoritaria", que incluye los valores atípicos y, al mismo tiempo, reduce la influencia de ellos.

El mismo análisis equilibrado está apoyado por distintos métodos estadísticos que se mencionan a continuación. Las tablas de las "Frecuencias normalizadas ajustadas por media de la frecuencia absoluta real observada" permite la apreciación de tamaño (escala) a la medida de la frecuencia observada, sin perder el mérito de la comparabilidad. Los gráficos con la "Desviación desde la mediana" permiten la observación de la tendencia general. En la "Tabla concentrada de manera diagonal" podemos interpretar los movimientos coherentes entre las filas, las columnas y la distribución de frecuencias en la tabla con el mismo concepto único, por ejemplo "tradición - innovación". Al mismo tiempo, analizamos los valores de distancia de filas y columnas por el "Análisis de agrupamiento por intervalo". Con la ayuda de los colores de puntos en los gráficos, sabemos a ciencia cierta el "Grado de significatividad" de cada frecuencia. Preparamos un diagrama de líneas para observar cambios dinámicos y un diagrama de puntos para observar variaciones estáticas.

En este estudio, hemos aplicado estos métodos estadísticos al análisis de cambios y variaciones de la estructura de verbo y pronombre átono en la historia de la lengua española.

FIN