

# スペイン語弱勢代名詞の位置の 歴史・地理・社会的変異

## 言語データ分析の問題と方法

上田博人

(2024)

# 1. 序論

本研究は、スペイン語の弱勢代名詞と動詞の位置関係について、その歴史・地域・社会における変化・変異を扱う。その位置関係には

- (1) 〈代名詞+活用形〉 (*lo+tiene*)
- (2) 〈活用形-代名詞〉 (*tiene-lo*)
- (3) 〈非活用形+代名詞〉 (*tener-lo*)
- (4) 〈代名詞+非活用形〉 (*lo+tener*)
- (5) 〈命令形+代名詞〉 (*ten-lo*)

という組み合わせがある。このように、すべての位置関係について *tener* と *lo* を使った例 (*lo+tiene*, *tiene-lo*, *ten-lo*, *tener-lo*, *lo+tener*) を示す。〈代名詞〉として〈弱勢人称代名詞〉 (*me*, *te*, *nos*, *os*, *lo*, *los*, *le*, *les*, *se*) だけを扱うので、以下では〈弱勢人称代名詞〉を単に〈代名詞〉 (p: pronombre) とする。動詞は〈活用形〉 (vc: verbo conjugado) と〈命令形〉 (ip: imperativo) と〈非活用形〉 (vn: verbo no conjugado) を区別する。〈命令形〉は肯定命令形であり、接続法を用いる否定命令形は含めない。〈非活用形〉のグループは、前置詞の結合について同じ統語法を示す不定詞 (if: infinitivo)、現在分詞 (g: gerundio)、過去分詞 (pv: participio verbal)、命令形 (ip: imperativo) を含む。代名詞が先行するときは「+」をつけて動詞の前に置き、後続するときは「-」をつけて動詞に続ける。これは、後述するように、動詞に後続する代名詞は動詞と一体化して単一語を成す、と考えられるためである。

現代スペイン語の代名詞は活用形の前に置かれるが (*le miraba*, *se detuvo*)、歴史的には動詞の後に付くことがかなり多かった (*mirábale*, *detúvose*)。動詞と離れて前置されることは *proclisis* と呼ばれ、動詞の後に付置されることは *enclisis* と呼ばれる。たとえば Benito Pérez Galdós (1843-1920) の作品には多数の *proclisis* と並んで少数の *enclisis* が見られる。*proclisis* と *enclisis* の出現は、主として動詞の形態(活用形・非活用形・命令形)と文中の位置(文頭と文中)という言語内的条件と、年代・地域・社会という言語外的条件に従っている。

以下では、これらの言語内的条件と言語外的条件に従って *proclisis* と *enclisis* を比較する。資料は CODEA の最新の歴史コーパス(4023 文書から年代、地域、社会を選択し約 180 万語を抽出)を使用する<sup>1</sup>。本研究は Ueda

---

<sup>1</sup> «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]

(2024)に続くものである。統計的方法の詳細については6.【付録】で扱う。

## 2. 方法

以下は Hanssen (1913), Menéndez Pidal (1976, 1980), Lapesa (1982), Penny (2006)の中の、中世スペイン語において優勢であった〈動詞-代名詞〉の語順と、文頭および接続詞 *e* (*et*)と *mas* の後で現れた〈活用形-代名詞〉という語順が言及されている箇所である(下線と[...]:筆者)。

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 vulgar 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)

Menéndez Pidal (1980)

- El pronombre átono va después del verbo, si éste encabeza la frase o va precedido sólo de la conjunción *et*: *faca nos Deus...* GIEmil 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)

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)

寺崎(2011: 188)

無強勢代名詞は前接語であり、原則として定型動詞の後に付加する。つまり後置される。休止の後、つまり文頭に置くことはできない。言い換えると、無強勢代名詞で文を始めることはできない。文頭でない場合でも、無強勢である接続詞 *e / mas* の直後に無強勢代名詞を置くことはできない。

中岡(1993: 35)

中世スペイン語では、人称代名詞弱勢形は前接語[動詞に後置する接語]であった。従って、これは休止の直後に位置して、文や句の先頭には位置しえなかった。。つまり、近代語の"Lo vi todo"「私はそれを全部見た」という語順は、中世語にはなかった。

中世スペイン語で人称代名詞弱勢形の位置は、原則として次のようになる。

(A) 動詞が文頭に立つときまたは休止の直後にくるときには、動詞の直後に置いた。Partios' de la puerta [...] (*Cid* 51) / Armaronse e tañeron vna tronpeta. Al son della, oyéronlo los de la gran galera [...] (LLEM [F. González Ollé, *Lengua y literatura españolas medievales, textos y glosario*. Ariel. Barcelona, 1980], 318; Victorial)

(B) 強勢をもつ語が動詞に先行するときには、動詞の前に置いた。[...] Decid me uostros nombres, nom los querades celar. [...] (*Auto RM*. 81)

一方、中岡(同: 40)は中世語で〈代名詞+動詞〉が僅かではあるが機能していた、と述べている。

中世語では基本的には前接辞[動詞に後置する接語]であった代名詞弱勢形は、僅かながらも、後接辞[動詞に前置する接語]として機能している場合もみられる。従って、この近代語的な特徴もかなり早くから潜在的にあった、ということになる。(下線と[...]:筆者)

この後に、文学作品の中から前接辞のある例文だけを 3 例載せている (*l'anda, m'an, te demostraré*)。引用した「近代語的な特徴」(〈代名詞+動詞〉)が「かなり早くから潜在的にあった」という説明は示唆的である。筆者は、本研究において、文学作品に代えて公証文書を調査すれば、この潜

在的な特徴が顕在化することを実証するつもりである<sup>2</sup>。

スペイン語史の先行研究は文学語(*lengua literaria*: 文学作品の言語)を扱っているようであるが、文学語は伝統的な古い規範に従っていることが多い。一方、本研究は過去の言語社会の実相を反映する公証文書を扱う。はたして、中世語の公証文書において〈代名詞+動詞〉(*lo+tiene*)は一定の条件のもとで限られたのであろうか、または基本的に現代語と同じように〈代名詞+動詞〉が多く使われていたのであろうか。本研究では大量の非文学資料を使って、この問題を明らかにする。

ここで、〈動詞-代名詞〉を〈代名詞後置〉(*enclisis del pronombre*)と呼び、〈代名詞+動詞〉を〈代名詞前置〉(*proclisis del pronombre*)と呼ぶことにする。すなわち中世では〈代名詞後置〉が一般に認められているが、近世以降(1500-)では逆に〈代名詞前置〉になる、と言われている。よって、次の歴史的变化が想定されている(V:動詞, p:代名詞)。

歴史的变化：[中世]〈代名詞後置〉→[近世]〈代名詞前置〉

この歴史的变化の想定は正しいであろうか。

中世語の代名詞の位置について Penny (2006)は動詞の〈活用形〉(*verbo conjugado*)と〈非活用形〉(*verbo no conjugado*)を区別しないが、近代を含めると両者の出現頻度は大きく異なるので、全歴史を扱う本研究ではこれらを区別する。〈非活用形〉には不定詞・現在分詞・過去分詞を含める。命令形は形態論体系の中では活用形であるが、形態論のレベルにおいて、特殊な不規則形は命令形でのみ使われ、否定命令形では使われない(*ven, no vengas; vete, no te vayas, etc.*)。また、統語論(代名詞との位置関係)の面では、以下で見るように、むしろ非活用形(不定詞・現在分詞・過去分詞)と同じ変化・変異を示している。よって、代名詞の位置については、命令形と否定命令形は形態論(特殊な不規則形の有無)と統語論(代名詞の位置)のどちらのレベルでも異なる体系に属する。

---

<sup>2</sup> この研究を口頭で発表した際(2024/10/26)、川上茂信氏、福嶋教隆氏、寺崎英樹氏から貴重なコメントをいただいた。発表では、私は中岡(1993)の引用箇所を誤解していたが、本稿では、いただいたコメントに従って以上の3段落を訂正した。発表後に受信した福嶋氏からのメール(ご許可をいただいて以下に引用)の中の「今日引用された[中岡(同)] p. 40の文言は川上さんのご指摘のとおり、有標のケースとして扱っておられるのだと思います。でも、この有標のケースをきちんと指摘している点で、中岡先生のご本は『～入門』と題しながら、上田さんの主張を予言するような野心的な主張をそっと忍ばせている、と言えるのではないのでしょうか。」というご意見は中岡(同)の重要性をお認めになり、私には大きな励みになった。三名の方々に感謝申し上げます。

最初に次の仮説を設定する<sup>3</sup>。

【仮説】 中世から近世・現代まで通して〈代名詞+活用形〉が無標の語順であった。ラテン語から継承した〈活用形-代名詞〉は有標の語順であり、一定の条件のもとで最近まで残存した。一方、〈非活用形-代名詞〉は独自の発展を遂げた。

この仮説の妥当性を検証しながら、次の3つの問題の解決を試みる。

問題-1: 中世語(文献資料:1200-1500)では *enclisis (tiene-lo)*のほうが多く、*proclisis (lo+tiene)*は一定の条件のもとに限られたのであろうか。

問題-2: 全歴史(文献資料:1200-1800)を通じて動詞活用形において *enclisis (tiene-lo)*から *proclisis (lo+tiene)*への移行という変化は認められるであろうか。

問題-3: 逆に、全歴史を通じて動詞非活用形において *proclisis (lo+tener)*から *enclisis (tener-lo)*への移行という変化は認められるであろうか。

問題-4: 否定命令形と異なる肯定命令形の *enclisis (ten-lo)*の歴史理由は何か。

次が全資料の中で見つかった高頻度の語形とその総頻度である(絶対頻度の高頻度順の上位語)。

1. 〈代名詞+活用形〉 (*lo+tiene*): *se contiene* (頻度: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), ... [計: 35,986].

2. 〈活用形-代名詞〉 (*tiene-lo*): *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), *confírmoles* (22), *confirmámoslo* (22), *obligámosnos* (18), *abrióse* (15), *sépase* (15), *otorgámosles* (15), *entróse* (14), ... [計: 2,135].

3. 〈非活用形-代名詞〉 (*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), *dádoles* (22), *averle* (21), *darla* (21),

<sup>3</sup> これは本研究が達した結論であるが、最初に仮説として掲げることで以下の展開が明瞭になると思われる。

siéndole (21)... [計: 5,842].

4. 〈代名詞+非活用形〉 (*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 complir (11), vos redrar (11)... [計: 1,536].

5. 〈命令形-代名詞〉 (*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)... [計: 303].

文法カテゴリーには次の略号を用いる<sup>4</sup>。

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, sustantivo, 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.

以下では、上の 1 - 5 の順で行った「均衡分析」(→ 6. Apéndice)の結果を示す。グラフの出力は次のような形式になる。表題の p は pronombre, v は verbo, Pos は posición を示す。言語外変数として、地域(R: región), 年代(A: año), 社会(S: sociedad)を使用する。各分析では、次の 2 つの変数の組み合わせを使った。「<」の左が対象変数を示し、右が参照変数を示し、R, A, S はそれぞれ región, año, sociedad を示す。

(a) R:A<S 地域(R):年代(A)を対象変数とし、社会(S)を参照変数とする。

(b) S:A<R 社会(S):年代(A)を対象変数とし、地域(R)を参照変数とする。

線グラフのポイントでは次の基準に従って「有意性」のグレードを色で示した (acp: adjusted cumulative probability) (→ 6. Apéndice)。

- 1.青:  $acp > 0.99$  (有意に非常に大きい)
- 2.薄青:  $0.95 < acp \leq 0.99$  (有意に大きい)
- 3.白:  $0.05 \leq acp \leq 0.95$  (有意でない)
- 4.薄赤:  $0.01 \leq acp < 0.05$  (有意に小さい)

---

<sup>4</sup> p は本論の中では代名詞 (pronombre) を示し、資料の中では前置詞 (preposición) を示す。資料中の cl (clítico) は本文中では p とする。

5.赤:  $acp < 0.01$  (有意に非常に小さい)

縦軸は「安全頻度」を示す。安全頻度は実測値(出現頻度)、母数(総和)、安全率(=0.95)から期待確率を求め、その期待確率と基準数を使って求める(→ 6. Apéndice)。

「均衡分析」(→ 6. Apéndice)は、対象とする2つの変数(〈対象変数〉)を合わせた頻度と参照する変数(「参照変数」)を集計し、その平均(「大数平均値」)を2つの対象変数に分解する方法である。

すべての数値は実測値(出現頻度)の平均値に合わせて調整し、出現頻度の規模と等しくなるようにした。この方法によって、数値の絶対的な大小の実態がわかる。

資料 CODEA に収録された文書は1100年代までは少ないので、ここでは1200年代から1800年代までの文書を扱う: a1200, a1300, ..., a1800。年代の数値はそれぞれの開始年を示す(たとえば a1200: 1200-1299)。

本研究の対象とする地域的変異は次のように分類される<sup>5</sup>:

AN: Andalucía

AR: Aragón

CN: Castilla la Nueva

CV: Castilla la Vieja

EX: Extremadura

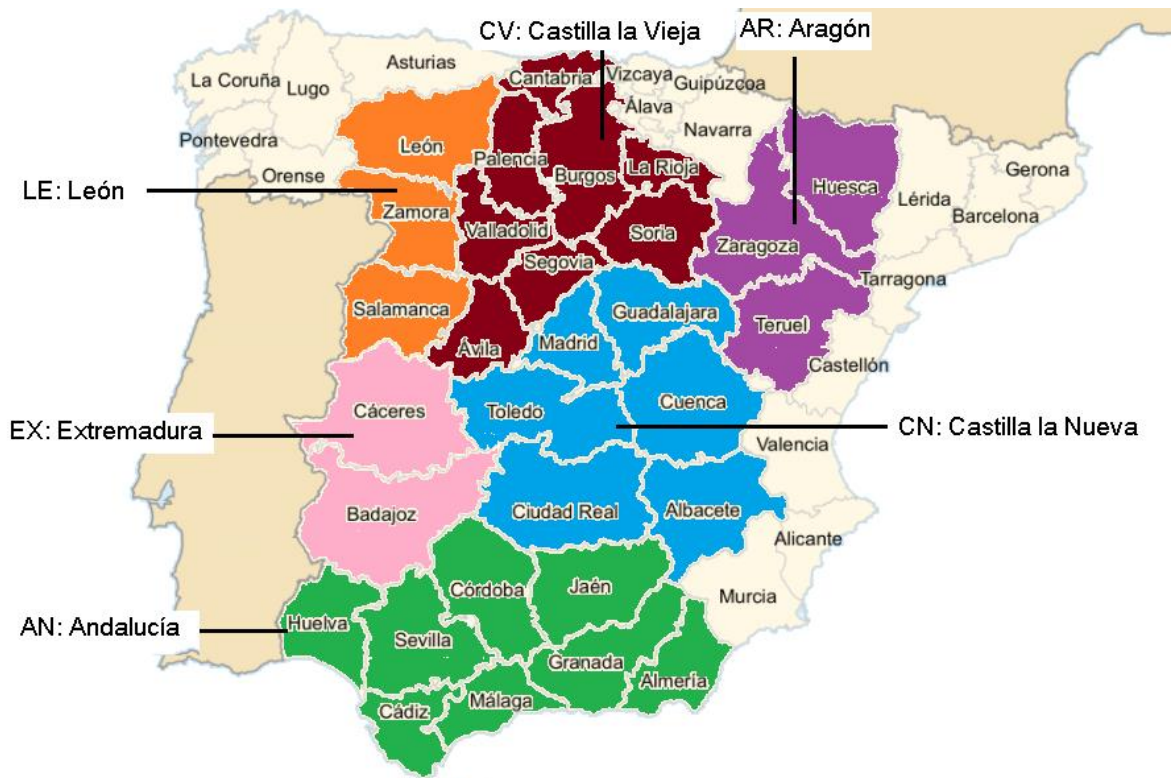
LE: León

---

<sup>5</sup> スペイン語歴史方言学の研究を行うための区分として現在の自治州よりも旧地方名のほうが適している。León, Castilla la Nueva, Castilla la Vieja は1983年まで使用されていた旧行政地方名である。Castilla la Nueva は主に現在の Castilla La Mancha 自治州と Madrid 自治州を含む。Castilla la Vieja と León は現在の Castilla y León 自治州に対応する。



それぞれの地方はそれぞれ次の県を含む。



【図-1a,b】本研究の対象とする旧地方名

次は総語数を地域変数と年代変数でクロスした集計表である。

| (1) | *      | a1200  | a1300  | a1400   | a1500   | a1600  | a1700  | a1800 |
|-----|--------|--------|--------|---------|---------|--------|--------|-------|
| 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 |       |

次は社会的変数の分類である。

C: documento canclleresco (王室文書)

E: documento eclesiástico (教会文書)

J: documento judicial (司法文書)

M: documento municipal (行政文書)

P: documento particular (私文書)

次は総語数を社会変数と年代変数でクロスした集計表である。

|     |   |        |        |         |         |         |         |        |
|-----|---|--------|--------|---------|---------|---------|---------|--------|
| (2) | * | a1200  | a1300  | a1400   | a1500   | a1600   | a1700   | a1800  |
|     | C | 62,486 | 48,572 | 68,242  | 71,492  | 24,844  | 657     | 184    |
|     | E | 72,567 | 78,106 | 109,047 | 53,243  | 29,980  | 2,812   | 2,759  |
|     | J | 7,487  | 5,986  | 17,922  | 68,605  | 86,451  | 82,786  | 7,034  |
|     | M | 5,052  | 8,913  | 10,103  | 37,124  | 13,299  | 9,821   | 21,528 |
|     | P | 31,801 | 65,612 | 77,629  | 159,930 | 214,521 | 134,256 | 97,366 |

次は総語数を地域変数と社会変数でクロスした集計表である。

|     |    |         |         |         |        |         |
|-----|----|---------|---------|---------|--------|---------|
| (3) | *  | C       | E       | J       | M      | P       |
|     | AN | 45,028  | 15,838  | 37,819  | 19,271 | 98,880  |
|     | AR | 3,389   | 64,535  | 17,979  | 3,945  | 93,328  |
|     | CN | 94,550  | 48,278  | 175,621 | 54,574 | 152,567 |
|     | CV | 119,973 | 107,543 | 28,561  | 11,559 | 112,227 |
|     | EX | 1,146   | 34,730  | 10,324  | 10,736 | 253,964 |
|     | LE | 12,391  | 77,590  | 5,967   | 5,755  | 70,149  |

分析には R (R Core Team: 2023) と ggplot (Wickham: 2016) を使用した。

### 3. 分析

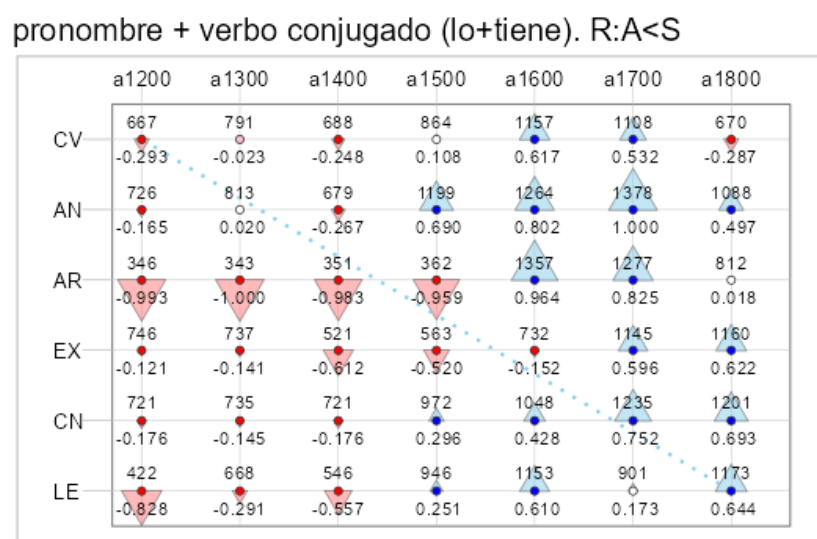
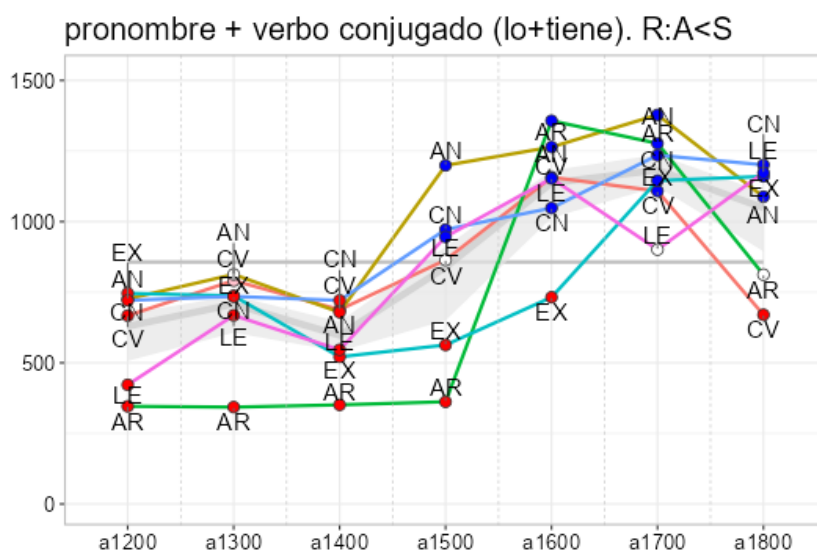
#### 3.1. 〈代名詞+活用形〉 (*lo+tiene*)

##### 観察

はじめに、全体で最も頻度の高い〈代名詞+活用形〉 (*lo+tiene*) について、均衡分析(1)-地域(対象変数=地域:年代、参照変数=社会)と均衡分析(2)-社会(対象変数=社会:年代、参照変数=地域)を行う。グラフは年代変化を扱う折れ線グラフと数値の変異を示すプロットグラフを用意する。折れ線グラフには全体の平均値を灰色の水平線で示し、それぞれの年代の「大数平均値」(上田 2024b)を濃い灰色の線で示し、上下の平均偏差で囲む領域を薄いグレーの面で示した。

**該当する形式**(頻度順): se contiene (頻度: 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)... [計: 35,986].

均衡分析(1)-地域. 対象変数=地域：年代、参照変数=社会.



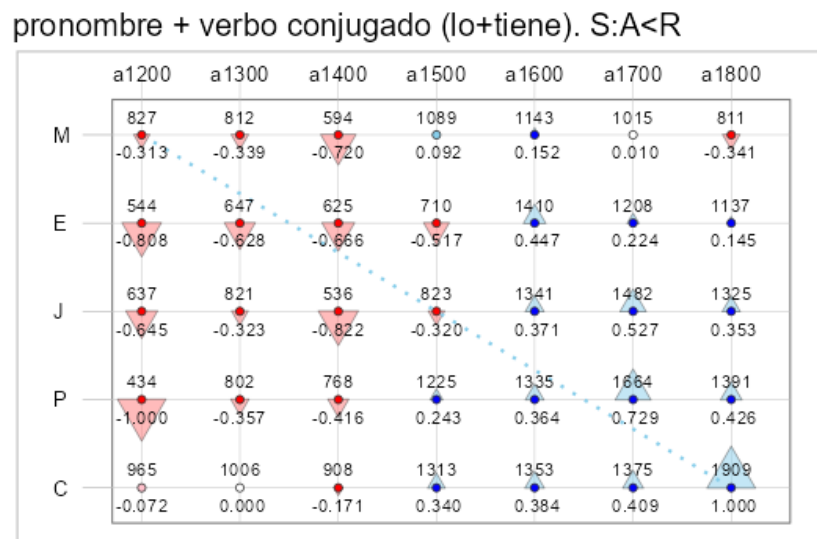
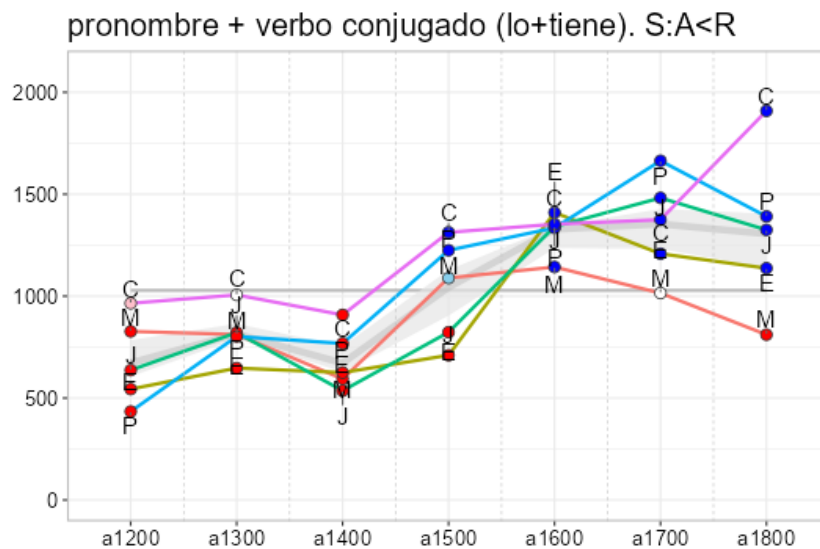
【図-2a,b】 〈代名詞+活用形〉地域:年代<社会 (平均: 857)

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

全体的に高頻度である。中世において〈活用形+代名詞〉(*tiene-lo*)が通常用いられ、〈代名詞+活用形〉(*lo+tiene*)は比較的少数であったことが先行文献に指摘されているが、本研究では上の図が示すようにかなり多数の〈代名詞+活用形〉(*lo+tiene*)が記録された(後述するように、全歴史を通して〈代名詞+活用形〉(*lo+tiene*)のほうが圧倒的に多かった→3.6)。しかし、〈代名詞+活用形〉だけを見るならば、中世では近世と比べて比較的少数であったことも確かである(グラフの有意な少数-赤)。よって、代名詞後置は中世の全

体的傾向(決定的な規則ではない)として認められる。

均衡分析(2)-社会. 対象変数=社会:年代、参照変数=地域.



【図-3a, b】 〈代名詞+活用形〉社会:年代<地域 (平均: 1028)

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

**観察:**全体的に高頻度である。社会変数については、中世では〈代名詞+活用形〉(lo+tiene)の使用は相対的に少なかったが、王室文書(C)では比較的多かった。a1500以降は全体的に上昇した。

## 考察

### (1) 〈代名詞+活用形〉に先行する語 (中世スペイン語)

次は、文頭(または休止の後) (#)・文中(=)という条件(x)を比較して〈代名詞+活用形〉 (*lo+tiene*) (y)の対称値(r)を計算した結果である。総語数(s)は関連する代名詞と動詞の組み合わせの合計値である(資料は a1200-a1400 の 2 つの語順 *lo+tiene*, *tiene-lo* を抽出したもの)。

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

【表-1】 文頭(#)・文中(=)と〈代名詞+活用形〉の関係

この結果は、条件が文頭(#)であるとき〈代名詞+動詞〉 (*lo+tiene*)が比較的出現しにくく、条件が文中(=)であるとき、出現しやすいことを示している<sup>6</sup>。しかし、それでも、文頭において〈代名詞+動詞〉 (*lo+tiene*)の語順のほうが(=19)、逆の語順(=17)よりも多かった。

次は文頭(または休止の後) (#)に現れた〈代名詞+活用形〉 (*lo+tiene*) のリストである。

# cl, v\_cj の語形: # 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.  
[計:19]

この語形の数(a=19)は、同じ位置(文頭、または休止の後)に現れた〈活用形-代名詞+〉 (*tiene-lo*)の語形の数(b=17)よりも多い。これは、先に見た先行研究の見解(「文頭または休止の後では代名詞は動詞の前に現れない」→2)と大きく異なる。

# v\_cl の語形: # Agola, # Áse, # concertámoslo, # Danles, # Fágovos, # mandámosla, # Mándome, # oblígome, # Vendémosle, # vendémosvos, # Vendiónos, # Vendívos, # Véndovollo, # Véndovos. [計:17]

次は、中岡(1993: 35)の「強勢をもつ語が動詞に先行するときには、動詞の前に置いた」という文学語に関する指摘が、ここで扱っている公証文書に当てはまるか否かを検証した結果である:

<sup>6</sup> ただし、文頭にある動詞句の例は少ない。

| 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 |

【表-2】先行語の弱勢(s.x=a)・強勢(s.x=t)と〈代名詞+活用形〉の関係

たしかに先行語が強勢語であるとき(t)、代名詞が前置することが多い(a)。しかし、逆に、先行語が弱勢語であるときも(a)、代名詞が前置することが多いのである(a)。対称値(r)は先行語が強勢語であるとき(t)のほうがやや高いが、その差は小さい。よって、この資料からは、むしろ、先行語の弱勢・強勢の違いに関係なく、両者とも代名詞が前置することが多かった、と言える。

次に、条件として先行する語の品詞を調べる。先行する語は高頻度語を取り上げる<sup>7</sup>。

| 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 |
| c      | cl, lv_cj | 3,884 | 9,819 | 3,294 | 590 | 6,525 | 546   | 10,955 | 0.696 |
| a      | cl, lv_cj | 162   | 9,819 | 134   | 28  | 9,685 | 1,108 | 10,955 | 0.654 |

【表-3】先行する語の品詞と〈代名詞+活用形〉の関係

このように、多くの品詞が〈代名詞+活用形〉(lo+tiene)を選択し、〈活用形-代名詞〉(tiene-lo)を避けている。後述するように、〈活用形-代名詞〉を後続する形式として優先的に選択する品詞はないので、とくに品詞の区別なく、〈代名詞+活用形〉の語順が一般であった、と言える。先に見たように(→2.)、そしてさらに後述するように(→4.1.)、〈代名詞+活用形〉の語順は4種の語順の中で最多数の語順であるので、これが無標項である。無標項であれば、その〈代名詞+活用形〉の語順が出現する条件を探索してもあまり意味はないかもしれない。むしろ、無標項である〈代名詞+活用形〉の語順が出現しにくい、語頭という位置に注目すべきである。

<sup>7</sup> 低頻度語で対称値が高くても有意性は低いからである。

Menéndez Pidal (1976: 403)と Lapesa (1982: 218)は特定の語形 *e* と *mas* を取り上げて、その直後で代名詞は動詞の後になる、と述べている。先に示した品詞のリストには関係代名詞と接続詞があるので、その中から比較的高頻度の *como, do, e, mas, porque, que, si* を取り上げて、それらに続く〈代名詞+活用形〉の語順との関係を調べると、次のような結果になった。

|        | 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 |

【表-4】 先行する語形と〈代名詞+活用形〉の関係

このように、*mas* の後の位置では〈代名詞+活用形〉になるので、上の検索の結果は Menéndez Pidal (*ibid.*)と Lapesa (*ibid.*)の見解と一致する。しかし、*e* には多くの例外があるので、一致していない。

従位接続詞(*como, si, do, porque, que*)は従属文を形成する。Castillo Lluch (1996: 142-197)が綿密に調べた中世文献の従属文中のほとんどのケースで代名詞が動詞の前に置かれていることを参照すると、ほとんど完全に〈代名詞+活用形〉の語順(無標)になることは不思議ではない。

## (2) 名詞句型パターン

本節(3.1)では、〈代名詞+活用形〉(*lo+tiene*)が高頻度であったことを見た。後述するように(→4.1)、この形式の頻度は〈活用形-代名詞〉(*tiene-lo*)と比較して圧倒的に多数であった。

〈代名詞+活用形〉(*lo+tiene*)が圧倒的に多数であった理由として、スペイン語の言語体系全体で〈弱勢語+強勢語〉という形態が多く見られることが挙げられる。すなわち〈弱勢語+強勢語〉という連続は〈定冠詞・所有詞+名詞・形容詞〉〈前置詞+名詞・形容詞〉〈接続詞+名詞・形容詞・動詞・副詞〉〈関係詞+名詞・形容詞・動詞・副詞〉に見られ、非常に多いのである。

これらの〈弱勢語+強勢語〉の中で、とくに定冠詞と所有詞は代名詞と形態的に類似しているので(*lo, la, los, las, mi-me, tu-te, su-se*)、前置活用形(*lo+tiene*)が圧倒的多数を占めるようになったと思われる。また、〈代名詞(範列)+活用形(屈折)〉と〈定冠詞(範列)・所有詞(範列)+名詞(屈折)〉は形態

変化(範列と屈折)の存在が共通している。

さらに、両者の屈折の要素の中の高頻度の-s が共通する(名詞の複数形、動詞の 2 人称単数形)。〈代名詞+活用形〉と〈定冠詞・所有詞+名詞(屈折)〉は〈機能語+内容語〉という構造を共有している。そして、両者共に機能語が、内容語の意味を限定している。

意味については、たとえば、*tu gusto, su casa* という名詞句と *te gusta, se casa* という動詞句は文法的な機能は異なるが、意味的にはパラレルである。両者の関係を認めなければ、その形態と意味は偶然に類似したことになるが、それは考えられない。この類似は歴史的にはラテン語、そして想定されるインドヨーロッパ祖語の共通語源に由来する。それが連綿と中世のカスティーリャ語に伝わり、その話者が、とくに名詞句と動詞句という文法範疇(現代の規範文法)を意識しなくても、根底にある共通の原理に基づいて言語を使用していたことが想像できる。

ここで、〈代名詞+活用形〉が〈弱勢語+強勢語〉と、とくに〈定冠詞・所有+名詞〉と、形態的・構造的・意味的に次のように類似することに注目したい(上田 2024)。

|     |  |          |
|-----|--|----------|
|     | 定冠詞・所有詞  | 名詞       |
| 名詞句 | [範列: <i>el, los, la, las, lo; mi, tu, nuestro, vuestro, su</i> ] | [屈折: -s] |
|     | 代名詞  | 活用形      |
| 動詞句 | [範列: <i>lo, los, la, las, le, les, me, te, nos, os, se</i> ]     | [屈折: -s] |

以下では、〈弱勢語+強勢語〉の頻度分布を他の組み合わせと比較しながら確認する。はじめに、資料中の全単語を弱勢語と強勢語に分類する。次の品詞が弱勢語である<sup>8</sup>。これら以外は強勢語とする。

定冠詞(*el, los, la, las, lo*); 関係詞(*adonde, cual, cuala, cuales, cuan, cuando, cuanta, cuantas, cuanto, cuantos, cuya, cuyas, cuyo, cuyos, do, donde, que, quien, quienes*); 接続詞(*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*); 接語(*la, las, le, les, lo, los, me,*

<sup>8</sup> 前置詞・関係副詞 *según* は強勢語である(Quilis 1981: 315-316)。弱勢語には次の名詞・形容詞・数詞を含める。*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 Blecua 1975: 443)。Dalbor (1969: 215)は *señor Martínez* を強勢語とするが、Stirling (1935: 60)は弱勢語としている。本研究では称号(título)のうち、歴史的に語末音脱落があった形式(*don, doña, san, santa, fray*)だけを弱勢語とする。



*mi, nos, os, se, se la, se las, se les, se lo, se los, te*);前置詞(*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 'hasta'*)、前置詞-定冠詞(*al, del*); 前置所有詞(*lur, lures 'de ellos', mi, mis, nuestra, nuestras, nuestro, nuestros, su, sus, tu, tus, vuestra, vuestras, vuestro, vuestros*), 記号 -, ~, ,, ;, :, !, i, ?, ¿, .., …, ·, ', ' , “, ”, «, (, ), [, ], {, }, §, ¶, /, #, ·, ^, +, =, |, —, | .

次は強勢語と弱勢語の総頻度である。

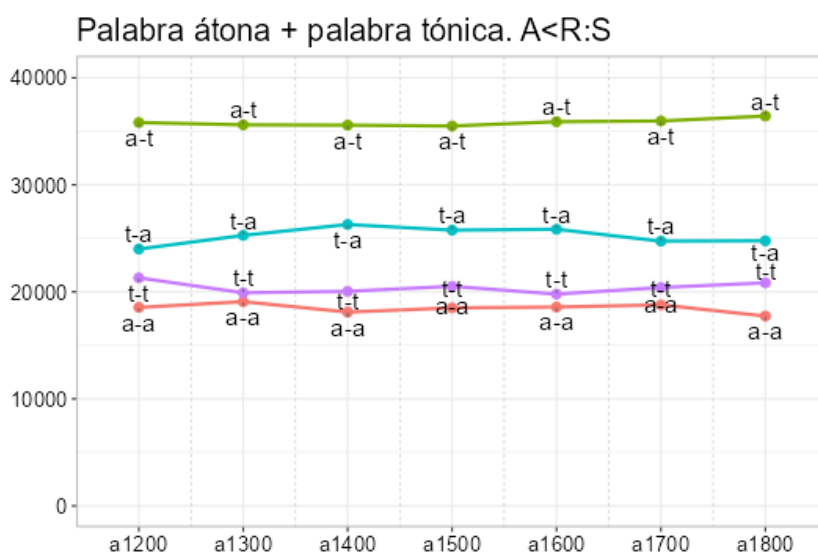
強勢語 (1,023,061: 57%), 弱勢語 (765,156: 43%) [計: 1,788,217].

次は、全資料中のすべての 2 語連続の組み合わせとその頻度を示すリストである。

- (1) 〈弱勢語+強勢語〉 *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* (1,046), *y dos* (1,032), *la villa* (971), *en ella* (960)... [計: 534,830].
- (2) 〈強勢語+弱勢語〉 *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)... [計: 425,960].
- (3) 〈強勢語+強勢語〉 *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)... [計: 361,697].
- (4) 〈弱勢語+弱勢語〉 *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)... [計: 215,352].

このように、(1) 〈弱勢語+強勢語〉が最大の頻度を示しているので、最重要の 2 語連続と考えられる。

次のグラフは上の 4 つの組み合わせの出現頻度の歴史的推移を示す(a:弱勢語, t:強勢語)。



【図-4a,b】弱勢語(a)と強勢語(t)の組み合わせの推移

上の図は全歴史を通じて常に〈弱勢語+強勢語〉(a+t)が最大の頻度であったことを示している。

次は、すべての2語の品詞連続の内訳を調べた結果である<sup>9</sup>。

すべての2語の連続の内訳: 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)... [計: 1,788,217]

このリストの中に、高頻度の〈定冠詞+名詞〉(ar\_df+n), 〈所有詞+名詞〉(psa+n), 〈代名詞+動詞〉(cl+v)が含まれている(太字で示した)。

以上の頻度データから、〈代名詞+動詞〉の語順が、最も高頻度の〈弱勢語+強勢語〉のパターンに合致し、その中でとくに形態が類似して高頻度の〈定冠詞・所有詞+名詞〉の語順に支えられていることが想定される。

次は、全資料(a1200-a1800)の中の高頻度の〈定冠詞+名詞〉, 〈所有詞+名

<sup>9</sup> 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.

詞〉，〈代名詞+活用形〉の語形のリストである。

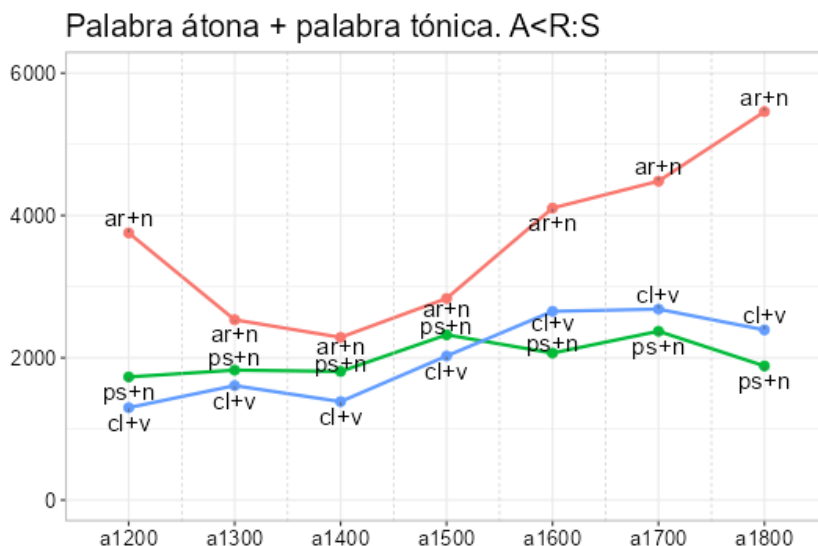
〈定冠詞+名詞〉：la villa (971), el señor (645), el rey (642), el día (579), la gracia (490), la ley (463), la iglesia (452), la casa (425), los bienes (417), el escrivano (396), la carta (388), la ciudad (386), las cartas (376), la orden (364), las partes (349), la iglesia (347), la ciudad (292), la manera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

〈所有詞+名詞〉：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)... [計: 34,872].

〈代名詞+動詞〉：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)... [計: 35,986].

このリストを見ると、〈短い語+長い語〉という語形のパターンが一部の少数の例外を除いて全体的に非常に類似していることがわかる。このことも〈定冠詞・所有詞+名詞〉と〈代名詞+活用形〉の間の相関を想定する理由である。

次のグラフは〈弱勢語+強勢語〉の名詞句と動詞句を選んで年代的推移を調べたものである(〈定冠詞+名詞〉 ar+n: artículo definido + nombre, 〈所有詞+名詞〉 ps+n: posesivo átono + nombre, 〈代名詞+活用形〉 p+v.cj: pronombre átono + verbo conjugado):



【図-5a,b】 〈弱勢語+強勢語〉の名詞句と動詞句の年代的推移

上のグラフは〈定冠詞・所有詞+名詞〉(ar+n, ps+n)と〈代名詞+活用形〉(cl+v)は共に、定常的に高頻度であったことを示している。

このように、代名詞と動詞の4つの組み合わせの中で圧倒的多数を占める〈代名詞+活用形〉(*lo+tiene*s)は、一般の〈弱勢語+強勢語〉のパターンに適合し、その中でとくに多数の〈定冠詞・所有詞+名詞〉のパターンと形態・構造・意味・出現頻度が類似することからその類推による影響を受けた無標の構造であった、と言える。

一般に〈弱勢語+強勢語〉が最大の頻度を示す理由として、次が考えられる。スペイン語の語形の最も頻度が高い強勢の配置は

...'CVC(V[s,n])# (C:子音、V:母音)

となる(上田 2021)。この括弧内の(V[s,n])#の出現は自由であり、先行するCは歯・歯茎調音である/r, s, n, l, d, z, y/の場合にほぼ限られる。歯・歯茎音以外の場合は母音が付加される(*capa*, *cabo*, *coche*, *tomo*, *rico*, *pago*)<sup>10</sup>。このように基本的に語末に強勢があることが強勢語の特徴である。

<sup>10</sup> *club*, *reloj*, *álbum*, *déficit* などの外来語・ラテン語は例外である。

語が結合して語群を形成するとき、その末尾は必ず強勢語でなければならない。Navarro Tomás (1972: 277-305)の音声表記には休止(*pausa*)の長さに応じて、|||, ||, |, ' の4種を区別している。たとえば、次の部分を取り上げよう<sup>11</sup>。

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. |||

この中にある様々な休止の前はすべて強勢語である(Navarro Tomás *ibid.* 278)。そして、文末には英語のように弱勢語が現れることはない(たとえば *Where are you from?*, *He is the friend I was thinking of*)。

このように、スペイン語には、[語・語群・文]の末尾は必ず強勢(語)となる、という「最終位置の強勢」という特徴がある<sup>12</sup>。この特徴が〈定冠詞・所有詞+名詞〉(ar+n, ps+n)と〈代名詞+動詞〉(cl+v)にも共有されていることも、両者間で近似した歴史的変化があったことの原因の一つとして考えられる。

周知のように〈定冠詞+名詞〉は古典ラテン語に存在せず、ロマンス諸語は古典ラテン語の〈指示代名詞+名詞〉*ille homine(m)*に由来する。この指示代名詞には強勢があったが、後にスペイン語では *ille homine* > *elo homne* > *el hombre* の変化の中で文法化して弱勢の定冠詞となった。所有詞前置形も初期スペイン語では強勢語(*mió / mío fiyo*)であったが、次第に弱勢化して今日の形態となった(*mi hijo*)。同様にして〈代名詞+活用形〉の代名詞も初期には指示代名詞として強勢があったが後に弱勢化した人称代名詞となった。

この3つの平行した構造(〈定冠詞+名詞〉〈所有詞+名詞〉〈代名詞+動詞〉)とその変化が偶然の一致ではないとすれば、それらの底流に〈弱勢語+強勢語〉という共通のパターンが存在したことが仮定できる。これには、1つの形態が他の形態の類推作用によって、類似した形態に合流したことも要因として考えられる。その文法化の合流は、中世における古形の残存状態と頻度分布(図-5)を考慮すると、〈定冠詞+名詞〉→〈所有詞+名詞〉→〈代名詞+動詞〉という順番が考えられるかもしれない。しかし、現実的には、互いの類推作用の結果、共通のパターンが形成され、その形成されたパ

<sup>11</sup> 引用したテキストは音声表記ではなく、Juan Valera. *Pepita Jiménez* の原文に Navarro Tomás の音声表記にある休止記号を付したものである。

<sup>12</sup> これは一般にロマンス諸語の特徴である。とくにフランス語に顕著に見られる。一方、英語は「開始位置の強勢」が顕著である。

タンが、次第に強化されて多数を占め、無標化したことが想定できる。

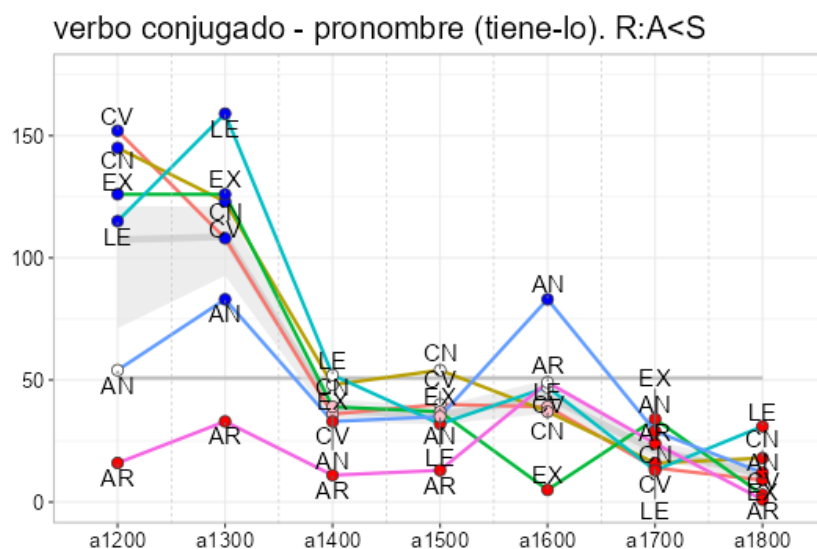
### 3.2. 〈活用形-代名詞〉 (*tiene-lo*)

#### 観察

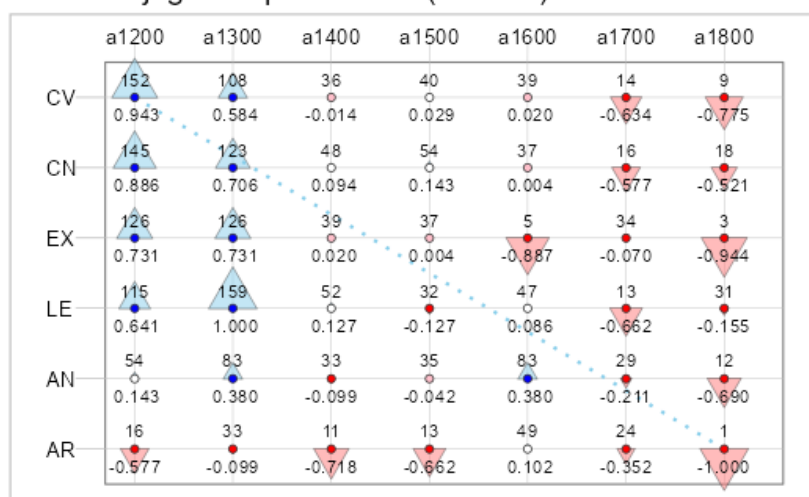
スペイン語史の研究で取り上げられている〈活用形-代名詞〉 (*tiene-lo*)は主に中世の資料で高頻度で観察されるが、実際は公証文書においては〈代名詞+活用形〉 (*lo+tiene*)と比較して非常に少数であった。

**該当する形式**(頻度順) : *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), *confírmola* (14)... [計: 2,135].

均衡分析(1)-地域. 対象変数=地域:年代、参照変数=社会.



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

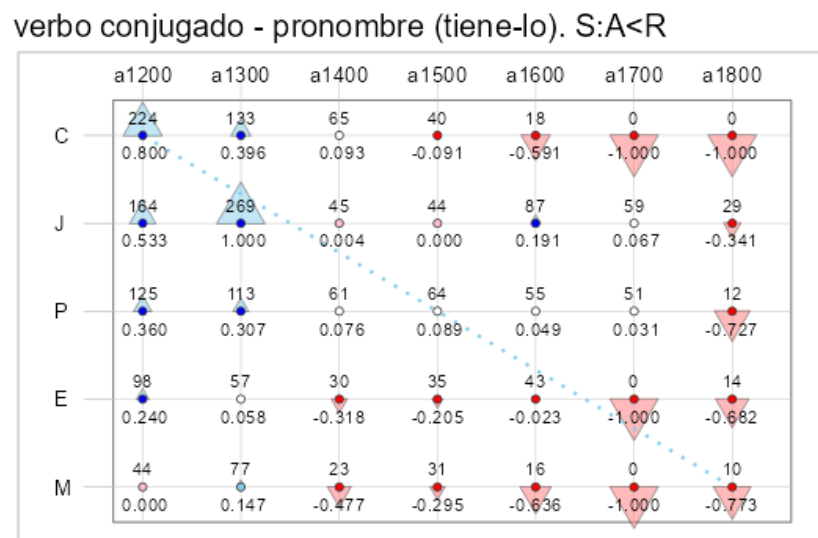
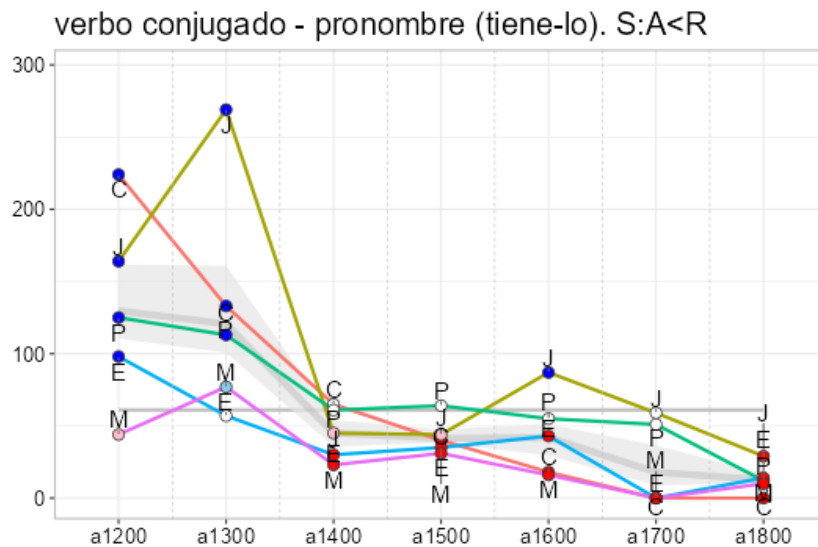


【図-6a,b】 〈活用形-代名詞〉 地域:年代<社会 (平均: 51)

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

後置活用形〈活用形-代名詞〉(*tiene-lo*)は少数ながら、中世(a1200-a1300)では近世(a1500-)に比べて頻度が比較的高かった。しかし、a1400以降減少し、a1800ですべての地方で消失した。〈活用形-代名詞〉はラテン語法に基づく伝統的な語順であるので、それが中世のCV, CN, EXで多く用いられた。a1600, a1700のAN, ARの用法は復古的な文体である可能性がある。

均衡分析(2)-社会. 対象変数=社会:年代、参照変数=地域.



【図-7a,b】 〈活用形-代名詞〉 社会:年代<地域 (平均: 61)

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

後置活用形〈活用形-代名詞〉(*tiene-lo*)は a1200-1300 に C, J に多かったが、a1400 以降は減少し、a1800 にすべての文書で終息した。このことは伝統的文体の特徴を示している。



## 考察

### (1) 〈活用形-代名詞〉に先行する語

先行研究は中世において文頭で〈活用形-代名詞〉の語順が見られたことを指摘している。たしかに、次の表によれば、対称値(r)は、先行条件(x)の文頭(#)で 0 に近く、これは〈活用形-代名詞〉(*tiene-lo*)と〈代名詞+活用形〉(*lo-tiene*)を同等に使用していることを示す。一方、文中(=)では〈活用形-代名詞〉(*tiene-lo*)がほとんど使用されていない。このように、文頭の位置(#)は〈活用形-代名詞〉(*tiene-lo*)を強く条件づけてはいないが、それでも文中の位置(=)と比べれば、より多く〈活用形-代名詞〉(*tiene-lo*)を選択したことがわかる。

| 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 |

【表-5】 文頭(#)・文中(=)と〈活用形-代名詞〉の関係

先行する語の品詞については、次の表のとおり、とくに〈活用形-代名詞〉を選択した品詞はない。

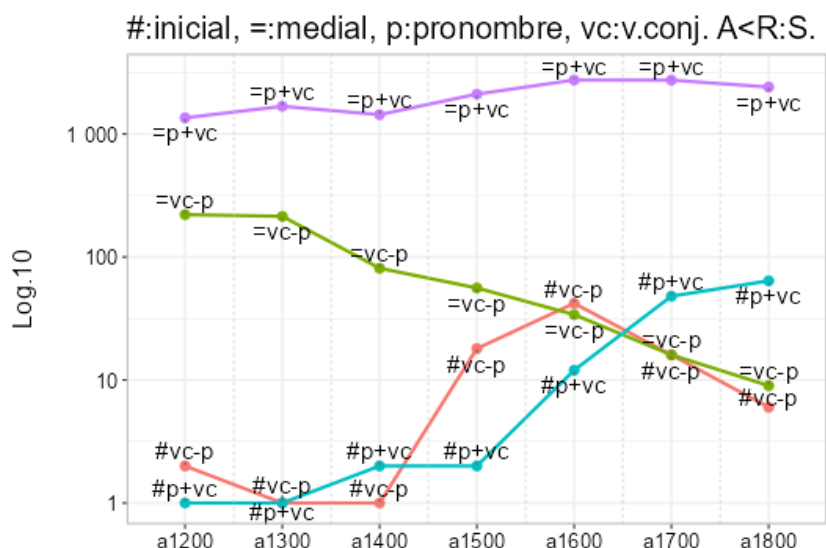
|        | s. x    | s. y  | x     | y   | a     | b     | c     | d      | s      | r |
|--------|---------|-------|-------|-----|-------|-------|-------|--------|--------|---|
| pst    | v_cj_cl | 22    | 1,136 | 4   | 18    | 1,132 | 9,801 | 10,955 | -0.636 |   |
| a      | v_cj_cl | 162   | 1,136 | 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 |   |

【表-6】 先行する語の品詞と〈活用形-代名詞〉の関係

よって、中世スペイン語の特徴である〈活用形-代名詞〉(*tiene-lo*)は、とくに文頭(開始および句読点の後)に出現した、と言える。この位置では弱勢

代名詞は支えとなる先行する強勢語がないために、動詞活用形(強勢)の位置に留まることとなったようだ。しかし例外も多かった。

次は、〈活用形-代名詞〉(*tiene-lo*)が文頭で現れる場合(#)と文中で現れる場合(=)を比較した年代推移を示すグラフである。



【図-8】 文頭(#)と文中(=)の〈代名詞+活用形〉 / 〈活用形-代名詞〉

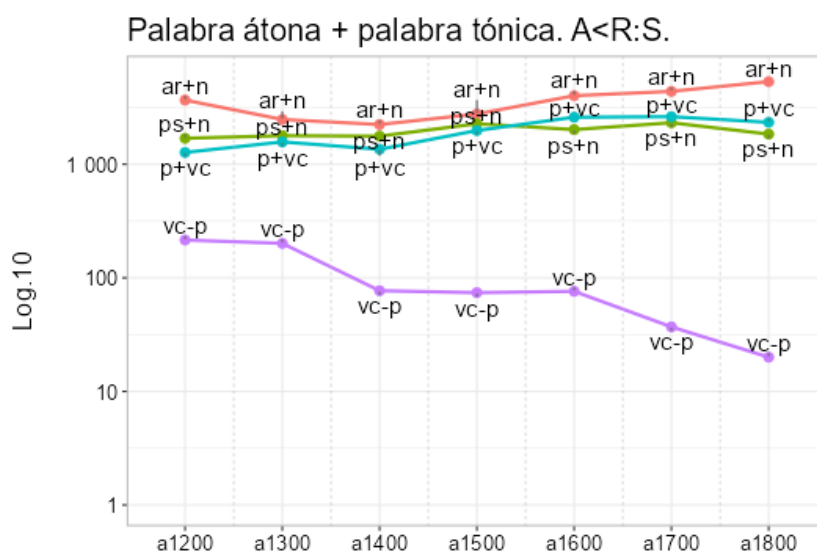
このように、文中(=)の〈代名詞+活用形〉(=p+vc: *lo+tiene*)は常に高頻度であった。文頭(#)の〈代名詞+活用形〉(#p+vc: *lo+tiene*)が中世ではほとんど見られなかったが、近世になって上昇した。文中(=)の〈活用形-代名詞〉(=vc-p: *tiene-lo*)はでは単調に減少し、a1800 にほぼ消滅しているが、文頭(#vc-p: *tiene-lo*)は、a1500-1700 で盛り返し、とくに a1600 に顕著に上昇している。これもやがて a1800 にほとんど終息する。

たとえば、Benito Pérez Galdós (1843-1920), *Marianela* (1878)の文頭の位置を手作業で集計すると、地の文で〈活用形-代名詞〉(*tiene-lo*)の個数が48、〈代名詞+活用形〉の個数が7であった。一方、会話文の中では、逆転して〈活用形-代名詞〉(*tiene-lo*)の個数が9、〈代名詞+活用形〉(*lo+tiene*)の個数が65であった(上田 2024)。このことから、〈活用形-代名詞〉が文語体、〈代名詞+活用形〉が口語体を示しているように思われる。これは中世からの伝統的特徴が文学作品の中で19世紀末まで持続していることを意味する。

## (2) 〈活用形-代名詞〉が終息した理由

〈活用形-代名詞〉 (*tiéne-lo*)の減少した理由は、競合(対立)する前置活用形 (*lo+tiéne*)の圧倒的多数を占める無標項であったため、逆に後置活用形 (*tiéne-lo*)が有標化したことが考えられる。有標項から無標項への変化は自然だからである。

ここで〈代名詞+活用形〉 (*lo+tiéne*)と同様に〈弱勢語+強勢語〉という構造になる〈定冠詞・所有詞+名詞〉 (*la / mi casa*)などは常に高頻度であったことを再度確認する。下の図は該当する名詞句と動詞句を比較したものである(対数軸, ar+n: artículo definido + nombre, ps+n: posesivo + nombre, p+vc: pronombre + verbo conjugado, vc-p: verbo conjugado-pronombre)。



【図-9】 〈弱勢語+強勢語〉と〈活用形-代名詞〉の年代的推移。

上の図が示すように、全年代を通して〈定冠詞+名詞〉 (ar+n: *la casa*)と〈所有詞+名詞〉 (ps+n: *mi casa*)が類似した高頻度で使用された。このことは〈代名詞+活用形〉 (p+vc: *lo+tiéne*)も同様である。一方、逆語順の〈活用形-代名詞〉 (vc-p: *tiéne-lo*)の推移は大きく異なり、全体的に頻度が低く、そのうえ、年代を追って顕著に減少した。しかし、完全には消失せず a1800 まで残存した。

次は中世(a1200-1400)における〈定冠詞+名詞〉, 〈所有詞+名詞〉, 〈代名詞+活用形〉の高頻度語形のリストである。

〈定冠詞+名詞〉: el rey (502), la iglesia (451), la gracia (388), la moneda (262), la villa (256), la carta (254), la ciudat (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)... [計: 18,139].

〈所有詞+名詞〉: 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)... [計: 12,270].

〈代名詞+活用形〉: se contiene (207), la confirma (148), la fiz (139), vos damos (123), me oblige (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)... [計: 9,819].

これらの語形には〈弱勢語+強勢語〉という強勢パターン、〈短い語+長い語〉という単語の長さのパターン、〈範列変化+屈折変化〉という形態のパターンが共通して見られる(→3.1.)。一方、本節で扱っている〈活用形+代名詞〉(*tiene-lo*)は次のようにまったく異なるパターンを示す。

〈活用形+代名詞〉: 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), tovímoslo (10), véndovoslo (10), otorgámosvos (9)... [計: 1,136].

このように、〈代名詞+活用形〉(*lo+tiene*)は〈定冠詞・所有詞+名詞〉と同調して圧倒的多数を示した。一方〈活用形+代名詞〉(*tiene-lo*)は〈定冠詞・所有詞+名詞〉と大きく相違する構造を有していたために比較的少数であり、さらに次第に減少して有標化した、と考えられる。

一方、〈活用形-代名詞〉(*tiene-lo*)は、中世では〈代名詞+活用形〉(*lo+tiene*)よりも少数ではあったが、比較的安定して使用された。この理由は、活用形と代名詞が結合して単一語となったことが考えられる。単一語であれば、〈弱勢語+強勢語〉という2語からなる強勢パターンに影響を受けずに、独自に残存することができたからである。ただし、圧倒的多数の〈代名詞+活用形〉(*lo+tiene*)の影響を免れることは困難であったため、徐々に衰退した。

Penny (2006: 163)は、*metiól en el mayor az, pagós mio Cid*のように、強勢語の後に付加された代名詞の語末の/e/が脱落したのは先行する語(主に動詞)と「音声的1単位」("una sola unidad fonética")を形成したためである、と

述べている。このことも *metiól* や *pagós* などが単一語であったことを示している<sup>13</sup>。

しかし、〈活用形-代名詞〉(*tiene-lo*)の単一語化は〈非活用形-代名詞〉(*tener-lo*)の単一語化と比べて、その完成度は低かった。その理由は活用形の部分が屈折を含むことであったと思われる。この屈折は〈活用形-代名詞〉(*tiene-lo*)において完全に機能し、人称・法・時制に応じて変化する。さらに、*tiene-lo* のように、語末から数えて3番目の音節の強勢母音に二重音化(e>ie)が生じている場合もある。これは本来の単一語にはないことである。一方、〈非活用形-代名詞〉(*tener-lo*)は非活用形は屈折しないので、固定した形態となり、たとえば、*parte, perla, puerta, colmo, Carlos* などと同じパターンとなり、ほとんど完全に単一語となった。この不完全な単一語化が〈活用形-代名詞〉の減少・衰退を引き起こした原因であると考えられる。単一語化については、次のセクションで詳述する。

### (3) ラテン語の人称代名詞

スペイン語の歴史を扱う本研究の範囲外であるが、ここで〈活用形-代名詞〉(*tiene-lo*)の起源について先行研究を参照しながら考察したい。

ラテン語の人称代名詞 *ego, mei:, mihi, me:, tu:, tui:, tibi, te:, no:s, nostri:, nostrum, no:bi:s, vo:s, vestri:, vestrum, vo:bi:s* と指示代名詞 *is, ea, id, e:ius, ei:, eum, eam, id, eo: ea:, ei:, eae, ea, eo:rum, ea:rum, ei:s, eo:s, ea:s, ea* はすべてインドヨーロッパ祖語の強勢語に由来する(Palmer 1974: 257-259)。ラテン語では、以下のように、強勢語となるときの弱勢語になるときがあった。

古典ラテン語の教科書(樋口・藤井 1963: 44)の練習問題に次の例文がある。

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

"The lady withdrew into the room with her children and slaves."

#### (2) *cur vos sic languidos video, pueri?* "Why do I see you so weary, boys?"

#### (3) *fuitne Marcus tibi amicus?* "Was Marcus a friend to you?"

(1)の *se recepit* は再帰動詞で、この再帰代名詞 *se* は動詞 *recepit* の直前にある。この再帰代名詞の位置は現代スペイン語と同じである。(2)の対格人稱代名詞 *vos* は動詞 *video* の前にあるが、離れた位置にある。(3)の与格人稱代名詞 *tibi* は動詞の後にあり、動詞と離れている。このように人稱代名詞が

<sup>13</sup> Penny (2006: 159)は弱勢代名詞は強勢語の前にあるときも、後にあるときも強勢語と単一語を形成する、と述べているが、私たちは強勢語の前にあるときは2語となり、強勢語の後にあるときは単一語となる、と考える。

動詞から自由な位置に置かれたのは、人称代名詞に強勢があったためである、と思われる。

この代名詞の自由な位置の 1 つとして文頭があった。Luraghi (2016)によれば、カエサル(BC 100-44)『ガリア戦記』の次のような文頭の代名詞には強勢があった。その英訳を見るとたしかに指示代名詞になっている。

(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).

一方、時代は下って『新約聖書』「マルコスによる福音」(Vulgata 382?)では下の例文(5), (6)のような代名詞(3 人称)の *enclisis* が非常に多くなる(Luraghi, *ibid.*)。例文(7)はキケロ(BC. 106-43)の『アッティクス宛書簡集』であるが(Luraghi, *ibid.*)、このような代名詞後置は古典ラテン語では少なかった(Pinkster 1995: 221)。この動詞の直後にある代名詞は弱勢語であった、と考えられる<sup>14</sup>。

(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).

上の〈動詞-代名詞〉(*enclisis*)の動詞部には活用形(*conspuent*, *flagellabunt*, *interficiunt*)と非活用形(不定詞: *increpare*, *subire*)がある。

Luraghi (*ibid.*)は 3 人称代名詞だけを取り上げているが、Yamamoto (1997)は『新約聖書』「マルコスによる福音」のラテン語訳、ギリシャ語訳、古英語訳における全人称の代名詞の位置を調べている。その集計の中からラテン語の部分合計すると、動詞の後にある代名詞は 247 例、動詞の前にある代名詞は 52 例となるので、やはり動詞の後にある代名詞のほうが圧倒的に多い。その例文を見ると、次のように代名詞は動詞の直後に置かれている。

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

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

---

<sup>14</sup> Bassols de Climent (1992: 120)は「動詞の補語として用いられた人称代名詞は強勢の場合と弱勢の場合があった」と述べている。

上の(8)では3人称単数形に *is, ei*:「彼」ではなく、*ille, illi*:「あれ」が使われている。古典ラテン語ではふつう *is, ei*:が用いられ、*ille, illi*:は強調形であった(Lausberg 1973: 164)。一方、上の(8)を見ると、聖書ラテン語では一般に「彼」の意味で使用されていたようだ。やがて、この *illi*:「彼に」がスペイン語の *le* となった。Menéndez Pidal (1968: 253)は、*illi*:の語頭の母音が脱落したのは *cantarunt-(i)llu, non-(i)llu* のような接合形が使われたためであった、と述べている。これらはラテン語形であるので、とくにラテン語で動詞の代名詞後置(*cantarunt-(i)llu*)が多かったことが影響した、と考えられる。

Luraghi (ibid.)と Yamamoto (ibidd.)の研究によって、古典ラテン語では代名詞の位置は自由であったが、俗ラテン語では動詞に後置するようになったことが想定できる。そこで、スペイン語史の研究において、とくに重要な俗ラテン語の状況を見るために、俗ラテン語に訳された『旧約・新約聖書』(*Vulgata*, 4世紀末)全体を高頻度動詞 *dicere* 'decir'に限って調べた結果(計6,181語)、(a)後置する代名詞は938例であった。他に、(b)動詞の後には ad+名詞 462例、(c)名詞(与格) 77例、(d)固有名詞人名(与格) 43例が見つかった。一方、(e)動詞に前置する代名詞はわずかに39例だけであった。よって、*Vulgata* の *dicere* だけの例ではあるが、俗ラテン語では与格目的語と ad+名詞は動詞の後に置かれることが普通であった可能性がある。そのことが影響して与格代名詞が動詞に後置したことが考えられる。

以下は、その代表的な例である(Weber und Gryson 1969, 現代語訳は 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).

この代名詞後置の高頻度の用法がロマンス諸語に継承され(Lausberg 1973: 171-208)、13世紀(a1200)の初期スペイン語(図-15)にラテン語形の残存として伝えられたようだ(Penny 2006: 161)。よって、文献が豊富にある13世紀(a1200)の初期スペイン語よりも以前では、enclisisがさらに多く見つかかり、俗ラテン語から共通ロマンス諸語への移行期には非常に多数のenclisisがあった、と想定できる。それが次第に減少し、13世紀(a1200)の初期スペイン語では、本研究で確認されたように、〈代名詞+動詞〉(proclisis)が圧倒的多数を占めていた。

### 3.3. 〈非活用形-代名詞〉 (*tener-lo*)

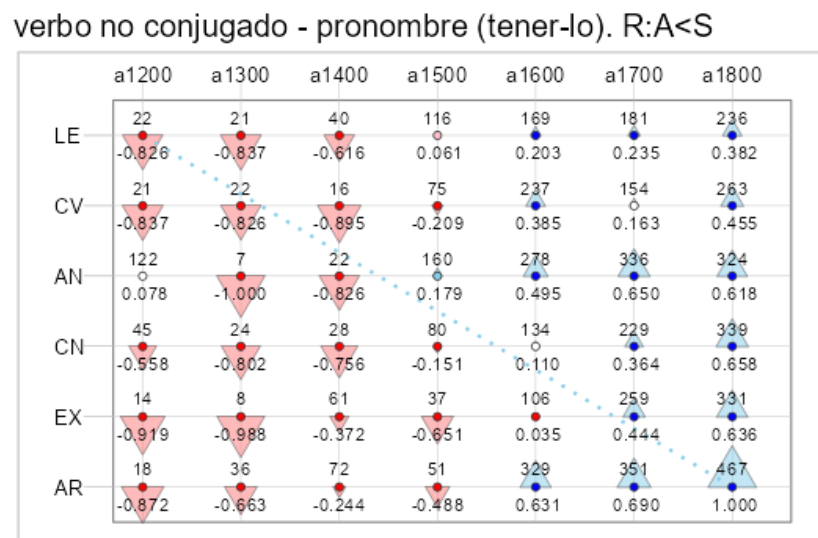
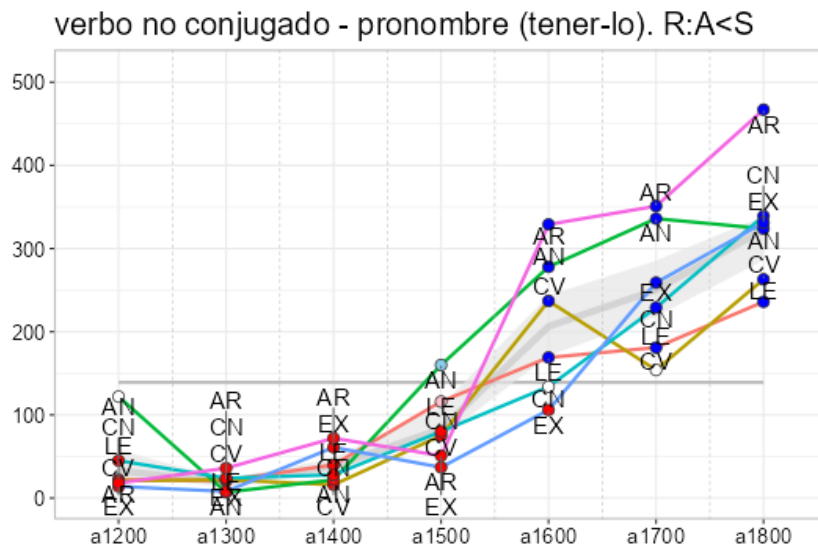
#### 観察

現代語では動詞の非活用形(不定詞・現在分詞)と代名詞の位置的關係は逆転して〈非活用形-代名詞〉(*tener-lo, teniéndolo*)となる。実は歴史的に見ると、この語順は中世ではほとんど見られず、近代に確立したものである。この節では、この歴史的経緯を観察し、その後で〈代名詞+活用形〉(*lo-tiene*)の語順の場合と歴史的頻度分布が異なる理由を探る。

**該当する形式**(頻度順) : 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)... [計: 5,842].



均衡分析(1)-地域. 対象変数=地域:年代、参照変数=社会.

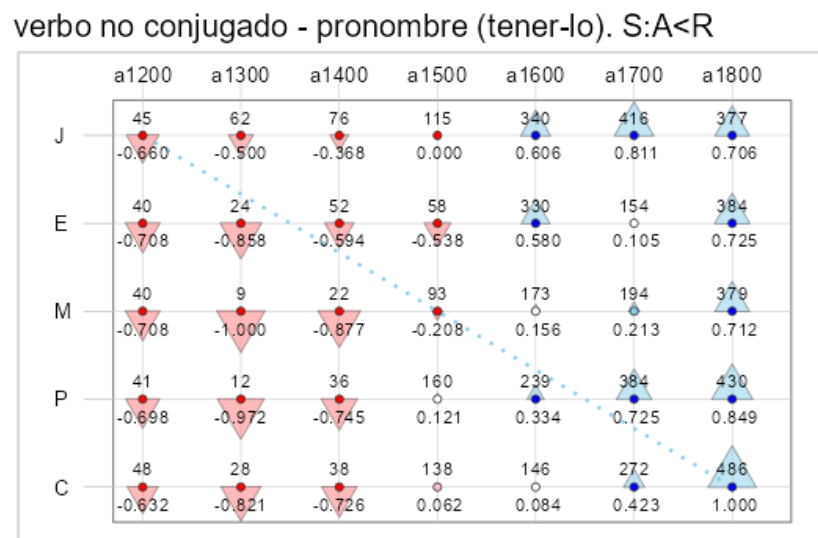
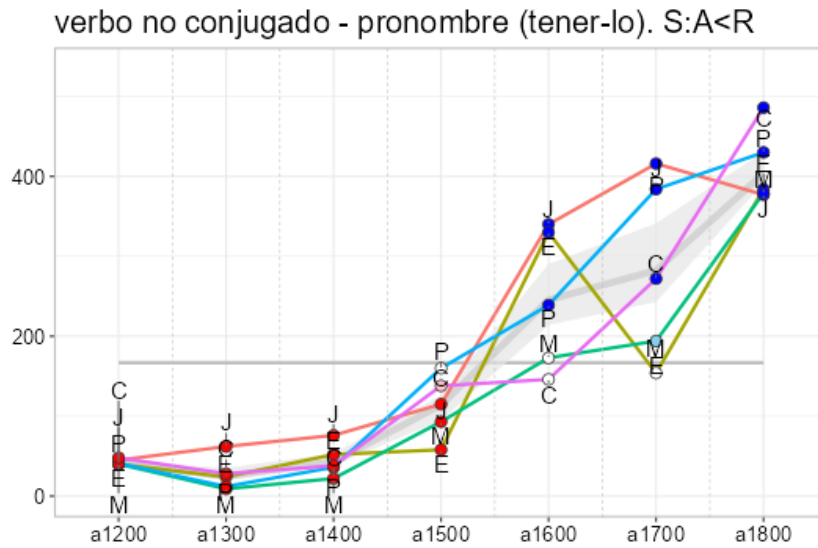


【図-10a,b】 〈非活用形-代名詞〉 地域:年代<社会.変化 (平均: 139)

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

全体的に、〈非活用形-代名詞〉(tener-lo)は a1200-a1500 であまり使用されなかったが、a1600 に有意に増加した。

均衡分析(2)-社会. 対象変数=社会:年代、参照変数=地域.



【図-11a,b】 〈非活用形-代名詞〉社会:年代<地域 (平均: 167)

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

全体的に、〈非活用形-代名詞〉(*tener-lo*)は a1200-1500 であまり使用されなかったが、a1600 に有意に増加した。

## 考察

### 〈非活用形-代名詞〉の単一性

〈非活用形-代名詞〉(*tener-lo*)の形式は中世では少数であったが、近世以降に急上昇した。この〈非活用形-代名詞〉(vn-p: *tener-lo*)の増加の動きは、

先に見た〈代名詞+活用形〉(*lo+tiene*)の安定した多数性とは性質が異なる。その理由は〈非活用形-代名詞〉(*tener-lo*)の代名詞(弱勢)が後置され、非活用形(*tener-lo*)の動詞は変化しないので、〈冠詞・所有詞+名詞〉構造(*la casa, mi casa*)と形態・構造的にまったく異なっていたことが挙げられる。

ここで、〈非活用形-代名詞〉(*tener-lo*)の〈強勢語+弱勢語〉という例外的統語構造は、これを〈強勢語-弱勢辞〉として単一語と見なせば統語的に例外でなくなる。以下では、〈非活用形-代名詞〉(*tener-lo*)が単一語であることを、言語体系(音韻論、形態論、統語論)と言語使用(非規範的表記)の面から検討する。

はじめに音韻の面を見る。〈非活用形-代名詞〉は単一の強勢を語末に近い位置に配置している: *tenerlo, teniéndolo*。このことはすべての強勢語(単一語)と同じである。

そして、先に述べた「最終位置の強勢」(〈弱勢語+強勢語〉)というスペイン語の原則(→3.1)は、〈非活用形〉+〈代名詞〉という2語の連続を嫌う。この代名詞が弱勢であるためである。この理由によって〈非活用形-代名詞〉が単一語となったと考えられる。

次のリストが示すように、〈非活用形-代名詞〉の音形の融合(*tener-lo > tenello*)が頻繁に起こったことも〈非活用形-代名詞〉が単一語であったことを示している。

非活用形と代名詞の融合: *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), *recebillos* (2), ..., [計: 200].

形態論の面では、〈非活用形-代名詞〉(*tener-lo*)の最初の要素(*tener*)が変化せず、末尾の要素(*lo*)が範列的变化を示していることを確認する: *tenerme, tenerte, tenerlo, etc.*。このような〈不変化部+変化部〉という形態は一般の名詞、形容詞、動詞(どれも単一語)と同じである。

次に統語面を見る。本研究では取り上げなかったことであるが、中世ではしばしば〈代名詞+活用形〉の間に *non* や主語代名詞などが〈内挿〉(*interpolación*)されることがあった(*nos non vala, les yo fago, vos yo vendo*) (Menéndez Pidal 1908, 1976: 409-411, Castillo Lluch 1996: 293-302)。次は本研究が使用している資料中の〈代名詞+内挿+動詞〉のリストである:

〈代名詞+内挿+動詞〉 *nos non vala* (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), [計: 930].

しかし、この内挿は〈動詞-代名詞〉の間には見られない(\**de tenernonlo*, \**para fazeryovos*)。このことも〈動詞-代名詞〉が単一語であったことに起因する。

〈代名詞+活用形〉(*lo+tiene*)が統語的に 2 語であり、〈非活用形-代名詞〉(*vn+p: tener-lo*)が統語的に 1 語(単一語)であることは次のような文法現象でも確認される。

RAE y AALE (2009: 1208)は次の例文を挙げて、*proclisis* と *enclisis* の等位構造の違いを説明している: *Lo leyó y resumió* / \**para leer y resumirlo*。

福畠(2021: 141)は、

2 つ以上の目的語が等位構造を成すとき、動詞に前置される場合は、最初の代名詞以外は省略できる。動詞に後置される場合は、どの代名詞も省略できない。

と述べ、次の命令形の例を加えて、(1)の 2 番目の *lo* は省略できるが、(2)のどちらの *lo* も省略できない、と説明している。*proclisis* と *enclisis* の等位構造の違いを示すためには、RAE y AALE (ibid.)よりも福畠(同)のように、後続する代名詞の省略可能性を見るべきである。

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

(2) Léelo y resúmelo cuanto antes.

この現象は動詞が不定詞(3)と現在分詞(4)のときも同様に起こる(以下は筆者の作例)。

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

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

これは〈非活用形+代名詞〉が動詞と結合して単一語を形成するので、その中の要素 *lo* を他の語と共有できないことを示している。このことは、たとえば、次のような例文が不可であることと同様である(筆者の作例)。

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

最後に、非規範的表記の中で言語使用の面を観察する。中世の文学・歴史書・司法文献における弱勢代名詞の位置を広範囲に研究した Castillo Lluch (1999)の豊富な収集例の中に多くの〈非活用形-代名詞〉(*tener-lo*)の結合形が見られる<sup>15</sup> : *pedirle, colgarla, ganarlás, ponerlas, etc.; otorgandolis, dandovos, pidiendol, etc.* 一方、〈代名詞+活用形〉(*lo+tiene*)のほとんどすべての代名詞は不定詞から分離されている<sup>16</sup> : *las fazer, gelos dar, se quejar, se mantener, delos menazar, nol amar, se defendiendo, nol conociendo, les dando, etc.* この違いは現代のような規範的正字法が成立していなかった中世において、書き手が〈非活用形-代名詞〉(*tener-lo*)を単一語と見なし、〈代名詞+活用形〉(*lo+tener*)を2語と見なしていたことを示している、と考えられる。

現代の王立アカデミーの正字法(RAE y AALE 2010: 530)は次のように述べている([...]と下線は筆者)。

[...] 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.*

ここで言及されている"una sola palabra gráfica"は表記だけの問題ではなく、言語体系の中でも"una sola palabra"である、と考えられる<sup>17</sup>。

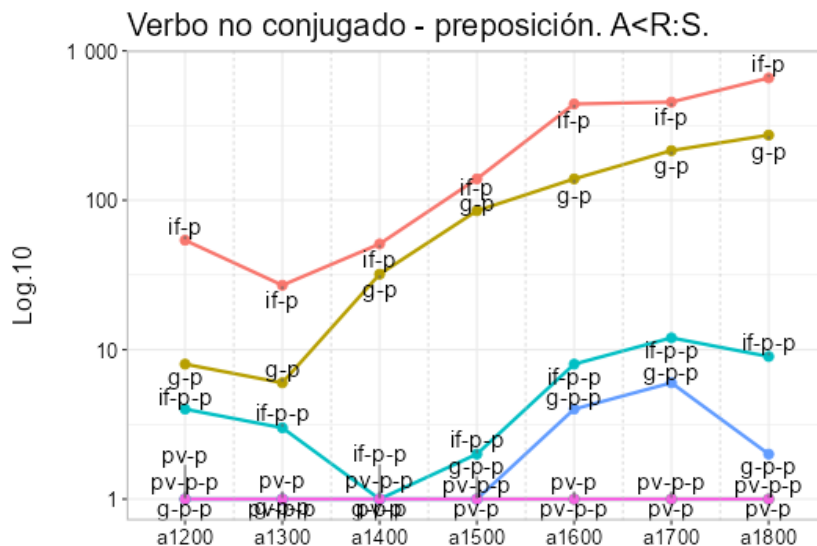
次のグラフは〈非活用形-代名詞〉の歴史的変化を示している(if: infinitivo, g: gerundio, ip: imperativo. pv: participio verbal)。

---

<sup>15</sup> 分離形も多く見つかる : *sacar les, uestir les, llamando la.*

<sup>16</sup> 唯一の例外は *L'auia dito Aristotus (Liber Regum, 9, 29)*である (Castillo Lluch. *ibid.*)。これ以外の〈代名詞+動詞(活用形・非活用形)〉の融合形・連続形は見つからない。

<sup>17</sup> Castillo Lluch. *ibid.*: 357)は次のように述べている: "Los casos de adjunción a la palabra siguiente como *uosdi* [...] son extremadamente excepcionales si los comparamos cuantitativamente con ejemplos como *quelo enterrase* [...] o del tipo *quel diera*, [...], que por otra parte, denotan no solo un ligamiento gráfico, sino morfológico." ここで *ligamiento morfológico* は代名詞が動詞に先行し、さらにその前の語に結合した場合であるが、代名詞が動詞の後に接合した場合も同様である。



Verbo no conjugado - preposición. A<R:S.

|        | a1200       | a1300       | a1400       | a1500        | a1600        | a1700        | a1800        |
|--------|-------------|-------------|-------------|--------------|--------------|--------------|--------------|
| if-p   | 54<br>0.407 | 27<br>0.185 | 51<br>0.333 | 139<br>0.630 | 441<br>0.852 | 454<br>0.926 | 658<br>1.000 |
| g-p    | 8<br>-0.037 | 6<br>-0.185 | 32<br>0.259 | 95<br>0.481  | 139<br>0.630 | 215<br>0.704 | 273<br>0.778 |
| if-p-p | 4<br>-0.333 | 3<br>-0.481 | 1<br>-0.704 | 2<br>-0.556  | 8<br>-0.037  | 12<br>0.111  | 9<br>0.037   |
| g-p-p  | 0<br>-1.000 | 0<br>-1.000 | 0<br>-1.000 | 0<br>-1.000  | 4<br>-0.333  | 6<br>-0.185  | 2<br>-0.556  |
| pv-p   | 0<br>-1.000 | 0<br>-1.000 | 0<br>-1.000 | 0<br>-1.000  | 1<br>-0.704  | 1<br>-0.704  | 1<br>-0.704  |
| pv-p-p | 0<br>-1.000 | 0<br>-1.000 | 0<br>-1.000 | 0<br>-1.000  | 0<br>-1.000  | 0<br>-1.000  | 0<br>-1.000  |

【図-12】 〈非活用形-代名詞〉の年代的推移

上のグラフの中で、とくに〈不定詞+代名詞〉(if-p)の増加が著しい。それに比べて、〈現在分詞-代名詞〉(g-p)の上昇はあまり大きくはない。この理由として〈不定詞+代名詞〉の強勢パターンは語末から数えて2番目の音節に強勢があるので(*comprar-lo*)、強勢の配置は無標の〈...CVC(V[s,n])#〉となる(→3.1.)のに対して、〈現在分詞-代名詞〉は語末から数えて3番目の音節に強勢があり(*teniendo-lo*)、有標の強勢パターンを示すことが考えられる。(上田 2024)。

また、次の表が示すように、不定詞の全頻度が現在分詞の全頻度よりも格段に多いことも考慮しなければならない。

| *       | ~      | ~.p   | ~p.p. | 計      |
|---------|--------|-------|-------|--------|
| 不定詞(if) | 33,363 | 3,900 | 167   | 37,430 |
| 現在分詞(g) | 6,100  | 1,599 | 182   | 7,881  |
| 計       | 39,463 | 5,499 | 349   | 45,311 |

【表-7a,b】 不定詞(if)と現在分詞(g)の頻度 (p:代名詞)

このように〈非活用形-代名詞〉(*tener-lo*)は、〈代名詞+活用形〉(*lo+tiene*)の分析において想定した〈弱勢語+強勢語〉による類推作用、その中のとくに〈定冠詞・所有詞+名詞〉による類推作用が形態的・構造的な相違から有効に働かず、独自の発展を遂げたようだ。その発展に寄与したのは〈代名詞+活用形〉の単一語化であった、と思われる。とくに多かった〈不定詞-代名詞〉の使用は、その強勢パタンの無標性(語末から数えて 2 番目の音節に強勢があること)と、不定詞そのものの使用頻度が高かったことが理由として考えられる。

さらに少数であるのは、代名詞が 2 つ結合した形式である(if-p-p, g-p-p, pv-p-p, ip-p-p)。これらも強勢の配置は無標の〈...'CVC(V[s,n])#〉に合致しない。よって、単一語化は不完全であったと考えられる。

次は、それぞれのグループに属する語形とその頻度のリストである。

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

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

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

if-p-p: avérselo (6), dársela (5), dárselo (5), haversele (5), avérsese (3), concertélo (3), fazérvoslo (3), avérsese (2), contrallárgela (2), dársese (2)... [計: 171].

g-p-p: aviéndosele (11), aviéndosele (3), haviéndosele (3), abiéndosele (2), aperziviéndosele (2), diciéndoselo (2), dándosele (2), dándoselos (2), habiéndosele (2), haviéndoselo (2)... [計: 74].

pv-p-p: dádosele (2), pagádoselos (1) [計: 3].

### 3.4. 〈代名詞+非活用形〉 (*lo+tener*)

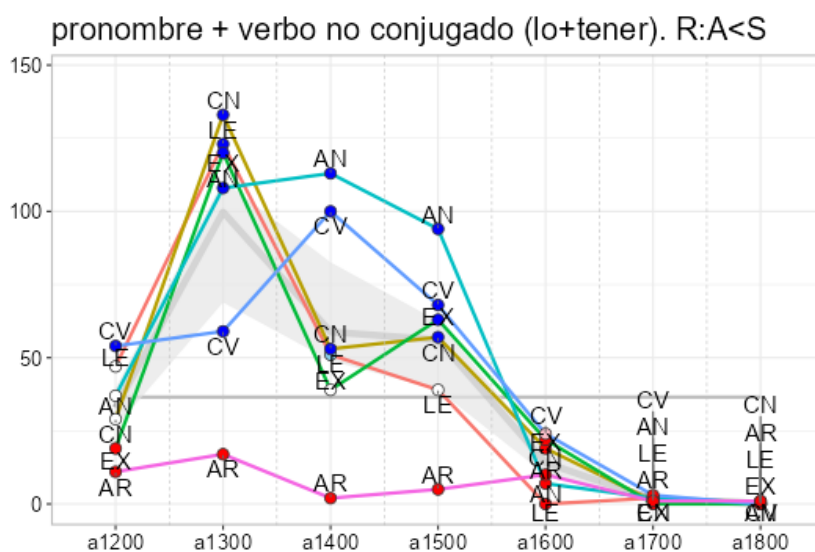
#### 観察

代名詞と動詞の連続の最後の場合は〈代名詞+非活用形〉 (*lo+tener*)である。これは先に見た〈活用形-代名詞〉 (*tiene-lo*)と同様に、ラテン語形の残存であり(→4.2.)、現代語では見られない。両者の語形だけを見ると共通点が見つからないが、どちらもそれぞれが競合する形態、すなわち〈非活用形-代名詞〉 (*tener-lo*)と〈代名詞+活用形〉 (*lo+tiene*)が多数使われているので、少数であった形態が有標化したことが共通している。この有標化が使用頻度の減少の原因である、と考えられる。4つの代名詞と動詞の組み合わせの中で、この〈代名詞+非活用形〉 (*lo+tener*)の語順は全歴史を通して頻度が一番低かった。

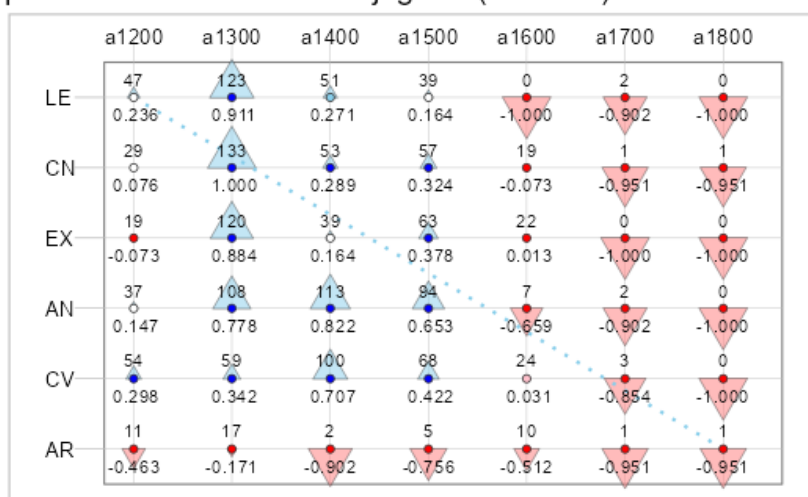
**該当する形式**(頻度順) : *lo fazer* (53), *vos fazer* (53), *les fazer* (44), *vos dar* (37), *les ir* (30), *lo vender* (22), *los tener* (21), *la fazer* (20), *vos defender* (18), *lo dar* (16), *los dar* (16), *lo aver* (15), *nos ayuntar* (15), *gela quebrantar* (13), *les passar* (12), *lo tener* (12), *le dar* (11), *les dar* (11), *lo cumplir* (11), *vos redrar* (11)... [計: 1,536].



均衡分析(1)-地域. 対象変数=地域:年代、参照変数=社会.



pronombre + verbo no conjugado (lo+tener). R:A<S

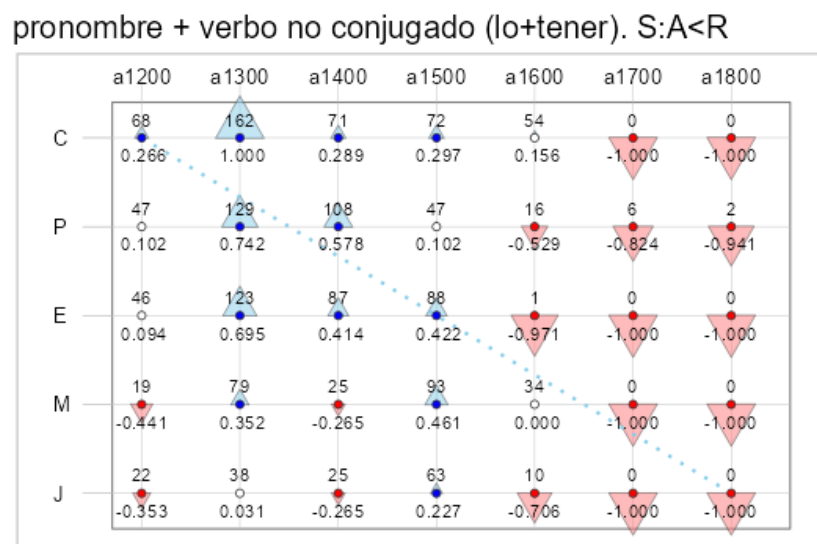
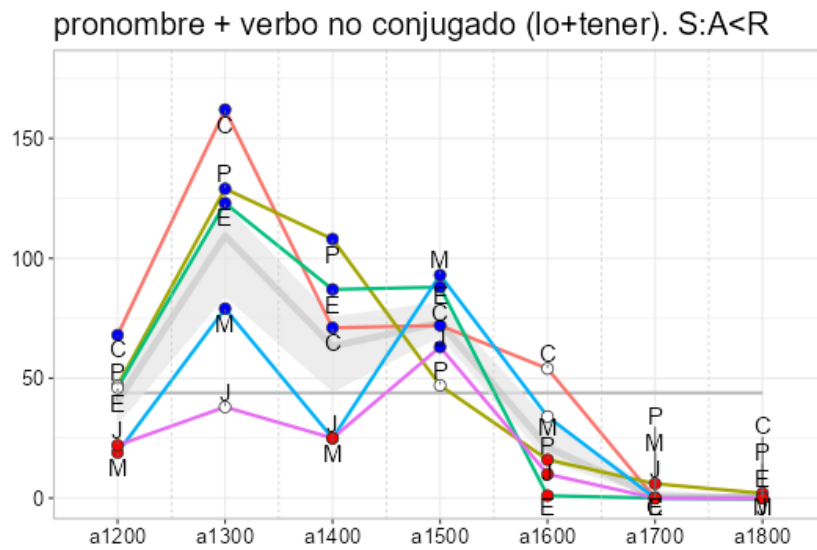


【図-13a,b】 〈代名詞+非活用形〉 地域:年代<社会 (平均: 37)

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

**観察:** 〈代名詞+非活用形〉 (*lo+tener*) は全体的に頻度が低く、さらに次第に減少し、a1600以降は皆無に近い。a1300にAR(最も革新的な地域)を除く全地域で有意に多数を占めた。全体を通してARでは僅少であった。よって、この語形は中世の伝統を守る年代・地域で比較的多く使われ、革新を示す年代・地域に比較的少なかった、と考えられる。

均衡分析(2)-社会. 対象変数=社会:年代、参照変数=地域.



【図-14a,b】 〈代名詞+非活用形〉社会:年代<地域 (平均: 44)

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

a1300にC, E, Pで有意に多数を占めた。全体的にM, Jでは僅少であった。よって、〈代名詞+非活用形〉(lo+tener)は伝統的な文体特徴である、と考えられる。

## 考察

### 〈代名詞+非活用形〉が減少した理由

〈代名詞+非活用形〉(*lo+tener*)は、競合(対立)する〈非活用形-代名詞〉(vn-p: *tener-lo*)が圧倒的に増加して無標項になったため(→3.3)、逆に有標化したことが考えられる。

〈代名詞+非活用形〉(*lo+tener*)は〈冠詞・所有詞+名詞〉(*la / mi casa*)の構造と比べると〈弱勢語+強勢語〉であることが部分的に類似するので、使用が増加することが予想されるかもしれない。しかし、動詞が非活用形(屈折しない)であることが形態的に〈冠詞・所有詞+名詞(屈折)〉(*las / mis casas*)と異なるので〈冠詞・所有詞+名詞〉による類推作用は作用しなかった、と考えられる。

次は〈弱勢語+強勢語〉の品詞の組み合わせのリストである。

〈弱勢語+強勢語〉の品詞の組み合わせ: 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)... [計: 643,601].

このリストのほとんどの 2 番目の要素は名詞(n), 動詞(v), 形容詞(a)である。これらは形態的に屈折する品詞である。一方、p+t, p+pn\_ps, p+np, c+av, c+a\_nm, pt+av の 2 番目の要素は屈折しないので、〈代名詞+非活用形〉(*lo+tener*)と類似するようだが、これらの 1 番目の要素は前置詞(p)と接続詞(c)である。この 2 つの品詞のメンバーには、以下に見るように、代名詞と類似する形態は存在しない。

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

前置詞: *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.*

代名詞: *la, las, le, les, lo, los, me, mi, nos, os, se, se la, se las, se les, se lo, se los, te.*

よって〈弱勢語+強勢語〉の品詞の組み合わせの中で〈代名詞+非活用形〉(*lo+tener*)に類推的に作用するような高頻度の形式は存在しなかったために支援を得られず、〈代名詞+非活用形〉(*lo+tener*)が有標化して消滅した、と考えられる。

### 3.5. 〈命令形-代名詞〉

このセクションでは、次のような〈命令形-代名詞〉(*ten-lo*)を扱う。命令形は人称に応じて活用するが、〈活用形+代名詞〉(*tiene-lo*)と異なり、歴史を通して減少することはなかった<sup>18</sup>。

**該当する形式**(頻度順) : *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)... [計: 303].

語形の出現数が少ないので、地域・年代・社会について頻度分布を観察することはできない。よって、このセクションは考察だけを行う。

#### 考察

##### 〈命令形-代名詞〉が安定して保持された理由

〈命令形-代名詞〉の動詞は人称に応じて活用することは〈活用形+代名詞〉と同じである。この類似性があるにもかかわらず、〈活用形+代名詞〉が終息し、〈命令形-代名詞〉が安定して保持された理由を、以下で探る。

はじめに、ラテン語の命令形の統語的特徴を確認しておくことが必要であろう。古典ラテン語統語論の研究書(Bassols de Climent 1992, Calero Calero 2003)の該当部分を参照すると、命令法動詞が文頭にある文例のほかに、次のように文頭にない多くの文例が載せられている<sup>19</sup>。

de palla memento (Plauto) "Recuerda el manto." [Bassols de Climent 1992:

---

<sup>18</sup> 稀に、次のような〈代名詞+命令形〉が見つかるが、非常に少数なので例外として、ここで扱わない: *la embiad* (3), *lo cumplid* (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) [計: 13].

<sup>19</sup> Cf. 樋口・藤井(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."

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.]

一方、俗ラテン語訳聖書(*Vulgata*, 4世紀末)を扱った専門書が載せている例文のすべての命令法動詞は文頭に置かれている(足立 2014: 244-251, 17例; 田淵 2006: 32例)。ここでは、文中のコンマ(,)と接続詞 *et* の後の位置も「文頭」に含める(現代語訳は Reina y Valera 1960)<sup>20</sup>:

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). [足立 *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). [同] / Benefacite his qui oderunt vos. "Benedicid a los que os maldicen." (S. Mateo: 5, 44) [田淵: 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). [同:48]

*Vulgata* 全体の中で、きわめて高頻度の動詞 *dicere* 'decir'の命令形(*dic* 'di', *dicite* 'decid')<sup>21</sup>を探すと、129例が見つかった。その内訳は、命令形の後の位置の後に現れる、(a)名詞(与格)が28例、(b)代名詞(与格)が21例、(c)固有名詞(与格)が13例、(d)ad+名詞・代名詞・固有名詞(対格)が14例である。これらの要素が命令形の前にある例は皆無に近い<sup>22</sup>。よって、*Vulgata* の *dicere* の命令形に限った資料であるが、命令形は基本的に代名詞を含めてすべての間接目的語を後続させている、と考えられる。次がその例である(Weber und Gryson 1969, 現代語訳は Reina y Valera 1960):

<sup>20</sup> たとえば、Medice cura te ipsum. "Médico, cúrate a ti mismo." (S. Lucas: 4, 23) [田淵: 131]の Medice は呼格なので、コンマ(,)を補えば cura は文頭になる。事実、Domine, doce nos orare "Señor, enséñanos a orar." (S. Lucas: 11, 1) [田淵: 54]ではコンマが付されている。田淵(同)の中の文頭でない命令形は次の例文だけである: Si mihi non vultis credere, operibus credite. "aunque no me creáis a mí, creed a las obras." (S. Juan: 10, 38).

<sup>21</sup> 古典ラテン語の命令形未来 *dicito*, *dicitote* は見つからない。

<sup>22</sup> 次が唯一の例外である: et Amasae dicite nonne os meum es et caro mea "Asimismo diréis a Amasa: ¿No eres tú también hueso mío y carne mía? [...]" (1° de Samuel: 19, 13).

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

これらの命令形は文頭(a-1, a-2, b-1, c-1: 全体で 20 例)、または接続詞 *et* の後で現れることが多い(b-2, c-2: 31 例)<sup>23</sup>。それ以外の場合でも、たとえば d-2 のように、命令形の主語が前置しているときなどは、その後に小さな休止があることが想定される。事実、現代語訳にコンマ(,)が使用されている。このように、*Vulgata* 中の *dicere* の命令形は基本的に文頭または文頭に準じた位置に現れている、と言える。一方、先に見たように、古典ラテン語では命令形の前に目的語やほかの要素がある例が多く見られる。よって、古典ラテン語の動詞命令形の位置は自由であったが、俗ラテン語の『聖書』ではほとんどが文頭に限られていたようである。このように、俗ラテン語において命令形が文頭に現れることは、民衆の言語の「下からの変化」(change from below)を示すようである。民衆が依頼したり、命令したりするとき、結びの動詞を先に表現し、その目的語や副詞などは後で補うようにしたことが想像できる。

ここでスペイン語の公証文書から〈命令形-代名詞〉だけを取り出して、その代名詞の位置と頻度分布、そして歴史的推移を観察し、命令形に代名詞が唯一的に後置した歴史的理由を探ることにする。

先に見たように、俗ラテン語の命令形は基本的に文頭に現れていたようである。そこで、スペイン語の公証文書の〈命令形-代名詞〉に先行する語

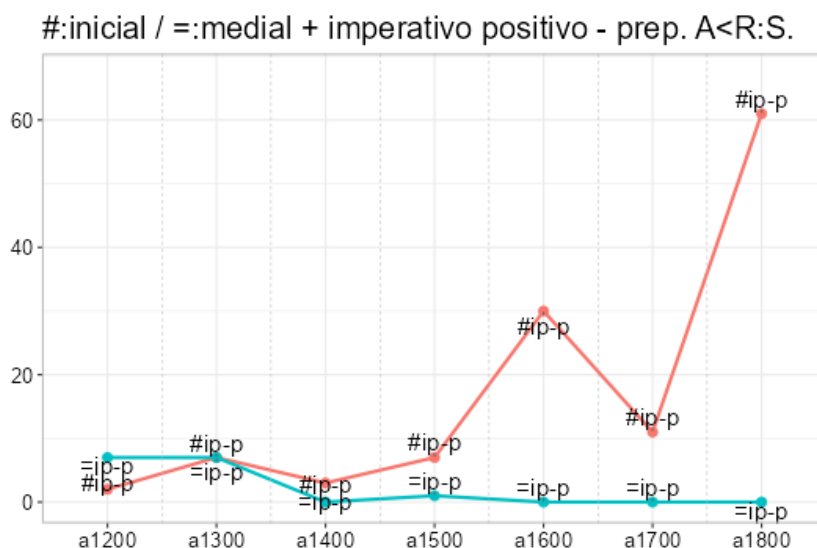
<sup>23</sup> *ac = atque* が使用されていることもある。

の品詞を見ると、その大多数が句読点(pt)と接続詞(c)である<sup>24</sup>。そして、接続詞の内訳を見ると、すべて等位接続詞であり、その大多数は *y*, *e* である:

先行語の品詞 : 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) [計: 303] // 先行句読点の語形: [.] (124), [,] (43), ] (4), “ (4), ” (2), ( (1), - (1) [計: 179] // 先行接続詞の語形: y (43), e (20), mas (1), o (1) [計: 65].

文頭の命令形の前等の位接続詞は単に付加された要素であり、必須の意味のない弱勢語である。よって、句読点の後と等位接続詞の後をともに「文頭」(#)として扱うと、命令形は主として文頭で用いられた、と考えられる。このように、公証文書全体に現れる命令形の位置が文頭であることは、俗ラテン語と同じである。

次は「文頭」(#: 句読点の後と等位接続詞の後)とそれ以外の位置(=:文中)にある〈命令形-代名詞〉の歴史的推移を示すグラフである。



【図-15】 〈命令形-代名詞〉の年代的推移

このグラフを見ると、〈命令形-代名詞〉は一般の(活用形-代名詞)と同様に、中世(a1200-11400)において文頭(#)と文中(=)で使用されていたように見えるが、出現数がわずかなので断定はできない。そこで、少ない事例なので、すべてを検索してその前文脈を調べると、それらは、次のように、

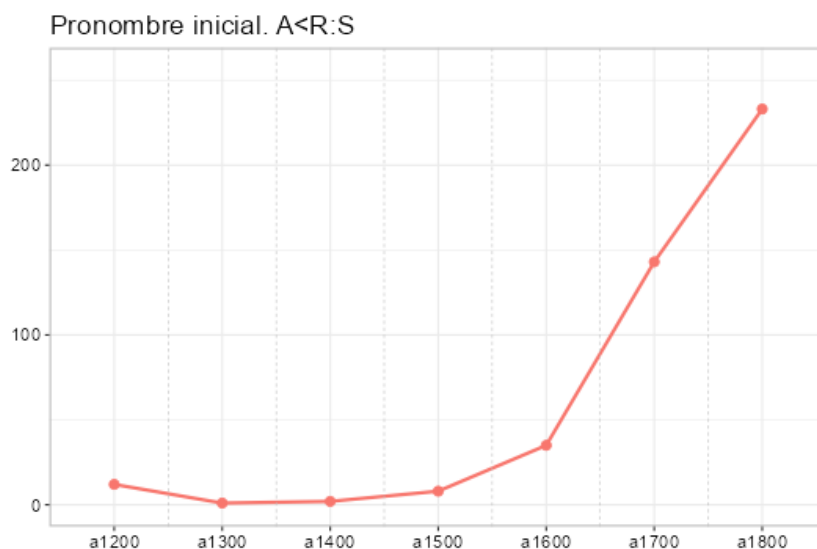
<sup>24</sup> 先行語の品詞として接続詞に続いて多いのは過去分詞の形容詞(pa: participo adjectival)である。その大多数は *La carta leída dádgela* という過去分詞構文であるので、*leída* の後に休止を想定すれば文頭として扱うことも可能である。

分詞構文、従属文、主題のいずれかであった:

分詞構文: e nós vista la carta diémosgelo (id-45, a1200) / La carta leída dádgela (id-76, a1200) // 従属文: E si assí non lo quisieren fazer tomadles (id-78, a1200) / ante que metades foz en las miesses fazétnoslo (id-150, a1200) // 主題: E lo que fiziéredes en esto embiátnoslo (id-136, a1300) / E los dichos maravedís dátgelos (id-537, a1400).

これらの前文脈の後で短い休止が想定されるので、その直後の命令形は準文頭の位置にあった、と言える。

次のグラフは語頭(休止の後)に現れた弱勢代名詞の出現頻度の年代推移を示している。このグラフによると、a1200-a1600の間、ほとんど代名詞は語頭で出現していない。このために、語頭で現れた命令形は代名詞を後に従えた、と考えられる。



【図-16】語頭(休止の後)に現れた弱勢代名詞

一方、近世以降は、文頭(#)に限って見ても、〈命令形-代名詞〉は上昇し、定着した。現代語でも、*mira, mire*のような談話標識を除けば、命令形が文中で使われることは稀で、ふつうは休止の後の文頭、または準文頭で使われる。

よって、命令形に代名詞が後置する理由は、俗ラテン語において、ほとんど排他的に〈命令形-代名詞〉が使用され、俗ラテン語から中世スペイン語、近世・現代スペイン語に至るまで、命令形は文頭および文頭に準じる位置に現れることがふつうであったので、弱勢であった代名詞は、文頭で代名詞がその支えとなる要素のない文頭に立つことができなかつたため



ある、と思われる。やがて、〈命令形-代名詞〉は、〈代名詞+命令形〉が皆無に近かったために、ほぼ唯一的に単一語(→3.2)となり、安定して保持された<sup>25</sup>。

このように、〈命令形-代名詞〉(*ten-lo*)と(活用形-代名詞) (*tiene-lo*)の大きな相違点は前者の出現位置が文頭に限られ、後者の出現位置は自由であったことである。〈命令形-代名詞〉(*ten-lo*)は、文頭に代名詞が出現しなかった a1200-a1500 で保持され、一方、(活用形-代名詞) (*tiene-lo*)は、同様に代名詞を後置するので a1200-a1500 で保持されたが、代名詞が文頭に現れた a1600-a1800 において競合する圧倒的多数の〈代名詞-活用形〉に負けて消滅した(→3.2)。

否定命令形は *non*, *no* が先行するので、代名詞はそれを支えとして、常に動詞の前に置かれ、一般的な名詞句のパターンと合致した(→3.1)。

## 3.6. 全体

### 年代の推移

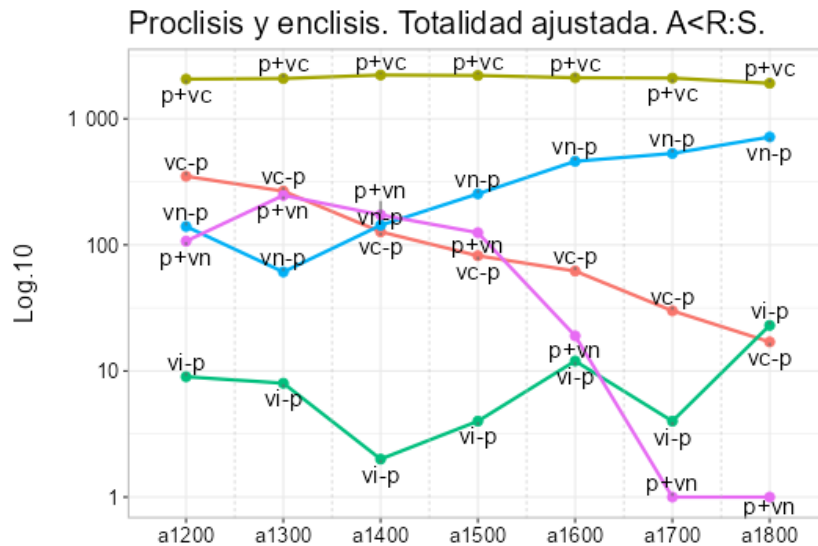
以上では、代名詞と動詞のすべての組み合わせのそれぞれについて、個別に分析してきた。そこで、このセクションでは全体を集計して比較することにする(p: pronombre, vc: verbo conjugado, vn: verbo no conjugado)。

次の図はそれぞれのケースの安全頻度を実測値の平均値によって調整した値である<sup>26</sup>。

---

<sup>25</sup> Real Academia Española y Asociación de Academias de la Lengua Española (2009: 1207)は次のような命令形の融合を"integración de los pronombres enclíticos en la base a la que se adjuntan"の例として挙げている: *pongámos + nos > pongamonos*, *digamos + selo > digámoselo*, *callad + os > callaos*。García de Diego (1970: 221)は中世スペイン語の *dadnos > dandos* の変化を単一語の *retina > riedna > rienda* と比較し、*dadla > dalda* の変化を単一語の *spatula > espadla > espalda* と比較している。

<sup>26</sup> パーセントにしても同じ曲線が得られるが、平均値調整頻度を使用することによって、実測値の規模で全体を観察できる。



【図-17】 全体の変化. 年代<地域：社会（平均値調整頻度：対数軸）.

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

上の図を見ると、活用形(vc)については、〈代名詞＋活用形〉(p+vc: *lo+tiene*)が恒常的に特出した最大の頻度を示し、〈活用形-代名詞〉(vc-p: *tiene-lo*)が減少していることがわかる。よって、全体の歴史的变化は代名詞後置(enclisis)→代名詞前置(proclisis)という変化ではなく、代名詞前置は変化していないので、むしろ、代名詞後置の減少とするほうが正しい。

代名詞後置(vc-p)、すなわち〈活用形＋代名詞〉(*tiene-lo*)は中世に比較的多く見られた古形である。それが古形であることから、古形を残す半島西部(CV, CN, LE)に多く分布し(→3.1.図-2, Andrés Díaz (2013: 651))、荘重な文体を使用する王室文書(C)、司法文書(J)、教会文書(E)に多用された(図-3)。

さらに時代が下って、19世紀の Pérez Galdós, *Marianela* (1878)でも代名詞後置(vc-p)は、会話文(話し言葉)には少なく、多くは地の文(書き言葉)の中で使われていることは先述したとおりである(→3.2.)。代名詞後置(vc-p)は現代スペイン語(20世紀以降)でも稀に見ることがあるが、やはり、古風な書き言葉に限られる(Keniston 1937: 68, Eberenz 2006: 174)。よって、代名詞後置の減少は、それが中世スペイン語の語形が残存した古形が有標化した「下からの変化」(cambio desde abajo)である、と考えられる。

非活用形(vn)については、中世において(a1300-a1400)、〈代名詞+非活用形〉(p+vn: *lo+tener*)が〈非活用形-代名詞〉(vn-p: *tener-lo*)よりも多かったが、近代になって(a1500-)、前者は減少し、後者が単一語化によって増加して逆転した。この場合は代名詞前置→代名詞後置という歴史的变化が認められる。

〈命令形-代名詞〉は頻度は少なく、年代的推移はほとんどない。最後の A1800 に、〈活用形-代名詞〉(vc-p)と〈代名詞+非活用形〉(p+vn)の減少・消滅により、〈代名詞+活用形〉(p+vc), 〈非活用形-代名詞〉に続いて第 3 位を占め、この上位 3 者が現代語の状態を決定した。

その全体をまとめると次の表のようになる。

| 形式         | 例                     | 変化           |
|------------|-----------------------|--------------|
| 〈代名詞+活用形〉  | p+vc: <i>lo+tiene</i> | 1 (圧倒的多数)    |
| 〈活用形-代名詞〉  | vc-p: <i>tiéne-lo</i> | 2→4 (下降)     |
| 〈非活用形-代名詞〉 | vn-p: <i>tener-lo</i> | 3→2 (上昇)     |
| 〈代名詞+非活用形〉 | p+vn: <i>lo+tener</i> | 4→3→4→5 (僅少) |
| 〈命令形-代名詞〉  | vi-p: <i>ten-lo</i>   | 5→3 (頻度は無変化) |

【表-8】 全体の変化 (数字は順位を示す)

全体的に、p+vc (*lo+tiene*)が常に圧倒的多数を占め、それに続いて vn-p (*tener-lo*)が常に上昇した。よって、全体的に見るならば、圧倒的多数を占めた p+vc (*lo+tiene*)と、定常的に増加した vn-p (*tener-lo*)の動きが 2 大潮流として認められる。

このグラフに基づいて全体を次の 3 つのグループに分類できる。

- (1) 〈代名詞+活用形〉：多数で安定(無標)
- (2) 〈非活用形-代名詞〉：上昇(無標): 〈言語的革新〉
- (3) 〈活用形-代名詞〉と〈代名詞+非活用形〉:下降(有標): 〈言語的伝統〉
- (4) 〈命令形-代名詞〉：少数で安定・無変化(標識なし)

5 形式の頻度(絶対頻度)・名詞句型パターン・単一語化・標識をまとめると次のようになる。

| 形式            | 例               | 頻度     | 名詞句型パターン | 単一語化   | 標識 |
|---------------|-----------------|--------|----------|--------|----|
| 1. 〈代名詞+活用形〉  | <i>lo+tiene</i> | 35,986 | V        | -      | 無標 |
| 2. 〈活用形-代名詞〉  | <i>tiéne-lo</i> | 2,135  | -        | V(不完全) | 有標 |
| 3. 〈非活用形-代名詞〉 | <i>tener-lo</i> | 5,842  | -        | V(完全)  | 無標 |
| 4. 〈代名詞+非活用形〉 | <i>lo+tener</i> | 1,536  | -        | -      | 有標 |
| 5. 〈命令形-代名詞〉  | <i>ten-lo</i>   | 303    | -        | V(不完全) | なし |

【表-9】 前置・後置 活用形・非活用形の絶対頻度

上表の 5 形式の中で、頻度が高く無標の 1. 〈代名詞+活用形〉(*lo+tiene*)と 3. 〈非活用形-代名詞〉(*tener-lo*)が二大潮流を成して現代スペイン語の姿となっ

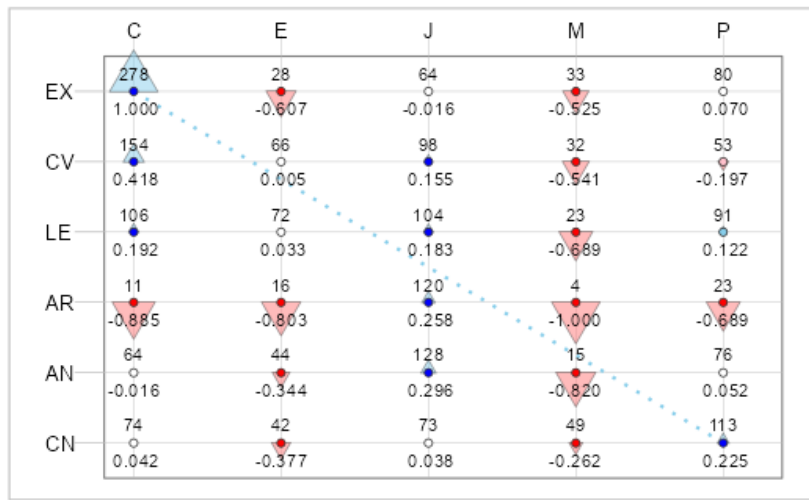
た。その中でとくに、1. 〈代名詞+活用形〉が〈範列変化+屈折変化〉という語順であったため、〈定冠詞・所有詞+名詞〉のパターンと一致したことにより、安定して圧倒的多数を占めた。それに続く 3. 〈非活用形-代名詞〉(*tener-lo*)は完全な単一語化によって、次第に勢力を増した。一方、3. 〈活用形-代名詞〉(*tiene-lo*)はラテン語の統語形式の残存である。これは、単一語化が不完全であったため、有標化し頻度が減少した。4. 〈代名詞+非活用形〉(*lo+tener*)もラテン語の統語形式の残存である。これは、スペイン語の名詞句のパターンと合致せず、代名詞が前にあるため単一語化も不可能であったため、a.1600 に消失した。5. 〈命令形-代名詞〉は文頭だけに現れ、独自に保持された。

### 地域・社会の変異

次に、地域：社会変数を対象変数とし、年代を参照変数とした二重均衡分析の結果に集中分析を実行する。地域変数と社会変数には特定の順番を前提としないので、行(地域)と列(社会)の集中化を行った。

|     |     |    |     |    |     |   |
|-----|-----|----|-----|----|-----|---|
| (4) | *   | 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



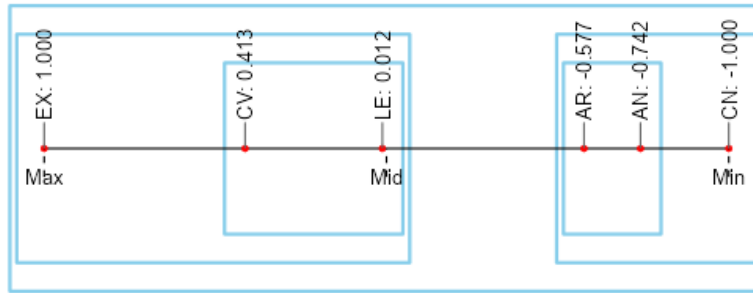
【図-18】 地域：社会<年代.安全頻度.集中分布

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

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

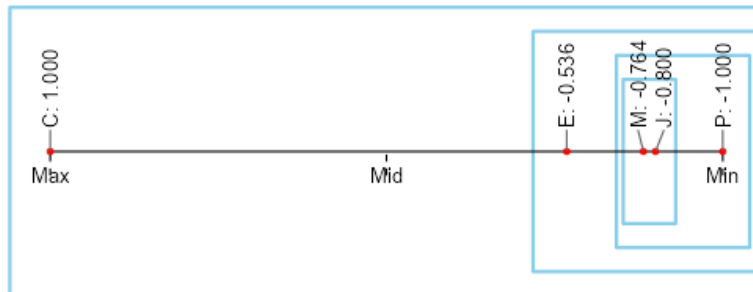
この集中分布図の左上の領域で[地域 CV-EX:社会 C]が大きなグループを成して enclisis を多用している。一方、[AR-AN-CN]が別のグループを形成している。前者が旧来の伝統的文体を示し、後者が新式の民衆的文体を示している、と考えられる。

次の図は上と同じ「地域：社会<年代.集中分布」の間隔クラスター図である。



【図-19】 地域：社会<年代. 地域.間隔クラスター

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



【図-20】 地域：社会<年代. 社会.間隔クラスター

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

地域変数は

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

の連続構造になり、社会変数(文書)は

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

という連続構造を示している。この社会変数の連続は「伝統→革新」という文体の推移を示していると思われる。それが正しいとすれば、地域と社会について同じグラデーションが存在すると考えられる。すなわち、 $[EX>[CV>LE]]$  (北西部地域) + C が伝統的な地域と社会のグループを形成する一方で、 $[[AR>AN]>CN]$  (南東部地域) +  $[E>[[M>J]>P]]$  も革新的なグループを形成している。

Cvで伝統的なC, Eの数値が高く、CNで革新的なM, J, Pの数値が高い。上の図を見ると、地域変数は[EX>[CV>LE]]と[[AR>AN]>CN]という2つのグループに分かれ、社会変数は[C]と[E>[[M>J]>P]]という2つのグループに分かれる様子が見える。地域について見れば、CVとEXは北と南東の周辺にあり、革新の中心CNは中央に位置する。CNを囲むLE, AR, ANはCNとともにM, J, Pの社会変数値が高く、民衆的革新のグループに入る。以上は全年代を通して観察した地域と社会との関係である。

## 4. 結論

以上で、2.方法に掲げた次の仮説が妥当であることを確認した。

【仮説】中世から近世・現代まで通して〈代名詞+活用形〉が無標の語順であった。ラテン語から継承した〈活用形-代名詞〉は有標の語順であり、一定の条件のもとで最近まで残存した。一方、〈非活用形-代名詞〉は独自の発展を遂げた。

以下に、2.方法に掲げた問題を再掲し、本論の中で得られた解決を「→」で示して載せる。

問題-1:中世語(文献資料:a1200-a1400)では *enclisis (tiene-lo)*のほうが多く、*proclisis (lo+tiene)*は一定の条件のもとに限られたのであろうか。

→中世語でも現代語のように *proclisis (lo+tiene)*が全体的に多用された。逆に *enclisis (tiene-lo)*のほうで、語頭などの特定の条件のもとで現れた。

問題-2:全歴史(文献資料:1200-1800)を通じて動詞活用形において *enclisis (tiene-lo)*から *proclisis (lo+tiene)*への移行という変化は認められるであろうか。

→当初(1200)から *proclisis (lo+tiene)*が圧倒的に多かった。よって、*enclisis (tiene-lo)*から *proclisis (lo+tiene)*への移行という変化はなく、*enclisis (tiene-lo)*が減少したという変化があった。*proclisis (lo+tiene)*が圧倒的に多かった要因は〈代名詞+活用形〉という形式が〈弱勢語+強勢語〉のパタン、とくに〈定冠詞・所有詞+名詞〉のパタンと類似し、両者が高頻度で出現したことが考えられる。一方、*enclisis*が減少した要因としてその形態の特殊性と使用頻度の有標性が挙げられる。

問題-3:全歴史を通じて動詞非活用形において *proclisis (lo+tener)* から *enclisis (tener-lo)* への移行という変化は認められるであろうか。

→動詞非活用形において *enclisis (tener-lo)* が全歴史を通じて次第に増加した。その要因として〈非活用形-代名詞〉(*tener-lo*)は単一語を形成し、2語である〈代名詞+非活用形〉(*lo+tener*)とは異なるパターンを形成したことが考えられる。

問題-4: 肯定命令形の *enclisis (ten-lo)* の歴史的理由は何か。

→命令形は文頭、または文頭の等位接続詞の後、または分詞構文、主題、従属文の後などの文頭に準じる位置で用いられる。文頭または準文頭の位置では、弱勢代名詞は支えとなる強勢語がないため、動詞(強勢語)の後に留まったため、肯定命令形の *enclisis* が生まれた。否定命令形は *non, no* が先行するため、それを支えとして、代名詞は動詞の前に置かれた。

先行文献は、中世で〈活用形-代名詞〉(*tiene-lo*)という語順がふつう一般に多く用いられ、それが近世以降より多かった、と指摘したが、〈代名詞+活用形〉(*lo+tiene*)の頻度を考慮していない(中岡 1993: 40 を除く)。本研究では、中世の〈活用形-代名詞〉(*tiene-lo*)を、当時多大であった〈代名詞+活用形〉(*lo+tiene*)と比較した。さらに、〈定冠詞+名詞〉(*la casa*)と〈所有詞+名詞〉(*mi casa*)と比較した結果、〈代名詞+活用形〉(*lo+tiene*)が文内の位置と大きな関係はなく、むしろ名詞句などの〈弱勢語+強勢語〉の一般的多数・増加の一環として捉えられる、という結論に達した。

一方、〈非活用形-代名詞〉(*tener-lo*)は2語の連続でなく、形態統語論的に融合した単一語として見なせば、〈弱勢語+強勢語〉という一般的なパターンの反例にはならないことを述べた。〈非活用形-代名詞〉を単一語と見なす根拠として動詞と代名詞の融合形、*Lo voy a leer y resumir. / Voy a leerlo y [resumir]lo / \*resumir*.などの統語的対立、中世スペイン語で *les yo fago* のような〈内挿入〉が *proclisis* で起こり *enclisis* では起きなかったこと、そして書き言葉に見られる多くの結合形を挙げた。

〈命令形-代名詞〉の語順は俗ラテン語において文頭の統語的特徴であり、それが中世スペイン語に継承された。これは、中世においては文頭に弱勢代名詞を置くことができなかつたためである。それが近代になって、文頭に弱勢代名詞を置くことが可能になったため、それまで主に文中に出現していた〈代名詞+活用形〉(*lo+tiene*)が文頭にも現れるようになった。一方、〈命令形-代名詞〉(*ten-lo*)の語順はそのまま保持された。その理由は、中世



スペイン語で文頭に代名詞は出現しにくいので、〈代名詞＋命令形〉は皆無に近かったためである。

一般に、言語変化には[A]→[B]という二項の完全な交代はなく、[A]→[AB](共存:Aが優勢)→[BA](共存:Bが優勢)→[B]という移行期を経由する。よって、〈命令形-代名詞〉→〈代名詞＋命令形〉という完全な交代は不可能であった。また、[A]→[AB](共存:Aが優勢)という過程も生じなかった。それは、話し言葉の命令形の直接的表現において動詞(命令形)が目的語(代名詞)に優先されたことと、〈非活用形-代名詞〉の単一語化が類推的に影響したことが相まって〈命令形-代名詞〉が定着していたので、それが逆転した〈代名詞＋命令形〉(\**lo+ten*)が生じる余地がなかったからである。

全資料を調査した結果、全歴史を通して〈代名詞＋活用形〉(*lo+tiene*)が圧倒的に多かったことと、〈活用形-代名詞〉(*tener-lo*)の頻度が次第に高くなったことを二大潮流として確認した。この2つの形式が無標であり、現代語の排他的分布を決定した。無標の〈代名詞＋活用形〉(*lo+tiene*)に対立する有標の〈活用形-代名詞〉(*tiene-lo*)と、無標の〈活用形-代名詞〉(*tener-lo*)に対立する有標の〈代名詞＋非活用形〉(*lo+tener*)は次第に減少し、19世紀(a1800)にはほとんど消滅した。地域変数を見ると、これらの有標の古形の残存は年代的には中世、地域的には半島の西部と北部、社会的には王室文書の特徴であることを観察した。

各形式の分析では、それぞれの形式の標識(無標/有標)を意識した。これは、それぞれの形式の頻度が大きく異なるからである。このような標識を認めると、言語体系の記述は有標項の特徴を示すことに限られ、無標項についてはとくに記述は不要になる。よって、活用形の有標項は〈活用形-代名詞〉(*tiene-lo*)であり、その特徴は〈古形〉であること、〈文頭位置〉に優先的に出現したこと、中世のCV, LE, EX, および中世初期のCに多く出現したことである。

一方、非活用形の有標項は〈代名詞＋非活用形〉(*lo+tener*)であり、その特徴は〈古形〉であることと、〈年代:中世初期(a1300, a1400)〉〈地域:LE, CN, EX〉〈社会:C, P, E〉という区分に集中することである。これらの有標項の特徴は新しく誕生した構造の特徴ではなく、旧来の構造が残存するときの条件として機能した。

本研究が達した結論は先行研究の見解と大きく異なる。これは、先行研究の大部分が伝統的な文体の文学作品に基づいているのに対し、本研究が

現実の言語社会を反映した公証文書を扱ったことに起因する可能性がある<sup>27</sup>。言語史研究においては、文学作品も重要であるが、現実の社会で使われた言語をとくに重視すべきであろう。

本研究では言語の形態・構造の問題を歴史・地域・社会の変数との関係を見ながら分析した。歴史・地域・社会と関係する言語特徴は「伝統」〈代名詞+動詞〉から「革新」〈動詞-代名詞〉へ、という大きな推移の中で把握された。この推移は言語・歴史・地域・社会という異なる領域に共通する統一的な変化である。

こうした歴史的洞察は現代スペイン語の理解に役立つ。たとえば、スペイン語の代名詞が動詞の前につく理由や、不定詞・現在分詞・命令形のとときに動詞の後につく理由が歴史的に説明される。また、比較的最近まで動詞活用形-弱勢代名詞が文語体の文頭に限って出現した理由も言語史を辿ることによって明らかになる。そのときの言語史は単に年代の推移を見る一次元的な直線ではなく、地域差・社会差を含めた多次元を構成するものでなければならない。

Saussure (1916, 1972)は「共時態」と「通時態」を峻別し、両者の混同を避けたが、言語の現在の姿は長く広大な歴史・地域・社会の変化・変異によって形成されたものである。よって、共時的・一面的に限定した分析だけでなく通時的・多面的な分析も必要である。共時的な特定の地域・社会の言語研究においても通時的多面的諸相を観察することが示唆的であるからである。

---

<sup>27</sup> Castillo Lluch (1996: 87)によると、15世紀の文献では代名詞前置(p+vc)は話し言葉の直接話法や手紙文で用いられ、文学作品では規範的な代名詞後置(vc-p)が用いられていた。

## 5. 参考文献

- 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. Klincksiek.
- 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. Hakuishisha.
- 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*. 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" in Paolo

- Pocetti (ed.), *Latinitatis Rationes: Descriptive and Historical Accounts for the Latin Language*, 246-261.
- [https://www.academia.edu/5774218/\\_The\\_position\\_of\\_weak\\_pronouns\\_in\\_Classical\\_Latin\\_prose\\_Submitted](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. 1993. 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. 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.
- 寺崎英樹. 2011. 『スペイン語史』 東京. 大学書林.
- Trask, R. I. 2000. *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 , pp. 51-105.. (en japonés)  
<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.
- 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](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.【付録】 統計的方法

### 6.1. 中央値偏差

数値データ行列の特徴を探るとき、単に大きな値だけでなく、小さな値にも注意しなければならない。なぜなら、全体の中で「語形が多く使用されている」という観察だけでなく、「語形が使用されていない」「使用率が低い」という観察も必要になるからである。そのとき、小さな数値の「小ささ」を大きな数値の「大きさ」に合わせて(同じ規模で)、負の数(マイナス)で示すとわかりやすい。そして、その範囲を[-1, 1]とすれば、全体の中で標準化されたスケールの中で判断することができる。

中央値はソートされたデータを二等分する中心点を示すので、データから、データの中央値を引くと、中央値より大きな数が正(プラス)になり、中央値より小さな数が負(マイナス)になり、中央値はゼロになり、正の数値の数と負の数値の数が等しくなる。そのように変換されたデータのそれぞれの値について、正の数は最大値で割ると、[0, 1]の範囲になる。負の数は最小値の絶対値で割ると、[-1, 0]の範囲になる。この方法を「中央値偏差」(*deviation from median*)と呼ぶ。

次は、簡単な例で計算した中央値偏差の作成法である。

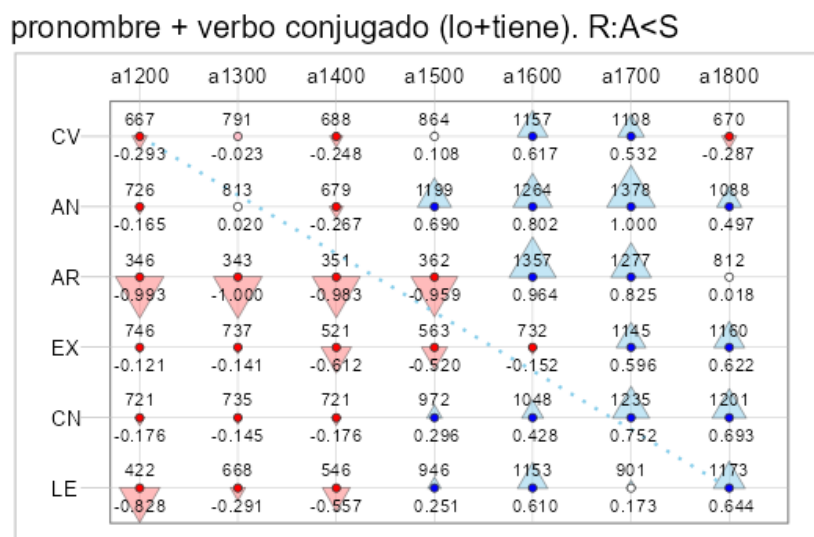
```
(5)  V=c(1,3,5,8,9)
      W=V-Median(V); W # -4 -2  0  3  4
      ifelse(W>0,W/Max(W),W/abs(Min(W)))
      # -1.00 -0.50  0.00  0.75  1.00
```

上の数値ベクトル  $V$  (1、3、5、8、9)の中央値は5である。 $V$  から5を引くと  $W$  (-4, -2, 0, 3, 4)になる。このとき、 $V$  の中央値5は  $W$  では0となる。 $W$  の最大値が1となるように全体を変換するために、0を超えた数値(3, 4)を  $W$  の最大値4で割ると、(0.75, 1)になる。中央値以下の数値(-4, -2, 0)は  $W$  の最小値(-4)の絶対値4で割ると、(-1, -0.5, 0)となる。このとき  $W$  の最小値(-4)で割ると、結果は(1, 0.5, 0)となるので、 $W$  の最小値(-4)の絶対値4で割る。

#### 応用

次のグラフは〈代名詞+活用形〉(*lo+tiene*)の均等分析(→6.7)の結果であ

る。それぞれのセルには地域と年代の正規化頻度(→2.8)と中央値偏差を載せてある。ポイントの色は有意度(→2.7)を示し、三角形の大きさは中央値偏差に従っている。



【図-21】 〈代名詞+活用形〉 地域:年代<社会

## 6.2. 平均値調整頻度

一般の相対頻度や正規化頻度、そして正規化頻度(後述→2.8)の計算では、一定の乗数を使って数値を拡大している。たとえば、相対頻度は  $30/100 = 0.3$  という小さな数値になるが、百分率(パーセント)にするには  $100$  という乗数を掛けて、 $0.3 * 100 = 30$  パーセントになる。正規化頻度では  $100$  万 ( $10^6$ )や  $10$  万 ( $10^5$ )などの乗数が選択される。このとき、積算の結果が元の実測値の規模に近い数になるような乗数を使うことが望ましい。

そこで、乗数として  $10^6$  や  $10^5$  などの切りの良い数値ではなく、相対頻度・正規化頻度の平均値と実測値の平均値が等しくなるような数値を使えば、出来上がった数値行列の規模が実測値の規模に等しい数値になるので、数値の実態をそのまま観察できる。具体的には正規化頻度・正規化頻度に次の乗数(m)を掛けることにする。

$$m = \text{実測値の平均値} / [\text{相対頻度} \cdot \text{正規化頻度}] \text{の平均値}$$

そのようにして調整した頻度を「平均値調整頻度」(*adjusted frequency by mean*)と呼ぶ。

次は、〈活用形-代名詞〉の地域\*年代の実測値(絶対頻度)のクロス集計と、

その平均値(mo)である。

| (6)         | * | 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) |   | 50.80952 |       |       |       |       |       |       |

次は、〈活用形-代名詞〉の地域\*年代の正規化頻度のクロス集計と、その平均値(ms)である。

| (7)         | * | 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 |       |       |       |       |       |       |

上の2つのデータを使って、平均値調整正規化頻度とその平均値(ma)を計算すると、次のようになる。

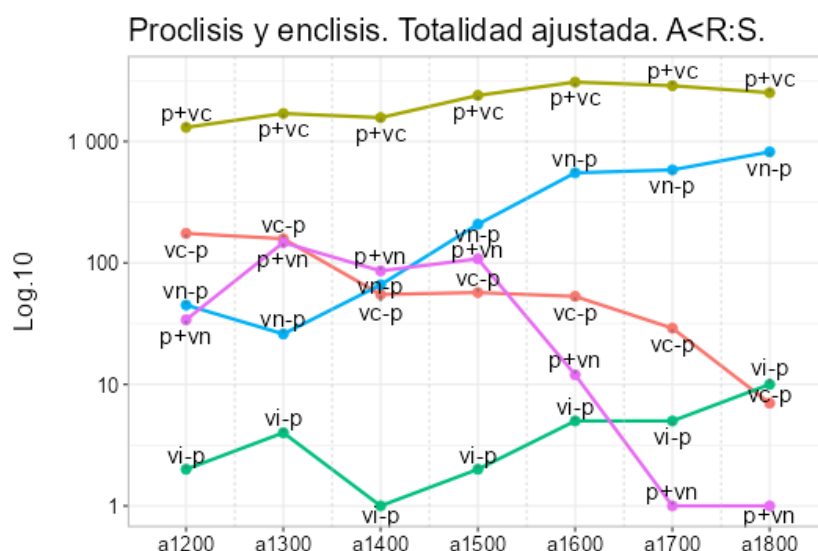
| (8)         | * | a1200    | a1300 | a1400 | a1500 | a1600 | a1700 | a1800 |
|-------------|---|----------|-------|-------|-------|-------|-------|-------|
| AN          |   | 117      | 40    | 33    | 33    | 58    | 37    | 21    |
| AR          |   | 12       | 21    | 20    | 14    | 70    | 38    | 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 |       |       |       |       |       |       |

たとえば、AN:a1200の平均値調整頻度(=117)は  $272 * mo / ms = 116.8216$  で求められる。このように、平均値調整頻度の平均値は実測値の平均値と一致するので((2), (4))、平均値調整頻度は相対的な頻度を実測値の規模に合致する現実的な実態を示している。よって、平均値調整をしない正規化頻度は数値の絶対的な大小を評価できないが、平均値調整頻度を使えば、それが可能になる。これは言わば「相対値の現実的な絶対化」である。



## 応用

代名詞と動詞のすべての組み合わせのそれぞれについて全体を集計して比較する。次の図はそれぞれのケースの正規化頻度を実測値の平均値によって調整した値である。



【図-22】 全体の変化 (平均値調整頻度). 年代<地域：社会.  
(p: pronombre, vc: verbo conjugado, vn: verbo no conjugado)

### 6.3. 対称値

言語形式が出現するときに、考えられる条件を  $x$  とし、出現形式を  $y$  として、それぞれの条件( $x$ )を変えて〈代名詞+活用形〉( $y$ )を検索しながら次の  $a, b, c, d$  を計算する。

- $a(x+, y+)$ : 条件( $x$ )があるときに  $y$  が検索される場合の数
- $b(x+, y-)$ : 条件( $x$ )があるのに  $y$  が検索されない場合の数
- $c(x-, y+)$ : 条件( $x$ )がないのに  $y$  が検索される場合の数
- $d(x-, y-)$ : 条件( $x$ )がないときに  $y$  が検索されない場合の数

最後に次の  $r$  という係数によって条件の有効性を数量化する：

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

この  $r$  のように分子を 2 項の差とし、分母を 2 項の和とする数値を「対称値」(contrast)と呼ぶことにする。対称値の範囲は $[-1, 1]$ である。対称値  $r$

は、 $a=0$  のとき最小値(-1)となり、 $b=0$  のとき最大値(1)となり、 $a=b$  のとき中間値(0)となる。対称値  $r$  が大きいほど条件と結果が強く関係していることになる。対称値  $r$  が 0 に近いほど、その関係がないことを示す。また、対称値  $r$  が -1 に近いほど、逆転した関係が強いことを示す。

この  $r$  では、 $a$  と  $b$  に限って簡単にしたが、他にも次のような係数が考案されている (Romesburg 1989, 1992: 177-202)。これらは「連関係数」(association score)と呼ばれる。

$(a+d) / (a+b+c+d)$  # Simple matching  
 $a^2 / (a^2+b+c)$  # Dice  
 $a / (a+b+c)$  # Jaccard  
 $a / s$  # Russel & Rao  
 $a^m / (a^m+b+c+d)$  # Russel & Rao.m  
 $((a+d)-(b+c)) / ((a+d)+(b+c))$  # Hamann  
 $(a*d-b*c) / (a*d+b*c)$  # Yule  
 $(a*d-b*c) / \sqrt{((a+b)*(c+d)*(a+c)*(b+d))}$  # Phi  
 $a / \sqrt{(a+b)*(a+c)}$  # Ochiai  
 $(a^2-b-c) / (a^2+b+c)$  # Ueda (私たちの提案)  
 $\log_2(a*s / (a+b) / (a+c))$  # Mutual information  
 $1 / \sqrt{a*(a-(a+b)*(a+c) / s)}$  # t-score

## 応用

次は、文頭(#)・文中(=)という条件(x)を比較して〈代名詞+活用形〉(*lo+tiene*)(y)の対称値(r)を計算した結果である。総語数(s)は関連する代名詞と動詞の組み合わせの合計値である(資料は a1200-a1400 の 2 つの語順 *lo+tiene*, *tiene-lo* を抽出したもの)。

|     |          |      |         |        |        |        |        |        |         |       |
|-----|----------|------|---------|--------|--------|--------|--------|--------|---------|-------|
| (9) | 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 |

【表-10】 文頭(#)・文中(=)と〈代名詞+活用形〉の関係

この結果は、条件が文頭(#)であるとき〈代名詞+動詞〉(*lo+tiene*)が比較的出現しにくく、条件が文中(=)であるとき、出現しやすいことを示している。しかし、それでも、文頭において〈代名詞+動詞〉(*lo+tiene*)の語順のほうが(=19)、逆の語順(=17)よりも多かった。

さらに、条件(文頭#)が満たされなくても出現したケースもある(上の c)。そして、条件が満たされなくて出現しなかったケースも多い(上の d)。しかし、この場合、c と d は非常に多く、これらを取り入れた係数を使用すると、c と d の影響を大きく受けた結果となり、本来の a と b の関係が見えなくなる。むしろ、条件を変えて、文中(=)の中で、a, b の対称値を見るほうがよい。

## 6.4. 大数平均値

次は 4 行 5 列のデータ例(Data)と、行の平均値(mean)と中央値(median: データをソートしてその中央にある数値)の出力(Stat)である。

| (10) | [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      |

このデータ(Data)には次第に大きな外れ値が含まれている(列 c5: 10, 100, 200, 300)。このように大きな外れ値のあるデータの平均値には、外れ値に向かって大きく偏りが見られる(r2, r3, r4)。そのような偏りがあるデータの平均値はデータを代表しているとは言えない。たとえば、上の r4 のようなデータの平均値(=62)は大多数のデータ(1, 2, 3, 4)から大きく乖離しているからである。

一方、中央値には外れ値の影響がないので、外れ値があるデータでは、その中心を示す代表値として平均値の代わりに中央値の使用が勧められている(たとえば Morales Vallejo 2008: 38, Vickers 2010: 4-5)。しかし、中央値には、その左右にある数値がどのように変化しても、それを感知せず常に一定である、という欠点がある: たとえば、(3, 5, 30, 32, 35), (10, 12, 30, 42, 45), (10, 20, 30, 420, 450), etc.の中央値はどれも 30 となる。そのような中央値は全体のデータの中心を適切に示しているとは言えない。また、外れ値の有無によって平均値から中央値へ切り替えるとき、外れ値の大小は連続的なので、その切り替えの閾値(threshold)の設定が恣意的になる。

データの外れ値の影響を少なくする方法として「トリム平均値」(trimmed mean)が考案された(芝・渡部・石塚 1984:173「調整平均」)。トリム平均値では昇順にソートしたデータの最小値から数えて k 個と最大値から数えて k 個が切除されるので、切除された要素の情報は失われる。当然、

切除された値がどのような数値であろうと感知しない。中央値と同様にトリム平均値の使用の基準(どれほどの外れ値があればトリム平均値を使用するかを決める基準)の設定は恣意的である。また、切除する個数(k)の設定も恣意的になる。

そこで、最小値と最大値付近だけをトリム(切除)するのではなく、データ数の半数以上が含まれるグループ(ソート済み)を左から徐々に切り出しながら、各段階の平均値を求め、その和の平均を求めて「大数平均値」(*Majority mean: M.mean*)とする方法を考える。以下は、昇順にソートしたデータ(1, 2, 4, 5, 55)を使ったプロセスを示している(NA: *not available*: 計算に含めない)。このデータ数は5なので、その過半数は3個になる。

|      |    |    |    |    |    |      |
|------|----|----|----|----|----|------|
| (11) | 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 |

大数平均値を計算するプロセスは次のとおりである:

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

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

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

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

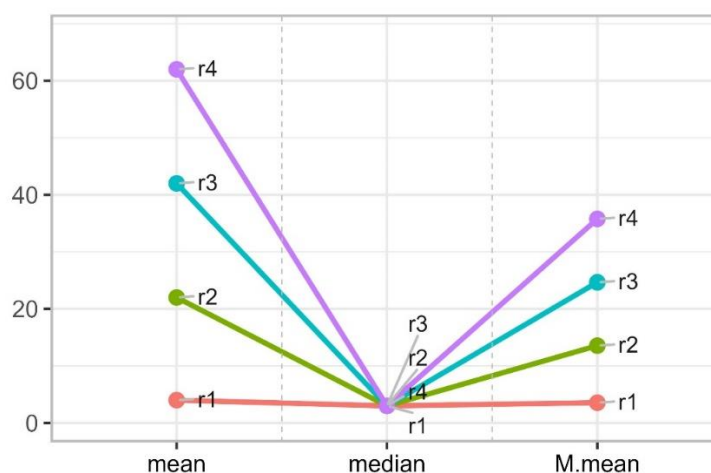
最初が一番左の要素 c1, c2, c3 の数値 1, 2, 4 の平均値(=2.3)を求める(r1)。次に、右に続く3要素 c2, c3, c4 の数値 2, 4, 5 の平均値(=3.7)を求める(r2)。最後に c3, c4, c5 の数値 4, 5, 55 の平均値(21.3)を計算する(r3)。このとき、3要素の連続は最後になるので計算を終了し、r1, r2, r3 で求めた平均値の平均を計算し(=9.1)、これをデータの大数平均値とする。

この大数平均値は全データの中から過半数である3連続の標本をすべて抽出しているので、データの抽出法に「偏り」はない。そして、データの中心にある中央値(=4)が一番多く含まれ、周辺にある数値を含む標本が次第に少なくなり、最小値と最大値を含む標本はそれぞれ1個に限られているので、平均値に求められる「中心性」が確保されている。

次がデータ行列 **D** の行の平均値(*mean*)、中央値(*median*)、大数平均値(*M.mean*)の出力である。

| (12) | [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     |

次は同じデータの平均値(*mean*)、中央値(*median*)、大数平均値(*M.mean*)を比較した折れ線グラフである。

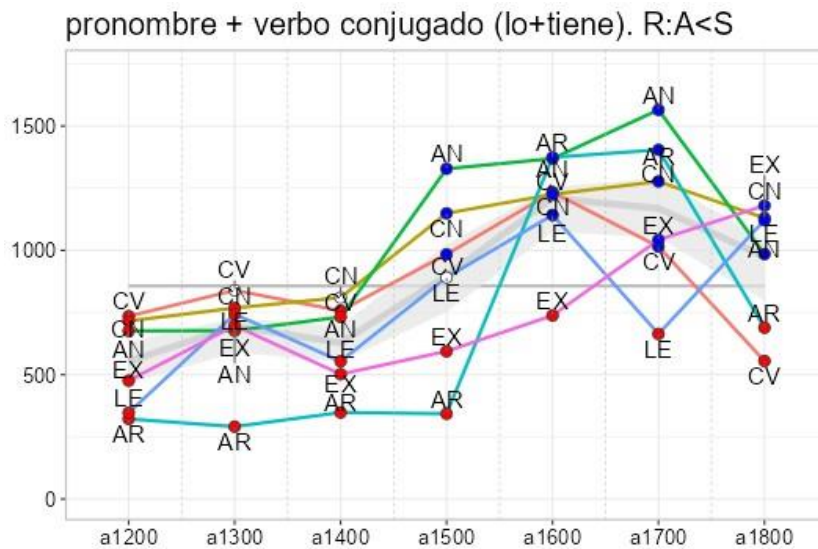


【図-23】 平均値(*mean*)、中央値(*median*)、大数平均値(*M.mean*)の比較

このグラフを見ると、平均値が外れ値の影響を強く受けていること、および中央値が外れ値の影響をまったく感知しないことがわかる。一方、大数平均値は平均値が外れ値の影響が少なく、その外れ値の大小を感知している。大数平均値はデータの分布の偏りが大きいとき(r3、r4)、平均値を大きく修正し、偏りが小さいときは平均値に近似し(r2)、偏りが僅少のときは平均値に一致するので(r1)、一般にすべてのデータを大数平均値を用いて比較することができる。

## 応用

次の〈代名詞+活用形〉(*lo+tiene*)の折れ線グラフには、灰色の水平線で全体の平均値を示し、灰色の折れ線でそれぞれの縦の大数平均値を示した。それを囲む灰色の面は大数平均値を超える値の半数を上限とし、大数平均値以下の値の半数を下限とした領域を示す。こうして、灰色の線と面を観察することによって、全体の傾向がわかる。



【図-24】 〈代名詞+活用形〉 地域:年代<社会

## 6.5. 集中分析

たとえば、次の表のデータ行列(D1)の行と列の順番を変えて、左上から右下に引いた対角線の近傍に集中する行列(D2)を作成する方法を「集中分析」(*Concentration analysis*)と呼ぶ。

| (13) | [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  |

上の[D1]は入力 of データ行列であり、[D2]は出力 of 集中行列である。D1ではデータの全体的傾向が見えないが、D2では数値が左上から右下に引いた対角線の近傍に高い数値が集中しているので、行 r3-r1-r2 と列 c1-c3-c4-c2 の連続が一定の同じ傾向(たとえば行:地域と列:年代など)を読み込むことができる<sup>28</sup>。

以下、次のサンプルデータ(D)を使って集中化の手順を示す。

| (14) | [D] | c1 | c2 | c3 | c4 |
|------|-----|----|----|----|----|
|      | r1  | 1  | 8  | 5  | 6  |
|      | r2  | 2  | 8  | 2  | 4  |

<sup>28</sup> 列に年代を指定して年代の推移を分析するとき、その列の順番を変えることはできない。そのときは行(たとえば語形、地域)だけの順番を変えて集中化する。

```

r3 9 2 4 1
A=D # 初期状態をAに保存する。

```

各行に与える原点からの距離を示すベクトル  $W$  を用意する(範囲[-1, 1])。以下では「原点(0)からの距離」を「原点距離」と呼ぶ<sup>29</sup>。

```
(15) -1.000 -0.333 0.333 1.000 ... W
```

データ(D)の各行に原点距離ベクトル  $W$  を掛けて加重する。これがそれぞれの数値の「原点(0)からの距離」を示す。

```
(16) *      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

```

この出力行列  $S$  の全体を 3 乗する(原点距離の 3 乗 :  $S_3$ )。

```
(17) *      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

```

この行列  $S_3$  の行和を計算する(原点距離の 3 乗の行和)。

```
(18)      r1      r2      r3 ... rowSums(S3)
200.667   37.333 -725.926

```

データ行列  $D$  の行和  $R_s$  を計算する。

```
(19) r1 r2 r3 ... Rs=rowSums(D)
20 16 16

```

3 乗の行列の行和  $R_s$  をデータ行列  $D$  の行和で割って平均値とする。

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

```

この平均値の 3 乗根を求め、元の単位に戻す<sup>30</sup>。これがそれぞれの原点距離(Minkowsky 距離)である。

<sup>29</sup> 集中化の係数として Minkowsky 距離(3 乗)を使用する。3 乗とするのは正負の別が保持されるためである(2 乗では正負の別がなくなる)。

<sup>30</sup>  $R$  の 3 乗根は負数を計算しないので(-27)、自作関数 `CubeRoot` を使用する:  
 $(-27)^{(1/3)} \rightarrow \text{NaN}$ ; `CubeRoot(-27) \rightarrow -3`.

|      |       |       |        |
|------|-------|-------|--------|
| (21) | r1    | r2    | r3     |
|      | 2.157 | 1.326 | -3.567 |

原点距離の昇順の順序(order)を求める(O)。r3 が 1 番であり、r1 が 3 番である。

|      |    |    |    |     |   |
|------|----|----|----|-----|---|
| (22) | r1 | r2 | r3 | ... | 0 |
|      | 3  | 2  | 1  |     |   |

原点距離の昇順の順序に従ってデータ行列の行の順番を変える。

|      |    |    |    |    |    |
|------|----|----|----|----|----|
| (23) | *  | c1 | c2 | c3 | c4 |
|      | r3 | 9  | 2  | 4  | 1  |
|      | r2 | 2  | 8  | 2  | 4  |
|      | r1 | 1  | 8  | 5  | 6  |

以上が最初の行の並べ替えである。「行の集中分析」であれば、ここで終了する。「行と列の集中分析」であれば、行の集中化と同様のことを列について行う。

|      |    |    |    |    |    |   |                |
|------|----|----|----|----|----|---|----------------|
| (24) |    | c1 | c3 | c4 | c2 | # | データ行列の列の順番を変える |
|      | r3 | 9  | 4  | 1  | 2  |   |                |
|      | r2 | 2  | 2  | 4  | 8  |   |                |
|      | r1 | 1  | 5  | 6  | 8  |   |                |

こうして、行と列の並べ替えを行った後、データ行列 D が並べ替えの前の行列 A と異なっていれば、再度、行と列の並べ替えを行う。このとき D が A と同じであれば終了する。

|      |     |    |    |    |    |     |    |    |    |    |
|------|-----|----|----|----|----|-----|----|----|----|----|
| (25) | [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  |

データ行列(D)は出力行列(C)のように集中化された。C の対角線近傍に大きな数値が見つかる。

## 応用

たとえば、二重均衡分析のプログラム(→2.6)は次のような 2 次元の行列を出力する。これは〈動詞活用形-弱勢代名詞〉という語形の頻度を地域・年代のクロス集計表の均衡分析の結果である(R:A<S: 平均値調整頻度)。

|      |   |       |       |       |       |       |       |       |
|------|---|-------|-------|-------|-------|-------|-------|-------|
| (26) | * | a1200 | a1300 | a1400 | a1500 | a1600 | a1700 | a1800 |
|------|---|-------|-------|-------|-------|-------|-------|-------|



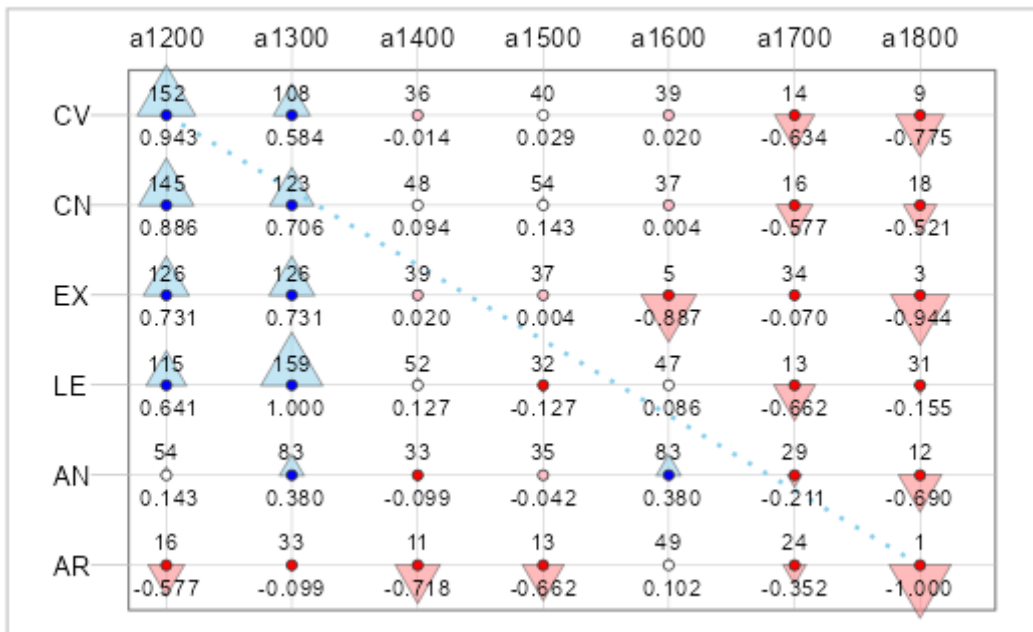
|    |     |     |    |    |    |    |    |
|----|-----|-----|----|----|----|----|----|
| AN | 54  | 83  | 33 | 35 | 83 | 29 | 12 |
| AR | 16  | 33  | 11 | 13 | 49 | 24 | 1  |
| CN | 145 | 123 | 48 | 54 | 37 | 16 | 18 |
| CV | 152 | 108 | 36 | 40 | 39 | 14 | 9  |
| EX | 126 | 126 | 39 | 37 | 5  | 34 | 3  |
| LE | 115 | 159 | 52 | 32 | 47 | 13 | 31 |

この行列の列(a.1200...a.1800)は年代なので、その順番は固定しなければならない。一方、行の順番(AN, AR, ..., LE)はアルファベット順になっており、これを変えて行列内の数値が左上から右下に引いた対角線の近傍に集中させることができる。この表によって、年代に即した地域の動きが行列の中で観察する。

|      |   |       |       |       |       |       |       |       |
|------|---|-------|-------|-------|-------|-------|-------|-------|
| (27) | * | 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     |

次のグラフのそれぞれのセルには、正規化頻度と、その下に、範囲が[-1, 1]の数値、正規化頻度を最大値=1, 最小値-1、中央値=0 となるように調整した数値(→2.1.中央値偏差)が示されている。

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

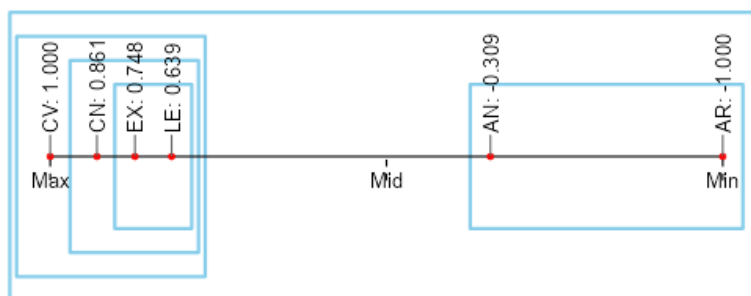


【図-25】 地域：年代<社会.正規化頻度.集中分布図

次は集中分析によって得られた縦軸の距離係数を直線上に配置し、個々の係数の間隔を計りながらグルーピングした結果である。

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

同様に、次のようなグラフを出力する関数を作成した。

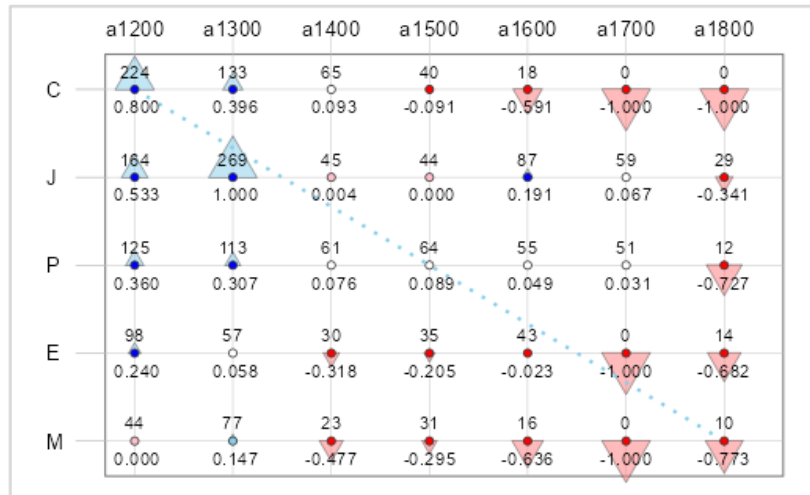


【図-26】 地域：年代<社会.正規化頻度. 地域の間隔クラスター

この図は集中分析に基づく地域の歴史的・言語的距離(間隔)の近さによる多重グループ構造を示している。数値の間隔が小さい部分は面積を拡大し、それが大きい部分は縮小することによって、グループの視覚化が明確になるようにした。このようなグルーピングは一般に各種の「クラスター分析」で行われるが、ここでは単純に数値の間隔だけを考慮した方法を試みた。この方法を「間隔クラスター分析」(*interval cluster analysis*)と呼ぶ。一般のクラスター分析のグラフは樹状図(*dendrogram*)で視覚化されるが、この間隔クラスター分析では上図のような内心構造で視覚化することにより、多重群構造によって次々に大きなグループにまとまる様子が直感的・視覚的に把握できる。

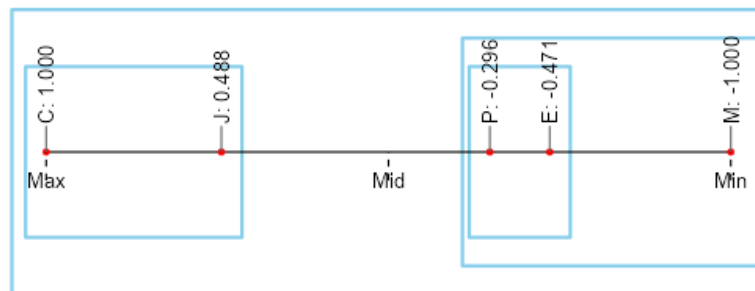
次は社会：年代変数を対象変数とし、地域を参照変数とした二重均衡分析の結果に集中分析を実行した結果である(S:A<R)。

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



【図-27】 社会：年代<地域.正規化頻度.集中分布

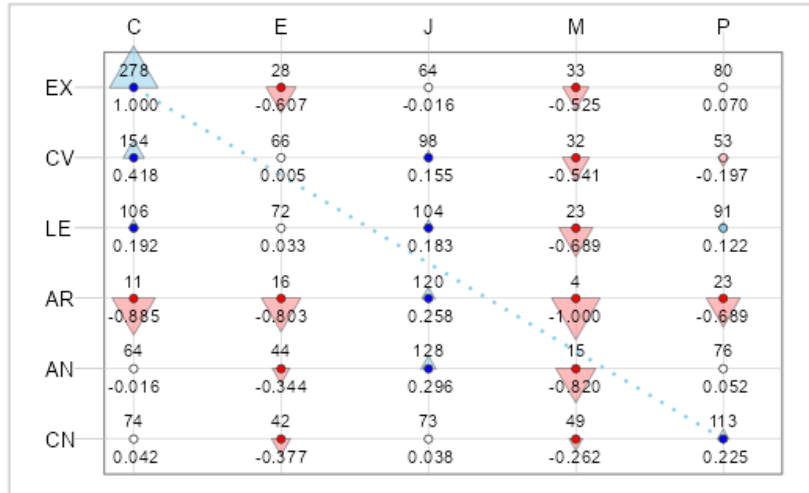
次の図は上の集中分布図の目的変数である社会の距離係数(上図の右の目盛)を、その間隔に基づいてグルーピングした間隔クラスターである。



【図-28】 社会：年代<地域.正規化頻度.社会の間隔クラスター

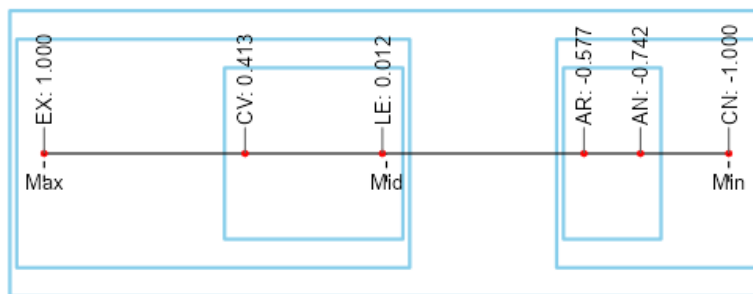
次に、地域：社会変数を対象変数とし、年代を参照変数とした二重均衡分析の結果に集中分析を実行する。今度は、地域変数と社会変数には特定の順番を前提としないので、行と列の集中化を行った。

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

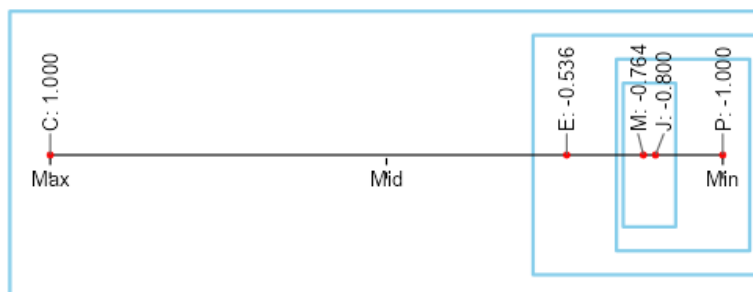


【図-29】 地域：社会<年代.正規化頻度.集中分布

次の図は上と同じ「地域：社会<年代.正規化頻度.集中分布」の間隔クラスター図である。



【図-30】 地域：社会<年代. 地域.間隔クラスター



【図-31】 地域：社会<年代. 社会.間隔クラスター

地域変数は

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

の連続構造になり、社会変数(文書)は

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

という連続構造を示している。

## 6.6. 有意度

たとえば、10 人のグループの中で 8 人が「(ある語形を)使用する」と答えた、とすると、この  $8/10 = 0.8$  という比率は「統計的に有意であるか否か」が問われる。このような場合の統計的検定には一般に「二項検定」(binomial test)が使われる。次は R 関数 `binom.test` の出力である(上側検定:greater)。

```
(28) 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
```

この出力の  $p$  値(*p-value*)は 0.05469 であり、これは「語形を使う人が 8 人以上」である確率の和を示している。語形を使うか否かはそれぞれ 0.5 の確率(ここで「期待確率」と呼ぶ)であるとする、それが 8 人以上である確率は 5%を超えていることになる。たとえば硬貨を 10 個投げて 8 枚以上が表になったときと同じである。それは偶然でも 0.05469 の確率(約 5.5%)で起こると予想される。はじめに有意水準として 5%を設定しておく、その 5%を超えているので「有意ではない」と判定される。ある語形を使う人が 9 人(以上)であれば、*binom.test* (9, 10, p=.5, 'greater')で  $p\text{-value} = 0.01074$  となり、そのようなことが偶然で起こる確率は約 1%なので、10 人中の 9 人は「有意な数値」である、と考えられる。なぜなら「10 人中の 9 人が使う」ということが偶然で起こる確率は非常に小さいからである(1%)。以下の出力

はそれぞれの二項検定の結果の p 値を示している。

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

このような統計的検定は、二項分布確率に基づいている。たとえば、1枚の硬貨を投げて、それが「表(おもて)」になる確率は 0.5 (50%)である。次の#1, #2 は 1枚の硬貨(サンプルサイズ=硬貨の数  $s=1$ )が「裏」( $x=0$ )である確率は 0.5 であり、「表」( $x=1$ )である確率も 0.5 であることを示している。このようなことは実験をしなくても、硬貨の面は「裏」と「表」であるから、それぞれの確率は 0.5 であることは直ちに理解できる。

```
(30) #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
```

しかし、2枚の硬貨を投げた場合の表が 0, 1, 2枚である確率については、少し考察が必要になる。このとき、すべての場合を列挙すると、「裏・裏」、「裏・表」、「表・裏」、「表・表」となるので、2枚の硬貨を投げた場合の表が 0枚, 1枚, 2枚である確率は、それぞれ、0.25 (「裏・裏」), 0.5 (「裏・表」、「表・裏」), 0.25 (「表・表」となる(下の#3, #4, #5)。さらに、たとえば、10枚の硬貨を投げて 9枚が表になる確率は、直感や考察では求めることは困難なので、関数 *dbinom* が必要になる。下の#6によれば、その確率は 0.009765625 である。

```
(31) 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
```

個別の確率ではなく累積確率(cp: cumulative probability)を求めるには、*pbinom*(q, s, p)を使用すればよい。たとえば 2個の硬貨を投げて表が 0枚である確率、1枚以下である確率(0枚と1枚が表になる確率の和)、2枚以下である確率(0枚と1枚と2枚が表になる確率の和)を求めるには以下の#7, #8, #9のように実行すれば、それぞれ 0.25, 0.75, 1 が得られる。

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

次の出力は、頻度(f)=0:2 (0, 1, 2), 和(s)=2, 期待確率(e)=0.5 として *dbinom* と *pbinom* を使って求めた確率密度(d)と累積確率(cp)と上側確率の和(up=1-

cp)を示している。ここで f が中央の部分(f=1)で最も確率密度(d)が大きくなり(d=0.50)、d の和が 1 であり、cp の到達点が 1 であることを確認できる。

| (33) | 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 |

次は、頻度(f)=0:5 (0, 1, 2, ..., 5), 和(s)=5, 期待確率(e)=0.5 として求めた確率密度(d)と累積確率(cp)と上側確率の和(1-cp)を示している。ここでも f が中央の部分(f=2, 3)で最も確率密度(d)が大きくなり(d=0.31)、d の和が 1 であり、cp の到達点が 1 であることが確認できる。

| (34) | 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 |

有意水準を 5% (0.05)と設定すれば、累積確率(cp)が 0.05 より小さい数値 f=0 が「有意に小さい」ということになる。この累積確率は次のように「二項検定」(binomial test)の P 値(p-value)になる。

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

「有意に大きい」数値を見つけるときは注意が必要である。上の表によれば f=4 のとき, cp=0.97 なので, その補数(up)は 0.03 になる。このことから、f=4 を「有意に大きい」数値とすることはできない。なぜならば、f=4 のときの累積確率の補数(up=0.03)は f=4, 5 の確率密度の和(0.16+0.03=0.19)ではなく、f=5 だけの確率密度(0.03)を示しているからである。f=4, 5 の確率密度の和(0.19)は f=3 の up の列にある。その理由は、累積確率(cp の列)が、f=(0, 1, ..., 4)の確率密度の和となり、f=4 の確率密度が含まれているからである。このとき、f=4 の確率密度を含めなければ(f=0, 1, 2, 3)、補数(up)は正しく f=4, 5 の確率密度の和となる。よって、「有意に大きい」数値を見つけるときの累積確率は f を f-1 とした累積確率 pbinom(f-1, s, p)を使用しなければならない。実際に「二項検定」(binomial test)の P 値(p-value)は次のようになる。

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

本研究では個々の頻度の有意性を二項検定の P 値を使って二項対立的に「有意である」と「有意でない」を判定するのではなく、頻度の有意性を連続的に計測する。そこで累積確率の補数ではなく、全体の下側 ( $p = \text{pbinom}(f, t, e) \leq 0.5$ ) では累積確率そのものを使い ( $\text{pbinom}(f, t, e)$ ), 上側 ( $p \geq 0.5$ ) では  $\text{pbinom}(f-1, t, e)$  とした「調整累積確率」(ap: *adjusted cumulative probability*) を使用する。次の ap がその出力である。

|      |   |   |     |         |         |         |
|------|---|---|-----|---------|---------|---------|
| (35) | 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 |

上の出力によれば、 $f=0:2$  までは累積確率(cp)は調整累積確率(ap)と一致し、 $f=3:5$  で  $f=2:4$  の累積確率(cp)と一致している。中央の  $f=2:3$  で同じ数値となり、その前後で同じ傾斜で下降する。その結果  $f=0$  の ap (0.03) と  $f=5$  の ap (0.97) を足すと 1 となり、 $f=1$  の ap (0.19) と  $f=4$  の ap (0.81) を足すと 1 となり、 $f=2$  の ap (0.50) と  $f=3$  の ap (0.50) を足すと 1 となるので、合理的である。ap が 0.05 より小さいときは「f は有意に小さく」、ap が 0.01 より小さいとき「f はとても有意に小さい」。ap が 0.95 より大きいとき「f は有意に大きく」、ap が 0.99 より大きいとき「f はとても有意に大きい」。

## 応用

本研究では、数値の集合の中のそれぞれの数値の有意性を知るために、調整累積確率(二項分布)を使用する。次が〈活用形-代名詞〉の地域:年代の正規化頻度を集中化した行列(Cc)である。

|      |    |       |       |       |       |       |       |       |
|------|----|-------|-------|-------|-------|-------|-------|-------|
| (36) | *  | 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     |

次は行列 Cc の各数値の有意度(調整累積確率)を示す行列(P)である。

|      |    |       |       |       |       |       |       |       |
|------|----|-------|-------|-------|-------|-------|-------|-------|
| (37) | *  | 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 |

たとえば、CV:a1400 の有意度 0.017 は次のように計算される。

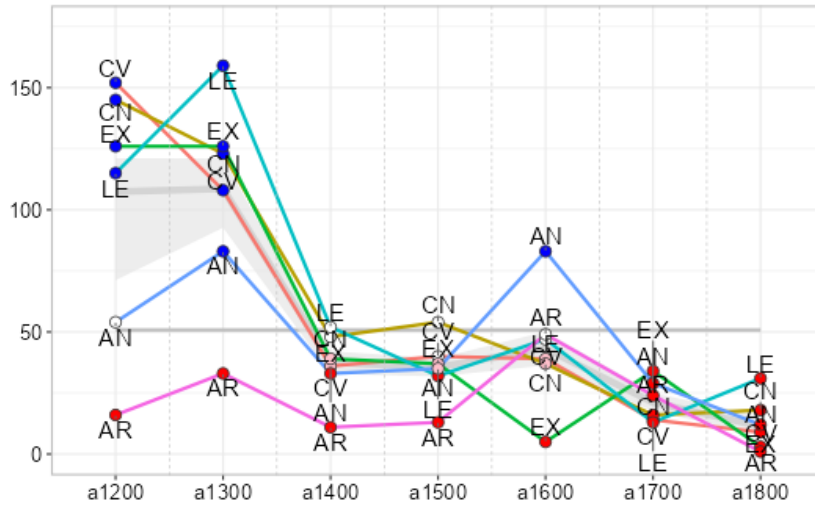
```
(38) BinP=function(f, t, e=0.5) {
      p=pbinom(f, t, e); ifelse(p<=.5, p, pbinom(f-1, t, e))
    }
E=L$cc
s=sum(E); s # 2134
nr=nrow(E); nr # 6
nc=ncol(E); nc # 7
e=1/(nr*nc); e # 0.02380952
BinP(E[1, 3], s, e) # 0.01725561
```

ここでは、有意度を求める関数 **BinP** を用意し、次に行列 **E** の和( $s=2136$ )を求め、**E** の行数 ( $nr=6$ ) と **E** の列数 ( $nc=7$ ) を求める。期待確率 (**e**) は  $e=1/(nr*nc)=0.02380952$  とする。この期待確率(**e**)は、それぞれのセルに分配される確率を示す。関数 **BinP** は頻度(**f**)、和(**s**)、期待確率(**e**)を受けて、有意度 (0.5452151) を返す。

次の図のポイントの青→薄青→白→薄赤→赤のグラデーションは有意度 (**s: significance**)の大きさを次のように反映している。

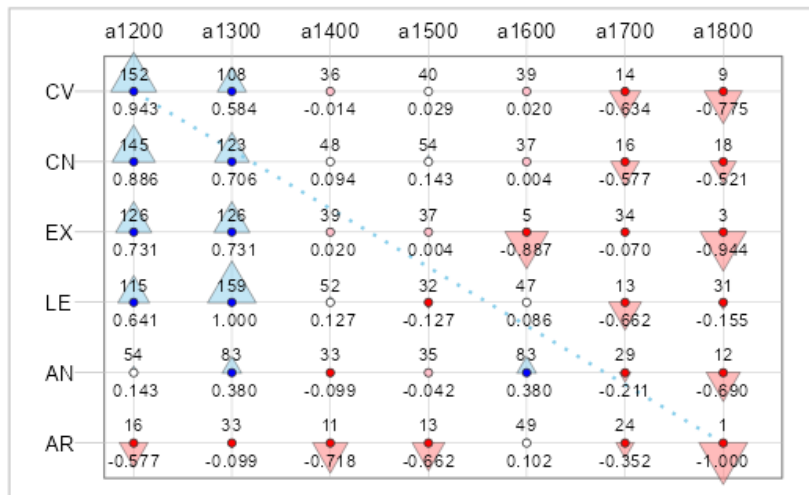
- 1.青:  $s > 0.99$  (有意に非常に大きい)
- 2.薄青:  $0.95 < s \leq 0.99$  (有意に大きい)
- 3.白:  $0.05 \leq s \leq 0.95$  (有意でない)
- 4.薄赤:  $0.01 \leq s < 0.05$  (有意に小さい)
- 5.赤:  $s < 0.01$  (有意に非常に小さい)

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



【図-32】 有意度. 地域: 年代. 年代変化

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



【図-33】 有意度. 地域: 年代. 集中分布

## 6.7. 均衡分析

一部の大きなグループの影響を除くために「比率の平均値」を使って全体の平均値を計算する方法を「均衡分析」と呼ぶ。はじめに「比率の平均値」を説明し、次に「比率の平均値」を使った「単純均衡分析」を説明する。最後に2つの変数を分析する「二重均衡分析」を扱う。

### 6.7.1. 正規化頻度

次は、年代ごとの *enclisis (tiene-lo)* の実測値(O), 総語数(S), 正規化頻度(N)を示す。

|      |   |         |         |         |         |         |         |         |
|------|---|---------|---------|---------|---------|---------|---------|---------|
| (39) | * | 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      |

たとえば、 $N:a1200 = 279$  (正規化頻度)は次のように計算されている。

$$N=O / S * 10^5 = 501 / 179393 * 10^5 = 279.2751$$

実測値(絶対頻度:O)は、その母数(総語数:S)が異なるので比較できないが、正規化頻度(N)は、それぞれの総語数(S)で割り  $10^5$  を掛けてあるので、比較が可能である。このことは、母数が異なる絶対頻度は比較できないが、パーセント(それぞれの母数で割り 100 を掛ける)にすると比較できることと同様である。

次に、クロス集計表の実測値、総語数、正規化頻度を計算する。

次が *enclisis (tiene-lo)* の地域と年代の実測値のクロス集計表(O)である。

|      |    |       |       |       |       |       |       |       |       |
|------|----|-------|-------|-------|-------|-------|-------|-------|-------|
| (40) | *  | 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     |       |

次は総語数の地域と年代のクロス表(S)である。

|      |    |        |        |        |         |         |        |        |       |
|------|----|--------|--------|--------|---------|---------|--------|--------|-------|
| (41) | *  | 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 |       |

そして、次は上の 2 つの表から計算した正規化頻度のクロス集計表(N)である。

|      |    |       |       |       |       |       |       |       |       |
|------|----|-------|-------|-------|-------|-------|-------|-------|-------|
| (42) | *  | 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    |       |

たとえば、AN:a1200 の正規化頻度(=272)は次のようにして求められる:

$$\begin{aligned} \text{正規化頻度(N)} &= \text{実測値(O)} / \text{総語数(S)} * \text{乗数(m: } 10^5) \\ &= 272 / 22429 * 10^5 = 1211.9693 \end{aligned}$$

### 6.7.2. 比率の平均値

たとえば、100 人(xb)のグループ(X)の中で 30 人(x)が「(ある語形を)使用する」と答え、200 人のグループ(Y)の中で 40 人が「(同じ語形を)使用する」と答えた、とする。次はそれぞれのグループの中での使用率を示す(#a1, #a2)。

|  |
|--|
| (43)   |
| 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) |

ここで、全体の使用率を求める必要があるとき、一般に、#a3 のように、全体の使用者の数(x+y=70)を全体の母数(xb+yb=300)で割る、という方法が勧められている。

$$(x+y) / (xb+yb) = 0.233 \text{ (#a3)}$$

これを「全比率」と呼ぶことにする。一方、次のような「比率の平均」を使ってはいけない、と言われている。

$$(rx+ry)/2 = 0.250 \text{ (#a4)}$$

なぜなら、グループを合体した全体で計算すれば、#a3 のように 0.233 になり、#a4 のような 0.250 にはならないからである。しかし、全比率(#a3)の 0.233 は Y の比率(0.20)に近く、X の比率(0.30)からは離れていることに注意したい。

たしかに、全体の比率としては、#a3 の方法が正しいのだが、実際のデー

タ分析では、大きなグループの影響を除いて、小さなグループも大きなグループも均等に扱う「比率の平均」(#a4)のほうが、全体の分布を代表した比率である、と言える。たとえば、X(全数:100)が「男性のグループ」であり、Y(全数:200)が「女性のグループ」であるとすれば、男性と女性を平等に扱う方法として、#a4 が妥当である(比率の平均)。市場全体で見た特定の商品の購買率を計算するときは、#a3 の方式(全比率)を採らなければならないが、「性別による偏りのない、語形の使用率」を計算するときは、#a4 の方法(比率の平均)でなければならない。なぜならば、#a3 の方式(全比率)では、女性(Y)の使用率の影響を強く受けているので、男性(X)と女性(Y)の使用率を均等に扱っている、とは言えないからである。この場合、全比率は「一般的な語形の使用率」というより、むしろ「Y(女性)における語形の使用率」に偏った数値を示していることになる。

このように母数が異なるときの分析は注意を要する。一方、次のようにそれぞれのグループの母数(xb, yb)が等しいときは、「全平均」(#c3)と「比率の平均」(#c4)は等しくなるので問題は生じない<sup>31</sup>。

|      |  |
|------|--|
| (44) | x=30; xb=100; rx=x/xb; rx #c1 0.30 (X) |
|      | y=20; yb=100; ry=y/yb; ry #c2 0.20 (Y) |
|      | (x+y)/(xb+yb) #c3 0.25                 |
|      | (rx+ry)/2 #c4 0.25                     |

### 6.7.3. 単純均衡分析

次は、年代ごとの *enclisis (tiene-lo)* の実測値(O), 総語数(S), 正規化頻度(N:→6.7.1)を示す。

|      |   |         |         |         |         |         |         |         |
|------|---|---------|---------|---------|---------|---------|---------|---------|
| (45) | * | 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      |

次は、正規化頻度の地域と年代のクロス表の下に、それぞれの列の大数平均値(M.mean→6.4)を加えた表である。

|      |    |       |       |       |       |       |       |       |     |   |
|------|----|-------|-------|-------|-------|-------|-------|-------|-----|---|
| (46) | *  | a1200 | a1300 | a1400 | a1500 | a1600 | a1700 | a1800 | ... | N |
|      | AN | 272   | 93    | 76    | 77    | 136   | 87    | 50    |     |   |
|      | AR | 29    | 50    | 46    | 33    | 164   | 89    | 10    |     |   |

<sup>31</sup> たとえば García Hoz (1953)や Juilland and Chang-Rodríguez (1964)の語彙頻度辞典は、それぞれの資料体の分野の語数を 10 万語に揃えているので、分野ごとの語彙の使用率も、全体の使用率も問題なくそのまま計算できる。

|         |     |     |     |     |     |     |    |
|---------|-----|-----|-----|-----|-----|-----|----|
| 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 |
| M. mean | 261 | 169 | 95  | 91  | 99  | 97  | 26 |

上の表を見ると、たとえば a.1200 年代の大数平均値(M.mean=261)は、それぞれの地域・年代で同じ方法で計算された正規化頻度の大数平均値なので、全地域を均等に評価して得られた数値になる。一方、こうした地域差を考慮しないで、年代だけの範囲で正規化頻度を計算すると、その数値は年代の特徴というよりも、高頻度の地域の特徴を主に示していることになる(出力:41)。実際に、地域差を無視した各年代の平均値調整正規化頻度(出力:\*\*の A)と、地域差を同等に考慮した平均値調整大数平均値(出力:\*の A)を比べると、その結果はかなり異なっている。このように、独立変数による正規化頻度ではなく、当該の変数(「対象変数」 *object variable* と呼ぶ)を他の変数(「参照変数」 *reference variable* と呼ぶ)とクロスさせた正規化頻度分布の大数平均値を求める方法「均衡分析」(*Equilibrated analysis*)と呼ぶことにする。

次に、複数の参照変数(地域と社会)を使う均衡分析の方法を説明する。次は列を年代とし(例:a1200)、行を地域+社会(例 AN:C)としたクロス表である<sup>32</sup>。

|      |      |       |       |       |       |       |       |       |
|------|------|-------|-------|-------|-------|-------|-------|-------|
| (47) | *    | 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    |

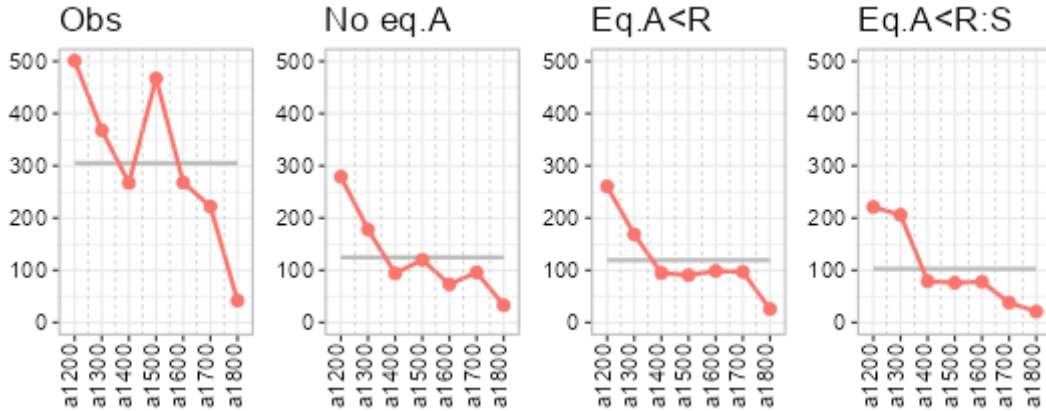
<sup>32</sup> アスタリスク(\*)は欠損値を示す。たとえば、AN:C:a1600 にはデータが存在しない。すべての列で欠損値が見られるが大数平均値(M.mean)の計算では、欠損値を除外したベクトルで計算している。

|         |     |     |     |     |     |     |     |
|---------|-----|-----|-----|-----|-----|-----|-----|
| 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  |

この出力の行名は地域変数と社会変数が組み合わせてある。たとえば、AN:C は AN (Andalucía) + C (documento cancelleresco)を示す。最後の行の大数平均値(*M.mean*)は、対象変数を年代とし、参照変数を地域+社会として計算した結果であり、これが均衡分析の出力となる。

次は、(1)*enclisis* の実測値、(2)年代を独立変数とした正規化頻度(No eq. [A])、(3)年代を目的変数とし地域を参照変数とした正規化頻度(No eq. [A<R])、(4)年代を目的変数とし地域+社会を参照変数とした正規化頻度(No eq. [A<R:S])を比較した表とグラフである。最終列にはそれぞれの行の平均値を添えてある。

| (48)      | * | 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  |



【図-34a、b、c、d】 平均値調整なし

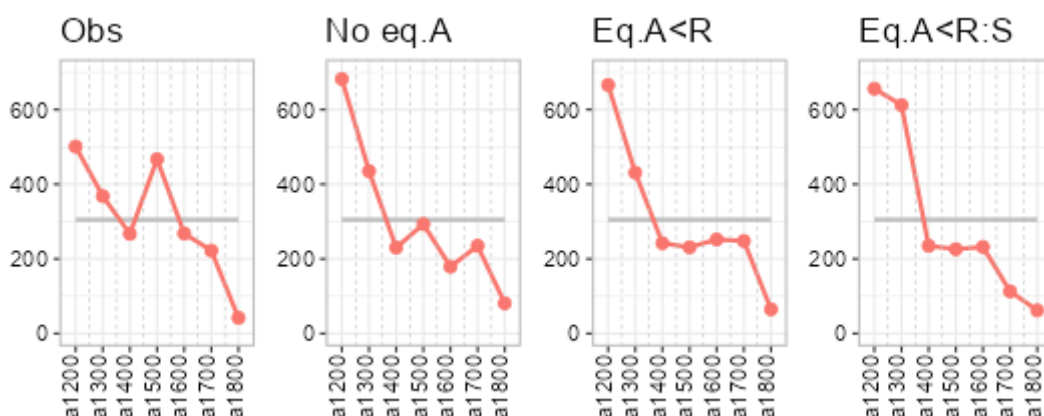
[a.実測値(Obs), b.均衡分析なし(No eq.A)、c.均衡分析(Eq.A<R)、d.均衡分析 (Eq.A<R:S)]

このように、均衡分析をしない分析(No eq. A)の頻度は正規化頻度そのものを使うので数値が比較的高く、参照変数を R とした場合は(Eq. A<R)、地域ごとの正規化頻度を使うので数値が比較的低くなる。最後に、参照変数を R:S とした場合は(Eq. A<R:S)、地域と社会の組み合わせの正規化頻度を使うのでさらに数値が低くなる。

よって、正規化頻度を使った均衡分析の結果をそのまま提示するのではなく、すべての正規化頻度の平均値を実測値の平均値と等しくなるように調整すると次のようになる(→6.2.平均値調整頻度, 下表の *mean* の列はすべて等しい)。

| (49)      | * | 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  |





【図-35a、b、c、d】 平均値調整頻度

[a.実測値(Obs), b.均衡分析なし(No eq.A)、c.均衡分析(Eq.A<R)、d.均衡分析 (Eq.A<R:S)]

上の実測値(Obs.)を除く3つの図はどれも大きな全体的傾向として下降線を示しているが、詳細はかなり異なっている。とくに均衡分析をしない最初のグラフ(No eq. A)の形状と規模が均衡分析をした後2者(A<R, A<R:S)と大きく異なる。

グラフの形状については、単数参照変数(地域)を用いた正規化頻度の平均値はa1200とa1300の間で大きな下降が見られるが(Eq.A<R)、複数参照変数(地域・社会)の場合では、両年代間の推移は小さく中世前期の比較的安定した状態を示している。最も大きな均衡性は3番目の複数参照変数(地域・社会)の場合に示されるので、その分析が妥当な結論になる(Eq.A<R:S)。

#### 6.7.4. 二重均衡分析

ここでは〈活用形-代名詞〉(*tiene-lo*)の対象変数(複合)を地域(R)+年代(A)とし、参照変数(単一)を社会(S)とした「二重均衡分析」(*Double equilibrated analysis*)を行う(A:R<S)。はじめに、比較のために、均衡化をしない従来の方法で地域(R)と年代(A)の単純なクロス集計表(実測値: O)を作成する(R:A)。

| (50) | * | 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     |     |   |

そして、次は正規化頻度のクロス集計表である。

| (51) | *   | 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    |       |       |

一方、ここで扱う二重均衡分析では、はじめに次のような地域：年代を行とし、社会を列とした正規化頻度のクロス表＋大数平均値(*M.mean*)を用意する。

| (52)     | *   | 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  |

この表のそれぞれの行の大数平均値(*M.mean*)が参照変数の社会変数(S)によって均衡化された地域+年代の変数の平均値である。このとき、それぞれの行は地域と年代が複合対象変数となる。たとえば、第1行の AN:a1200 は地域(AN: Andalucía)と年代(a1200)で構成されている。次に、この複合対象変数を地域と社会に分解した以下のクロス表を作成する。

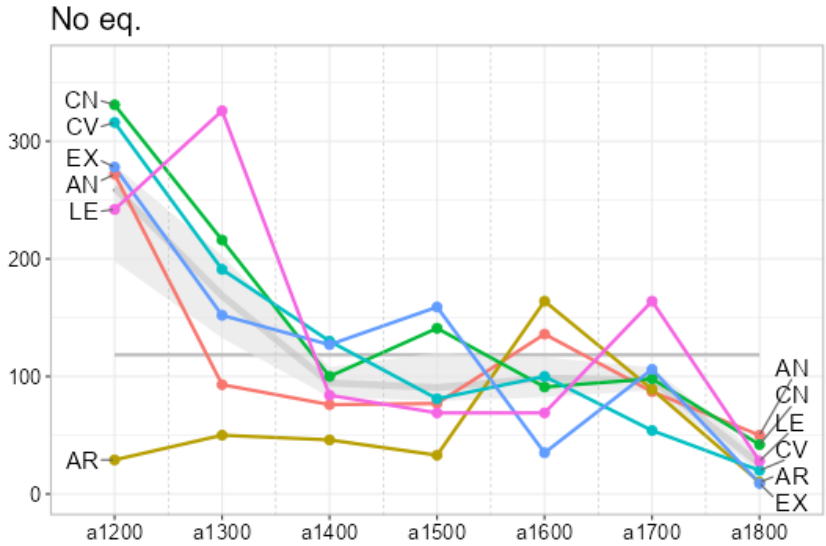
|      |     |       |       |       |       |       |       |       |
|------|-----|-------|-------|-------|-------|-------|-------|-------|
| (53) | *   | a1200 | a1300 | a1400 | a1500 | a1600 | a1700 | a1800 |
| AN   | 115 | 177   | 70    | 76    | 176   | 62    | 26    |       |
| AR   | 34  | 71    | 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    |       |

上の表の中の、たとえば AN:a1200 (=115)は先の(48)の第1行の大数平均値(*M.mean*)である。そのRとAを分解した(49)が二重均衡分析の結果として表示される。

しかし、上の表の数値のスケールはデータの実測値(絶対頻度)と一致しないので、次に実測値の平均値によって調整すると(→6.2)、次の表が得られる。この数値は実測値のスケールと一致するので、現実的な「頻度」を示している、と言える。よって、均衡分析ではこの「平均値調整正規化頻度」を使用する。

| (54) | *  | 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    |

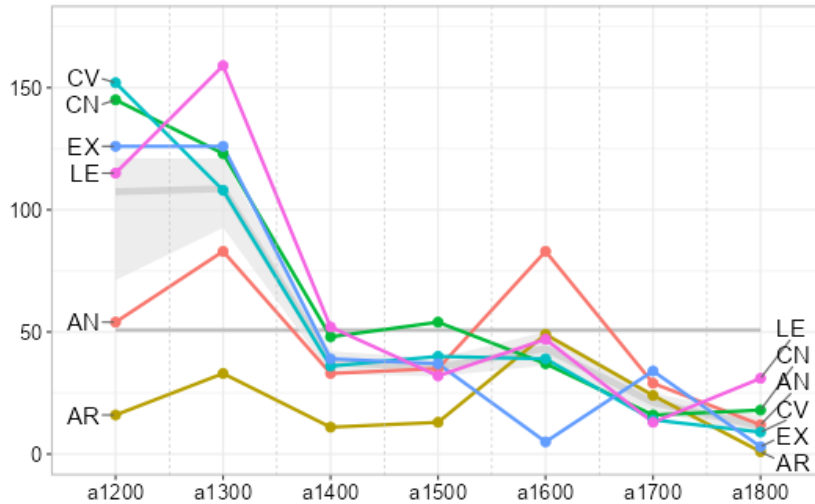
次の 2 つのグラフはそれぞれ均衡化をしないクロス表と、均衡化したクロス表から作成されたものである。薄い灰色の面の中央に引かれた灰色の太線はそれぞれの年代における地域の正規化頻度の大多数平均値の位置を示し、薄い灰色の面は上側平均偏差と下側平均偏差に囲まれた領域を示す<sup>33</sup>。これらによって中央の半数のデータの傾向がわかる。グラフの中の灰色の水平線はデータ全体の平均値を示す。



【図-36】 均衡分析をしない単純な集計. 地域：年代 R:A.

<sup>33</sup> 全体の傾向線を示すとき、一般に中央値と四分位を使用するが、先述のように(→2.4.大多数平均値)、中央値は全体の中心を適切に示さないので中心線を大多数平均値とした。このとき、中央値を使用しないので、中心の幅として四分位を使用することは整合性が失われるので、上下を区別する平均偏差を使用した。このとき標準偏差を使うと、上下の区別しないので、不正確である。

Eq.R:A&lt;S



【図-37】均衡分析．地域：年代&lt;社会 R:A&lt;S.

2つのグラフは近似しているが、細かく見ると、たとえば AN:a1200などは大きく異なる。先の表(48)で AN:a1200の行を見ると、C=298という高い数値が見つかる。しかし、この AN:a1200の行の大数平均値はわずかに115である。よって、均衡化しないデータの正規化頻度(47)は a1200の状態を示しているのではなく、むしろ C (cancilleresco)の特徴を表している、と考えられる。このように、均衡化されていないデータには社会変数の偏りがあるので不正確である。よって、分析は社会を参照した均衡化データで行うべきである。

## 6.8. まとめ

本研究の統計的方法で扱ったことは数量的変異を示すクロス集計表に一般に見られる問題である。私たちは数量的変異・特異性があるデータから、分布の特徴を探るのであるが、そのような特異な変異があるデータの分布には歪み・偏りがある、ということが問題である。そのとき、データ全体の分布は一部の特異な数値に左右され、全体の数値の分布を適切に表示していない。頻度分布の特異性を探ることによって分布の特徴を探るとき、その頻度分布が特異であることが支障になる、ということでは大きな(本質的な)矛盾になる。本研究ではこの問題に対して各種の分析法を使って解決する方法を提示した。

統計数理研究所が1953年から5年ごとに継続している「日本人の国民性調査」では、面接の対象者を抽出するために「層化多段抽出法」というサ

ンプリング方法が用いられている。このサンプリング法の具体的な手順は次のように説明されている<sup>34</sup>。

全国の市区町村を、区部・人口 20 万人以上の市部・人口 10 万人以上の市部・人口 10 万人未満の市部・郡部・沖縄県の 6 つに層化しました。次に、各層から合計 400 町丁字等を確率比例抽出で選びました。最後に、抽出した町丁字等の住民基本台帳から、その地点に割り当てた人数（平均 16）の標本を等間隔抽出法で選びました。

ここで規模が異なる 6 つの層からさらに小さな単位の町丁字等が選ばれている。このような方法を採用すれば、一般のランダム化サンプリングで起こりやすい大都市による偏りが防げるはずである。しかし、私たちの歴史資料では現存するデータしか入手できず、アンケート調査のようにサンプルをコントロールすることは困難である。そこで、人口規模を勘案してその偏りを排除した層化多段抽出法のように、規模の異なる地域や社会文体ごとに計算した正規化頻度を平均化することで地域と社会文体の偏りを回避した。

このように、すべての頻度と母数から計算される正規化頻度の過半数平均は、特定の地域や社会に影響を受ける値ではなく、すべての地域や社会が同じ基準と方法で均等に扱われた値である。この「均衡分析」と名付けた方法は次のような様々な方法と連携している。そこでは外れ値の影響を受ける平均値（平均値）ではなく、外れ値を含めてその影響を軽減した「大数平均値」を使った。そして、実際の実測値の平均によって調整された正規化頻度の数表を、比較の可能性を失うことなく、実測値と同じ規模で評価した（「平均値調整頻度」）。さらに、中央値からの偏差を含むグラフにより、全体的な傾向を観察した（「中央値偏差」）。対角線上に集中した表により、同一の概念、たとえば〈伝統と革新〉を使って、表内の行、列、および頻度分布の間の一貫した動きを観察した（「集中分析」）。集中分析で得られた行と列の距離係数は「間隔クラスター分析」によってその連続的階層構造に位置づけた。グラフ上のドットを「有意度」によっての色分けした。グラフは、動的な変化を観察するための折れ線図と、静的な変異を観察するためのプロット図を用意した。

---

<sup>34</sup> <https://www.ism.ac.jp/kokuminsei/page9/page10/index.html>