

東京スペイン語学会 2023/10/28

スペイン語計量文体論の問題と方法

語長・語形成・語彙

Inmaculada Martínez

上田博人

2023

はじめに

私たちは 2021 年に Santander, España の自由会話の録音資料の語彙目録と形態素目録を作成しました(Martínez y Ueda 2021, 2023)。これらの目録は単語の使用頻度を列挙したのですが、これには様々な用途がありますが、その一つとして使用語彙から話し言葉の文体的特徴を明らかにすることが考えられます。

話し言葉の特徴は書き言葉と比較することで明確になります。そこで、先行研究による資料(Juilland and Chang-Rodríguez, 1964 ; Justicia 1995)を用いて各種の出版物と小学生の自由作文での使用語彙と比較します。私たちの資料を含めて全体の資料が扱う語の数は 111 万語以上になりました。このような多数の資料が示す数値はかなりの信頼性があると考えられます。

この研究は資料(先行研究+私たちの研究)の性質から語と、語の構成要素(接頭辞・接尾辞・合成辞)を単位としています。当然、文体的特徴は語だけでなく、文法範疇、統語、文、段落、談話などにも認められます。これらすべてを含めて一挙にスペイン語計量文体論を構築することは非現実的なので、私たちの研究の第一歩として、ここに語長・語複合語彙を扱うことにしました。

私たちの研究の進展は随時、次のサイトにアップロードいたします。

<https://h-ueda.sakura.ne.jp/kenkyu/index.html>

目次

はじめに	2	3.4. 地域差	52
1. 問題と方法	6	3.5. 年代差	55
1.1. 計量文体論	6	3.6. 文体差	57
1.2. 問題	7	4. 語彙	59
1.3. 資料	7	4.1. 機能語	60
1.4. 機能語と内容語	10	4.2. 内容語	62
1.5. 語形成要素	12	4.3. 単一語	64
1.6. 方法	15	4.4. 複合語	66
2. 語長	17	4.5. 地域差	68
2.1. 序論	17	4.6. 年代差	71
2.2. 短縮の法則	19	4.7. 文体差	75
2.2.1. 先行研究	19	4.8. 文体・年代・地域	77
2.2.2. 辞書の見出し語	20	5. 結論	80
2.2.3. 使用頻度	23	6. プログラム-R	82
2.2.4. 先行研究	31	6.1. 相対標準偏差	82
2.2.5. 内容語	31	6.2. 規定距離	83
2.2.6. 複合語	33	6.3. 単調性	83
2.3. まとめ	36	6.4. L字型分布	83
3. 語形成	40	6.5. 集中分析	84
3.1. 接頭辞	40	7. 参考文献	84
3.2. 接尾辞	44		
3.3. 合成語	51		

要旨

本研究ではスペイン語の語彙の形態・機能・意味の問題を語長・語種・変異の観点から分析し考察する。資料は異なる性質をもつ4種(Juillard-Chang-Rodríguez 1964, Justicia 1995, Davies 2006, Martínez y Ueda 2021, 2023)を使う。これらによって同質の資料が示す偏りを避ける。

一般に言われる「短縮の法則」(「使用頻度が大きくなると語長は短縮する」)が主張する使用頻度と語長の逆関係は、語彙を機能語(冠詞・代名詞・前置詞・接続詞など)と内容語(名詞・形容詞・副詞・動詞など)に峻別することによって明らかになる。機能語は語彙的意味を表示せず、文法的機能を果たすので短い弱勢語になることが多い。特に非常に高頻度の定冠詞・前置詞・接続詞はすべて短い弱勢語である。一方、内容語の意味表示には一定の音形が必要である。この機能語と内容語の対照的な違いが短縮の法則の要因である。よって、短縮の法則は機能語と内容語の文法的・語彙的対立の結果であって、その要因ではない。つまり、高い使用頻度が語形を短縮するのではなく、逆に、語形が短い機能語・単一語が高頻度で使用されるのである。

表面的な音形の増加は本質的に加重される形態素(接頭辞・接尾辞)の増加によるものである。そこで形態素が加重されることから有標性を帯び、その結果頻度が減少する。有標的形態素は各種の変数に従って様々な変異を生む。それらの変異は計量文体論の研究で定量的・定質的に分析される。ここでは資料が示す学年・出版物・性・年齢・学歴という変数に依存する文体特徴を探る。そして、言語地図に現れる「地域」という変数も一部で観察する。

Resumen

En este estudio, vamos a analizar y discutir cuestiones de forma, función y significado del vocabulario español desde los puntos de vista de la longitud de las palabras, los tipos de palabras y sus variaciones. Vamos a utilizar ocho materiales con diferentes propiedades: 1. García Hoz (1953), 2. Juilland and Chang-Rodríguez (1964), 3. Ueda (1987), 4. Justicia (1995), 5. Moreno et al. (2005), 6. Davies (2006), 7. Martínez y Ueda (2021a, 2013), 8. Real Academia Española (2023b).

En la relación inversa entre frecuencia de uso y longitud de la palabra, que generalmente defiende la "ley de acortamiento" ("a medida que aumenta la frecuencia de uso de la palabra, disminuye su longitud") apoyada por la "ley de menor esfuerzo", hay que ver la diferencia entre las palabras funcionales (artículos, pronombres, preposiciones, conjunciones, etc.) y las palabras de contenido (sustantivos, adjetivos, adverbios, verbos, etc.). Las palabras funcionales no poseen un significado léxico y, en su lugar, cumplen una función gramatical, por lo que suelen ser palabras cortas y débiles por no presentar un significado dotado de una forma distintiva. En particular, los artículos definidos, las preposiciones y las conjunciones, muy frecuentes, son palabras átonas bastante cortas. Por otro lado, se requiere una determinada forma sonora para representar el significado de una palabra de contenido. Esta diferencia contrastante entre palabras funcionales y palabras de contenido es la causa de la ley de acortamiento, que es el resultado, no su causa, de la oposición gramatical y léxica entre palabras funcionales y palabras de contenido.

El aumento de las formas sonoras se debe principalmente al aumento de los morfemas adicionales (prefijos y sufijos). Dado que los morfemas están ponderados, las palabras que los utilizan adquieren carácter distintivo y, como resultado, su frecuencia disminuye, por ser especiales, es decir, por estar marcadas. Los morfemas específicos producen diversas variaciones de acuerdo con diversos factores. Estas variaciones se analizan cuantitativa y cualitativamente en el estudio estilométrico.

Exploraremos características estilísticas que dependen de variables como grado escolar, publicación, sexo, edad y nivel educativo indicadas en los materiales. También observaremos la variable geográfica que se realiza en los mapas lingüísticos. Para todo ello, utilizamos programas elaborados en R y nuestro sistema Lyneal.

1. 問題と方法

1.1. 計量文体論

言語の計量的研究(lingüística cuantitativa)には計量方言学(dialectometría), 計量音韻論(fonometría), 計量語彙論(lexicometría)などのように合成辞-metríaが付された研究分野があり, その一つに計量文体論(estilometría)がある。計量文体論は各種の「文体」を数量の観点から研究する分野であり, その文体とは言語表現(話し言葉と書き言葉)の変異を意味する。言語の体系・構造を対象とする理論的言語学の研究では数量的方法は必須ではないが, 言語の変化・異種を対象とする実践的研究では必須である。ここで扱う計量文体論は数量的方法を用いた言語の変化・異種の研究である。様々な言語で行われているが, 私たちの環境ではとくに日本語と英語についての研究成果を参照することが多く, その方法は参考になる。とくにコーパス言語学の分野にはすでに多くの研究成果がある。スペイン語についてもすぐれた研究がある(Gutiérrez 1978, Kock 1983, 2001; Moreno Fernández 1992, Irizarry 1997, Parodi 2007; Nieto Caballero y San Segundo 2020; Rojo 2021)。

一般に文体論は特定の文学作品や著者の文体特徴が扱われてきた(波多野 1958, 1965, 1988; Molinié 1992 [1994])。個別の作品や著者が対象となるときもあれば(Tonko Milic 1967; 安本 1977; 林四郎 1979; 林巨樹 1979; Hartzfeld 1972; Frago 2015), 複数・多数の作品・著者が比較されることもある(安本 1960, 1965, 1977, 2009; Miles 1964; Williams 1970; 中村 1977; 樺島・寿岳 1979; 齊藤 1992)。さらに文学作品に限らず, 広範囲の書き言葉や話し言葉までを含めて論じられることもある(林 1966; 山本 2014)。前者は文学研究, 後者は言語研究として区別されることがあるが, その境界を取り除いてすべての言語活動を含むことも可能である。

文体論の研究は定質的にも定量的にも行われているが, 「質」の問題と「量」の問題を排他的に選択するのではなく(Hough 1969 [1972]: 82-88; 福井 1978: 376-379), 相補的に総合することも考えられる(Guiraud 1957 [1959]: 122-124; 中村 2016, 161-166)。ここでは, 情報科学・統計学の手法を取り入れた定量的分析を行うが, 分析結果の解釈においては数値・グラフを提示するだけでなく数値・グラフが示している意味を質的に解釈することで定量的研究と定質的研究の総合を目指す。このとき, 両者を同時に行うのではなく, 量的分析→質的解釈という方向と質的仮説→量的検証という 2 つの方向を考える。

1.2. 問題

私たちの計量文体論は次の問題を取り上げる。

- ・ 文体特徴を決定する。
- ・ 文体特徴と変数(書き手・話し手・作品)の関係を観察する。

ここでいう「文体特徴」(rasgos estilísticos)とは、他と区別される特定の資料・書き手・話者の有意な言語的特徴を指す。よって、文体特徴には十分な頻度(有意性)と変異(異種性)がなければならない。統計的には和(または平均)と偏差(相対標準偏差→プログラム-R)を使用する。

文体特徴には、語、形態素(語根・接頭辞・接尾辞)、語形成(単一語・派生語・合成語)、文法的区分(機能語・内容語)、文(平叙文・疑問文・感嘆文；単文・複文)、品詞、屈折、語種(俗語・民衆語・教養語)などが挙げられるが、ここでは語・形態素・語形成・文法的区分(機能語・内容語)を取り上げる。

1.3. 資料

この研究では次の 8 種の資料を用いる：1. García Hoz (1953), 2. Juilland and Chang-Rodríguez (1964), 3. 上田 (1987), 4. Justicia (1995), 5. Moreno et al. (2005), 6. Davies (2006), 7. Martínez y Ueda (2021a, 2013), 8. Real Academia Española (2023b). これらを総合した資料を以下では「8 種資料」と呼ぶ。

1. García Hoz (1953 ; GH とする)は 1940 年代のスペイン人の言語生活を次の 4 つの分野に分け、それらに対応する言語資料をそれぞれ 10 万語を収集した。

- (a) 家族生活 … 個人の手紙
- (b) 一般的社会生活 … 新聞
- (c) 規範的社会生活 … 公文書
- (d) 文化的生活 … 書籍

2. Juilland and Chang-Rodríguez (1964: JCh)は 20 世紀前半にスペインで刊行された五種の出版物(演劇作品・小説・随筆・新聞雑誌・科学技術文)をそれぞれ 10 万語ずつ収集し、見出し語と変化語形の使用頻度をまとめたものである。その中から見出し語の頻度を抽出し、それぞれの出版物の使用語彙の頻度分布を計算した。総使用頻度が 5 以下の語は収録されていない。

3. 上田 (1987: Manual) は García Hoz (1953)と Juilland and Chang-Rodríguez (1964)を整理して対照頻度表を作成し、加えてスペイン、フランス、アメリカ合衆国で発行された中級程度のスペイン語教科書・参考書 7 冊(1957-1977)

の使用語彙と比較した。ここでは、その中から教科書・参考書の使用語彙を使用する。

4. Justicia (1995) は東アンダルシア(Almería, Granada, Jaén, Málaga)の小学生(6~13 歳) 3,402 名が 20 分間に書いた自由作文の統計的語彙目録である。私たちは定冠詞・代名詞などの性・数の区別を合算するなど、一部の語彙の集計を変えて、A1(6~7 歳), A2 (8-10 歳), A3 (11~13 歳)という記号を付けた。それぞれの学年(A1, A2, A3)の児童の使用語彙は語彙の発達(desarrollo del vocabulario)を研究するための基礎的な資料とされるが、それぞれの学年を変数とした計量文体論研究にも有用である。

5. Moreno et al. (2005) は 500 人の話者による自然な公式・非公式な発話の録音を転写した資料である。さらに、学術的な講演の転写資料、健康相談の文書を加えた。

6. Davies (2006) はスペイン(43%)とラテンアメリカ(57%)の合わせて 2000 万語の話し言葉・書き言葉・文学・新聞の使用語彙の頻度と拡散を示している。資料は 1970-2000 年の範囲に限定し、とくに 1990 年代のものが多。

私たちのサンプルコーパス 7. PRESSEA-Santander (Martínez y Ueda, 2021, 2023: Santander)は性(2)*年齢(3)*学歴(3)=18 名の録音資料(録音時間 45 分以上)を収録している。その使用語彙は次のように分布している(Martínez y Ueda (2021b) :

(1) サンプルコーパス PRESSEA-Santander の語数

Sexo:	H: hombre			H.total	M: Mujer			M.total	Total
	E1	E2	E3		E1	E2	E3		
Nivel: N1	6 380	6 395	8 160	20 935	6 347	6 971	5 141	18 459	39 394
Nivel: N2	4 250	8 319	7 633	20 202	3 337	10 278	5 459	19 074	39 276
Nivel: N3	7 699	5 437	5 888	19 024	4 343	3 710	9 352	17 405	36 429
Total	18 329	20 151	21 681	60 161	14 027	20 959	19 952	54 938	115 099

H (男性) / M (女性)

E1 (20~34 歳) / E2 (35~54 歳) / E3 (55 歳~)

N1 (初等教育 ~10 歳) / N2 (中等教育 ~16-18 歳) / N3 (高等教育 ~21-22 歳)

この資料は性(H,M)・年齢(E1,E2,E3)・学歴(N1,N2,N3)のすべての組み合わせをカバーし、全体的に偏りのない言語資料を構成している。また、たとえば性差を見るときはそれぞれの性で年齢(3)と学歴(3)を組み合わせた 9 人ずつとなるので、偏りのない比較が可能になる。年齢差・学歴差を見るときも同様にそれぞれ 6 人ずつとなる。

8. Real Academia Española (2023: Colpes) は最新(2023)の最大規模

(395,000,000 語)の資料である。21 世紀(2001~2023)のスペイン(30%)とラテンアメリカ(70%)で発行された書籍・新聞・定期刊行物・インターネットの使用語彙の見出し語・品詞・頻度が示されている。

以上の資料を総合した「8 種資料」は、以下のような割合で目的値を設定し、計 100 万語としてある。たとえば、Colpes は総頻度は数字・句読点を除き、過去分詞形の形容詞を動詞に統合させるなどの調整をすると、計 352,000,000 語になるが、これを 500,000 として、それぞれの見出し語の頻度を 50 万語あたりの相対頻度に変える。その結果、全体の 100 万語の中でその半分(50 万語)が Colpes の頻度になる。以下の資料も同様である。見出し語数は各種資料の相対頻度の和のランク 1~5000 とした。よって、「8 種資料」の全体の見出し語数は 5000 であり、その全体の相対頻度数は 100 万である。

(2) 資料の構成

Datos	Frecuencia	Objetivo	%
CORPES	352,899,385	500,000	50.0%
Davies	17,326,289	250,000	25.0%
GH	387,524	60,000	6.0%
JCh	460,065	60,000	6.0%
Moreno	685,905	60,000	6.0%
Justicia	507,384	30,000	3.0%
Santander	111,118	25,000	2.5%
Manual	98,460	15,000	1.5%
TOTAL	372,476,130	1,000,000	100.0%

それぞれの資料の相対頻度(Frecuencia relativa)はその資料の使用頻度(Frecuencia de uso)と特徴を考慮して決めた。どの相対頻度も使用頻度を超えないようにした。Colpes はそのラテンアメリカを重視し、規模が最大であり、スペイン語の現在の姿を反映しているので、最大の比率(50%)を与えた。Davies の資料はやや古く、規模は Colpes よりかなり小さいので 25%とした。Justicia と JCh の規模は類似し、その性質は大きく異なる(特殊性と一般性)。それぞれを 7.5%とした。GH はかなり古いデータなので 6%とし、Santander の話言葉の録音は貴重であるが、地域が限定され、総頻度は大きくないので 2.5%とした。最後に Manual は特殊な目的のための人工のテキストであり、総頻度も多くないので 1.5%とした。

このように、性質が異なる複数の資料を総合した理由は分析資料の偏りを避けるためである。実際に各種資料の頻度分布を比較すると、多くの見出し語の使用順位は近似するが、かなり異なる場合も少なくない。純粋に 1 つの

データに限った分析からは一般性を保証する結論を導出できず、「…においては」という条件をつけなければならない。その場合は他のデータでの検証が必要になる。一方、現在利用できる資料を多数集めることによって、より一般性のある結論が得られる可能性が高い。もっとも、いくら多数の資料を集めても、その集められた資料は自ずと限界があることは確かであるが、少しでも結論の一般性が高められるはずである。少数の純粋なデータによって得られた特殊な結論は貴重であり興味深い。一方、多数のデータによって得られたより一般性のある結論も重要である。

以下ではこの 8 種資料を使って使用語彙の語長を総合的に調査して考察する。次に、語形成と語彙の使用分布を比較するときには、それぞれのデータを個別に抽出して比較する。

1.4. 機能語と内容語

本研究では機能語と内容語を区別して分析する。Stubbs (2002 [2006]: 53-55)は語彙の総体を内容語(content word)と機能語(functional word)とに大別し、それぞれの特性の違いを次のように説明している。

(3) 内容語：名詞・形容詞・副詞・本動詞

- ・テキストがどういった内容であることを示す。
- ・数が多く、さらに新語を増やすことができる。
- ・語形変化をする。

(4) 機能語：助動詞・法助動詞・代名詞・前置詞・限定詞¹・接続詞

- ・内容語を互いに関連づける。
- ・数は少なく、新語を増やすことは非常にまれである。
- ・語形変化をしない。

私たちは機能語の限定詞に含まれると思われる数詞を内容語に含め、さらに内容語に間投詞を含める。これは数詞と間投詞は内容語を互いに関連付けるというよりも、テキストの内容にかかわる語であり、その語数が多く開かれた集団であるという理由に因る。一方、機能語に疑問詞と関係詞を加える。どちらも内容語の語彙的特性ではなく、機能語の文法的特性を有するためである。

¹ 石橋(1973: s. v. determiner)によれば冠詞・属格・指示形容詞・不定形容詞・数詞など。田中春美(1988: s. v. determiner)によれば冠詞・所有詞・指示詞・数量詞(*both, any, some, many, much, more, few, every, each, no* など)。Lindquist (2009 [2016]: 39 注)は *which, what* を疑問限定詞としている。

そして、内容語の第一の特性はむしろ話題語のものであるので、代わりに内容語の第一の特性を「独立した意味・情報をもつ」とする。さらに、内容語の数が多く機能語の数が少ないというときの「数」は異なり語数を指しているので、述べ語数(使用頻度)については逆に「個々の内容語の使用頻度は低く、個々の機能語の使用頻度は高い」をそれぞれの特性として加える。さらに、内容語の語形は語形成によって拡張した複合語が多いが、機能語では複合語は少なく、弱勢の単一形態素語の使用頻度が高いことも挙げる。

このように機能語と内容語はまったく対照的に異なる特性をもつので、一般に語彙の分析は両者を区別して行わなければならない。

次が 8 種資料の機能語とその使用頻度のリストである^{2,3} :

- (5) アルファベット順: a.prep, a.prep, algo.ind, alguien.ind, alguno.ind, ambos.ind, ante.prep, aquel.dem, aunque.conj, bajo.prep, bastante.ind, cada.ind, cierto.ind, como.conj, con.prep, contra.prep, cual.rel, cualquiera.ind, cuando.conj, cuanto.rel, cuyo.rel, cuál.interrog, cuándo.interrog, cuánto.interrog, cómo.interrog, de.prep, demasiado.ind, demás.ind, desde.prep, donde.conj, durante.prep, dónde.interrog, el.art, él.pro, en.prep, entre.prep, ese.dem, estar.v, este.dem, excepto.prep, haber.v, hacia.prep, hasta.prep, lo.pro, mas.conj, me.pro, mediante.prep, menos.ind, mi.pos, mientras.conj, mismo.ind, mucho.ind, más.ind, mí.pro, nada.ind, nadie.ind, ni.adv, ninguno.ind, no.sn, nos.pro, nosotros.pro, nuestro.pos, o.conj, os.pro, otro.ind, para.prep, pero.conj, poco.ind, por.prep, porque.conj, pues.conj, que.conj, quien.rel, quién.interrog, qué.interrog, se.pro, según.prep, ser.v, si.conj, sin.prep, sino.conj, sobre.prep, su.pos, sí.pro, sí.sn, tal.ind, también.adv, tampoco.adv, tanto.ind, te.pro, ti.pro, todo.ind, tras.prep, tu.pos, tú.pro, un.art, uno.ind, usted.pro, vario.ind, vos.pro, vosotros.pro, vuestro.pos, y.conj, y/o.conj, yo.pro.

次のリストは上のリストの使用頻度を降順で並べ替えた結果である :

- (6) 使用頻度降順: el.art 使用頻度:118997, de.prep 76116, que.conj 36827, y.conj 32648, a.prep 30109, en.prep 28910, un.art 22875, ser.v 21015, se.pro 17597, no.sn 13365, con.prep 11419, su.pos 10826, por.prep 10656, lo.pro 10348, para.prep 7918, haber.v 7382, este.dem 6284, estar.v 5950, como.conj 5796, más.ind 5307, me.pro 5206, todo.ind 4643, ese.dem 4472, o.conj 4305, pero.conj 4297, él.pro 3541, otro.ind 3249, si.conj 2857, mi.pos 2806, yo.pro 2631, mucho.ind 2257, porque.conj 2163, qué.interrog 2067, sin.prep 2042, te.pro 2017, sí.sn 2001, cuando.conj 1996, también.adv 1955, nos.pro 1873, sobre.prep 1853, uno.ind 1748, entre.prep

² 内容語のリストは膨大になるのでここで扱うことはできない(→Martínez y Ueda 2023)。

³ 品詞 : a: adjetivo, adv: adverbio, art: artículo, cif: cifra, conj: conjunción, dem: demostrativo, extr: extranjerismo, ind: indefinido, interj: interjección, interrog: interrogación, n: nombre, num: numeral, pos: posesivo, prep: preposición, pro: pronombre, prop: nombre propio, rel: reletivo, sn: *sí-no*, v: verbo.

1712, alguno.ind 1622, mismo.ind 1591, tanto.ind 1559, pues.conj 1452, hasta.prep 1397, desde.prep 1396, nuestro.pos 1253, tu.pos 1243, poco.ind 1177, ni.adv 1123, nada.ind 1013, cada.ind 964, aquel.dem 953, usted.pro 913, donde.conj 884, algo.ind 866, menos.ind 839, cómo.interrog 824, aunque.conj 772, quien.rel 728, mí.pro 702, durante.prep 697, sino.conj 673, tal.ind 631, tú.pro 576, contra.prep 576, hacia.prep 572, mientras.conj 564, según.prep 551, nosotros.pro 521, cierto.ind 517, ninguno.ind 514, cual.rel 504, ante.prep 485, cualquiera.ind 458, vario.ind 454, nadie.ind 391, dónde.interrog 342, cuyo.rel 342, tras.prep 340, cuanto.rel 327, quién.interrog 314, tampoco.adv 273, bastante.ind 261, cuándo.interrog 259, demás.ind 254, cuál.interrog 246, ambos.ind 239, alguien.ind 213, demasiado.ind 211, bajo.prep 200, os.pro 182, ti.pro 177, cuánto.interrog 168, mediante.prep 152, vuestro.pos 85, vos.pro 71, vosotros.pro 66, mas.conj 56, y/o.conj 17, sí.pro 14, excepto.prep 11.

このように弱勢語で語長が短い語の頻度が高く、逆に強勢語で語長が長い語の頻度は少ない。この理由を以下で考察する(→2.2)。

1.5. 語形成要素

ここで扱う語形成要素は基本的に接頭辞・接尾辞であり、それに合成語を加えた。接頭辞・接尾辞については、言語の理論的な分析とは別に、次のような実際的な分析を採用した。たとえば Hualde, Olarrea y Escobar (2001: 202-203)は *increíble* を [[in.creí].ble]ではなく [in.[creí.ble]]と分析する⁴。その根拠として *creíble* は存在するが **increer* は存在していないことを挙げている。よって、**increer* > *increíble* という派生ではなく、*creíble* > *increíble* という派生を考えることになる。しかし、Juilland and Chang-Rodríguez (1964)の頻度辞典によれば *creer* (使用頻度: 567), *increíble* (10)は収録されているが *creíble* は収録されていない(DLE には載せられている)。よって、実際に? *creíble* > *increíble* という派生は考えにくい。そこで、派生を *creer* > *increíble* [in.[creí].ble]として、*in...ble* という「接周辞」(circunfijo)を考えることができる(「接周辞」: Whaley 1997 [2006]: 123)。接周辞は *anochecer*, *enriquecer* などの *formas parasintéticas* (García de Diego 1970: 287-288; Serrano-Dolader, 1999: 4701-4730; Real Academia Española y Asociación de Academias de la Lengua Española, 2009: 605-615)に相当する (*a-noch(e)-.ecer*, *en-riqu[rico]-ecer*)⁵。DLEによれば *increíble* の語義は"Que no puede creerse; Muy difícil de

⁴ García-Page Sánchez (2011: 85, 143)も *inadmisible*, *insoportable* を同様に分析している。

⁵ DLE: s.v. *parasíntesis*, Del gr. παρασύνθεσις parasynthesis.
1. f. Gram. Proceso de formación de palabras en que intervienen simultáneamente la composición y la derivación; p. ej., pordiosero, picapedrero.

creer."であり、*Que no es creíble ではない。このことも接頭辞 *in-*ではなく接周辞 *in...ble* の設定を支持する。この接周辞派生が正しいならば、計量文体論において、*increíble* は *creíble* と比較するのではなく、*creer* と比較して分析しなければならない。資料全体を見ると(「8種資料」), *in-(negativo) ... -ble* は 18 例見つかる。その中で *in-*がない語形の方が頻度が高いのは次の 3 例である : *imposible.adj* (頻度 173:*in-* 頻度 77), *invisible.adj* (37:*in-*28), *irresponsable.adj* (9:*in-*.5). 一方、次の 18 例では、*in-(negativo)*がある語形の方が頻度が高い : *imperdible.sus* (0:*in-*.1), *inagotable* (0:*in-*.8), *inalcanzable* (0:*in-*.1), *inconfundible* (0:*in-*.6), *incontable* (3:*in-*.1), *indefinible* (0:*in-*.22), *indestructible* (0:*in-*.1), *indiscutible* (0:*in-*.11), *indispensable* (0:*in-*.41), *ineludible* (0:*in-*.11), *inevitable* (0:*in-*.19), *infalible* (0:*in-*.1), *insoportable* (1:*in-*.12), *invencible* (0:*in-*.10), *irresistible* (0:*in-*.9).これらの多くの例でも接頭辞 *in-*ではなく接周辞 *in...ble* を設定できる。実際にほとんど使用されていない語形(肯定形: *perdible, agotable, ...*)からその否定形(*imperdible, inagotable*)を派生させることは合理的でないためである。よって、これらの否定語形は肯定語形に否定接頭辞を加えて派生したのではなく、動詞から直接派生して語彙化したものと考えられる。

一方、*posible > im-posible, visible > in-visible, responsable > i-rresponsable* という接頭辞付加は当然認められなければならない。このように、同じ接頭辞が異なる派生プロセスを持つことが不自然であるならば、両方ともに共通して *in-*の接頭辞付加を認め、*imperdible, inagotable* には *-ble* が同時に付加される、とする方法も考えられる。この方が現実に即した考え方であろう。そのとき、*imperdible, inagotable* に接頭辞・接尾辞と並べて接周辞を設定するのではなく、接周辞の 2 つの要素をそれぞれ別々に接頭辞と接尾辞に含めることになる。この理由は接周辞の第 1 要素はその語形・意味・機能において一般の接頭辞と語形・意味・位置が共通し、その第 2 要素はその語形・意味・機能において一般の接尾辞と共通するためである。また、一般の接頭辞・接尾辞よりも頻度はかなり少ないので、それらと同列に扱えない、という理由のためでもある。本研究ではこの方法を採用した。

他に、単一形態素動作名詞(*comprar > compra, estudiar > estudio*)と短縮語(*palabra abreviativa: colegio > cole, bicicleta > bici*)も種類・頻度が少ないために、どちらも単一語(*base*)に統合した。

合成語(*composición*)は独立した語が結合したものとされる : 例 *alto + voz > altavoz, girar + sol > girasol*.よって、たとえば *neogótico* の *neo-*は独立して使われないので合成語にはならない(Camus 2022: 82)。 *neogótico* が合成語では

2. f. Gram. Proceso de formación de palabras en que intervienen simultáneamente la prefijación y la sufijación; p. ej., embaldosar.

なく派生語であるとするれば *neo-*は接頭辞になる。一方、*anglo-árabe*, *cardiovascular* などの *anglo-*, *cardio-*も独立して使われないが、これらは合成語とされる(同: 84)。次の Real Academia Española, *Diccionario de la lengua española*. (DLE, 23a ed.) (s.v. *composición*)の説明も同様である：

DLE (s.v. *composición*): "9. f. Gram. Proceso morfológico por el cual se forman palabras a partir de la unión de dos o más vocablos, de dos o más bases compositivas cultas o de la combinación de palabra y base compositiva. Se forman por composición *rompeolas*, *neuralgia* y *nefrotóxico*."

私たちは *neogótico*, *angloárabe*, *cardiovascular* の *neo-*, *anglo-*, *cardio-*をすべて接頭辞ではなく、「語」の意味と形態を備えた Real Academia Española, *Diccionario de la lengua española*. (DLE, 23a ed.)の用語'*elemento compositivo*' (たとえば s.v. *cefalo*)に従って、「合成辞」(*compositivo*)とし、これらの語形を派生語(*palabra derivada*)ではなく合成語(*palabra compuesta*)と見なす。たしかに *altavoz* の *alto*, *voz* は独立可能な自由形態素、*neogótico* の *neo* は独立不可能な拘束形態素であることが、大きな違いのように見える。しかし、ここに「異形態」(*alomorfo*)という概念を導入すれば、拘束形態素 *neo-*は自由形態素 *nuevo* の異形態と見なせる。よって、自由形態素 *nuevo* との連絡によって「語」の資格を得る。*anglo ~ inglés*, *cardio ~ corazón* も同様である。仮に *neo-*を接頭辞とすると、他の一般的な接頭辞(*des-*, *in-*, *pre-*など)が語彙的意味を持たないので、両者には大きな齟齬が生じる。私たちが「合成辞」という単位を設けた理由は *neo-*のような独立した語(*nuevo*)の異形態を *altavoz* の *alto* のような独立語と同様に扱うためである。

よって、語彙全体を単一語と複合語に分類し、複合語の下位分類として派生語・合成語を区分し、派生語の中で接頭辞・接尾辞を区分した。

(7) 語彙の分類

全体	{	単一語	
		複合語	{
		派生語：接頭辞・接尾辞	
		合成語：語・合成辞	

すべての資料(「8種資料」)で収録された語彙(1,112,360語)はこの区分に従うが、接頭辞と接尾辞は重複を避けるために、外側の接辞だけを計算した。たとえば *imposibilidad* は接頭辞 *in-*と接尾辞 *-dad* だけを計算に加え、(*ibil >*) *ble* は除外した。

派生・合成に伴って発生する異形態はすべて形態素の代表形にまとめた。たとえば *posibilidad* は *pos.ibil.idad* として分析し、*idad* は *dad* の異形態とす

る。 *pos* は単一語 *pod(er)* の異形態とする。

concebir, *recibir*, *percibir* などに接頭辞 *con-*, *re-*, *per-* を認め、これらを単一語ではなく派生語と見なす。 *con-*, *re-*, *per-* は他でも共通して使われ (*conducir*, *combinar*; *remover*, *retrasar*; *perdonar*, *perdurar*), *ceb*, *cib* は共通して異形態 *cep* を持つからである (*concepción*, *recepción*, *percepción*)。以上の共通性はラテン語の語形成規則で確認される (Gómez de Silva, 1988)。一方, *concebir*, *recibir*, *percibir* を分析せず全体を単一語とすると, これらの共通性を偶然の一致としなければならない。

-*ar* は *corto* > *cort-ar* のように接尾辞になることがあるが, *comprar* は *compra* から派生したのではなく, 逆に *compra* は *comprar* から逆派生 (*derivacio/n regresiva*) したものである。そのような逆派生語には「...の動作」 (*acción de ...*) という意味になる。ここで *compra* は単一形態素の語であり, *compr-a* (名詞化) という分析はしない。-*a* に「名詞化」という積極的な意味は認められず, -*ción*, -*miento* と同列には扱えないためである。一般に接尾辞は強勢を帯びるが, -*a* には強勢がないこともこれを接尾辞としない理由になる。

語彙の代表形(「見出し語」「レマ」: 名詞・代名詞・冠詞・形容詞などの男性単数形, 動詞の直説法現在 1 人称単数形)を単一語(語根)・接頭辞・接尾辞・合成辞に区分するには, 関連する語形を比較するとともに, Saporta (1959), Colominas (1976), Gómez de Silva (1982), 秦(1999), 太田(2012), 寺崎(2019), 岡本(2021)を参照した。

1.6. 方法

語(単語)を計量的文体研究の対象とするとき, その長さ(文字数・音節数・音素数)が計測され, それぞれの長さの語の使用頻度の総和や平均が求められる。一般に単語(の長さ)が長いと, それだけ使用頻度が減少することが知られている。逆に短い単語の使用頻度は多いので普通に使われる。一方, 頻度が低い長い単語は目立つので有標であると考えられる。そして, 長い単語が多用されると短縮される, という(「短縮の法則」)。語長は単語の形態的・文体的特徴として重要な問題なので, この「法則」が真に言語的に正しいか否かをはじめに検証しなければならない。このとき頻度分布の様子を観察することが必要である。

言語研究者はすべての単語の長さを一律に機械的に計測する方法は採用しない。すべての単語は文法的観点から様々な品詞に分類され, 語彙的観点から様々な語形成法に分類されるからである。そのような言語的分類法を無視して行う機械的計測は言葉の言語的側面の解明には繋がらない。そのよう

な機械的計測は、たとえば、ある地域の哺乳類と魚類の体長を混ぜて計測しその平均値を求めたり個体を比較するようなことと同様である。機械的計測ではなく言語的計測を行うには、品詞分類と重なる部分があるがそれと同一ではない機能語・内容語の区別が重要である。この区別に立脚した測定をすれば、これまで機械的に行われてきた測定では見えなかった単語の長さに関する事実が明らかになる(→2.2)。

次に、単語を長くする要因を語形成の面から考察する。音素や音節が物理的に延長して単語の長さを決定しているのではなく、単語の内部構造を検査すればそこに様々な語形成要素(語根・接頭辞・接頭辞・合成辞)が構造的に繋がって長い単語を形成していることがわかる。それらの語形成要素はゼロ→多数まで様々なグレードがあり、それに比例した有標性を示している。その有標性は語形成要素・使用頻度・使用領域・文体という多角的側面から観察しなければならない。

数千のすべての単語の使用頻度分布を作成しても、その全体的分析は困難であり、無駄になることも多い。なぜならば、僅かな頻度しか持たない語の分布パターンは偶然である可能性が高いからである。また、変数間の使用頻度の変異(分散)が小さい単語は全体に平均して使用されているので、文体差を示さない。よって、使用頻度の和と分散(標準偏差)の両者を勘案した総合値を求め、その総合値の上位にある語形を分析する。

使用頻度は極端なべき乗分布に近い分布を示すので、対数に変換する必要がある。そして、標準偏差は一般に平均値の大きさに左右されるので、標準偏差を平均値で割った「変動係数」(coefficient of variation)を用いるべきである。さらに変動係数を範囲[0, 1]の中に収めるために「相対化標準偏差」を用いることが望ましい(→プログラム-R)。

文体特徴の定量的実態はその頻度分布の中で観察される。このとき、データの姿そのものの解釈は困難であるので、分布をわかりやすくパターン化する。そのパターン化の方法の一つとして、私たちが開発した「集中分析法」(concentration:→プログラム-R)を採用する。この方法によって、使用頻度の分布状態が左側に集中するタイプ・中央に集中するタイプ・右側に集中するタイプに区分され、その順番に従って語形が上下に並べ替えられる。このような集中分析法を使って語形選択のパターンを観察する。

分布パターンの分析は機能語・内容語と単一語・複合語に分けて行う。そこで特異な集中分布が見られれば、それが「文体特徴語」(characterizante estilístico)である可能性がある。そのとき、集中分布を示す単語がすべて文体特徴語であるわけではなく、それが「話題語」(palabra temática)である可能性もある。たまたまある変数(学年・出版物・性・年齢・学歴)に集中していても、文章のテーマ(話題)に従って出現した単語であることも多い。とくに

名詞の選択は文章の記述内容に大きく依存する(金 2016: 63)。よって、文体特徴語を決定する際には他の資料と比較したり、母国語話者に直感・経験に基づく意見・印象などを求めたりして確認する必要がある。

分析には各種の統計的方法を用いる。既存の方法では難解な部分もあるので、なるべく簡単に直感的に理解できる方法を選び開発した⁶。

2. 語長

2.1. 序論

計量的言語研究でしばしば取り上げられるテーマに使用語(単語)の長さがある(以下で「語長」*longitud léxica* とする)。語長は一般に短いほど頻度が高く、語長が長くなるにつれて使用度数が減少する傾向がある、と言われている。このことは自明とされているが、言語学的見地から再検討を要する問題である。この問題は計量的言語研究だけでなく、語長を文体特徴の一つとする計量文体論にとっても重要である。なぜなら長い語の使用頻度が少なければ文体的に有標であると考えられるからである。たとえば L. M. Montgomery. *Anne of Green Gables* 『赤毛のアン』(村岡花子訳新潮社:123)の主人公アンが「(...)もしダイアナがあたしを好きにならなかつたらどうしよう。あたしの生涯における最大の悲劇的な失望となるでしょうよ」"What if she shouldn't like me! It would be the most tragical disappointment of my life."と言うと、養母マリラは「(...)そんなながつたらしい言葉を使ってもらいたくないもんだ。小さな女の子が使うと、とってもこっけいに聞こえるからね。」"I do wish you wouldn't use such long words. It sounds so funny in a little girl."と言ってたしなめている(New York. Bantam Books. 85)。ここでマリラは *the most tragical disappointment of my life* という長い大袈裟な表現を話題にしているようだが、それは *tragical, disappointment* というフランス語に由来する長い語形成のことなのかもしれない。両方を区別なく問題にしている可能性もある。

私(上田)が最初にスペイン語を学んだとき、多くの単語が英語の単語よりも長いように感じた。たとえば *ing. put, eat, write, work* は 1 音節(3~5 文字)であるが、スペイン語は *po.ner, co.mer, es.cri.bir, tra.ba.jar* は 2~3 音節(5~8 文字)である。竹蓋(1981: 179-182)は各種の英語資料の使用語彙(述べ語数)⁷の

⁶ 以下の統計処理では R を使用した(R Core Team 2021)。グラフの作成には ggplot2 (Wickham 2016)を使用した。プログラム-R も多用した(→プログラム-R)。

⁷ 語数として「述べ語数」(token, running words)と「異なり語数」(type, different words)を区別する。たとえば、{a,a,a,b,b,c}という語の集合の述べ語数は 5 で

語の長さ(平均)を文字数で測ると 4.0~5.6 文字であったことを報告し、「文字数の多い語のほうが少ないものよりむずかしい語と考えることができるので、この基準による語彙の比較も無視すべきではない」と述べている。吉岡(1996: 201)の調査では、新聞記事(2種)・小説(2種)・高等学校教科書(2種)の平均値はそれぞれ 5.2, 5.0; 4.2, 4.3; 3.9, 4.4 文字であった。一方、私たちがスペイン語の単語(述べ語数)の平均文字数について、Juilland and Chang-Rodríguez (1964), Justicia (1995), Martínez y Ueda (2021, 2023)に収録されている語彙(延べ語数)の平均文字数を調べるとそれぞれ 7.4, 7.4, 7.3 文字であった。よって、たしかに英語の単語文字数(4-5 文字)よりもスペイン語の単語の文字数(7 文字)が多い。

スペイン語の語長が長い理由として、音韻の弱化による短縮が抑制されたことと、語形成のプロセスが生産的であることがあげられる。「語が多用されると短縮される」としばしば言われてきたが、スペイン語ではそのようなことはあまり起こらなかった。むしろ豊富な接頭辞・接尾辞などによる語形成のプロセスが新語を生み出し、その結果比較的長い単語が堅固に(短縮されることなく)使用されている。

私のはじめて出会った単語の中でとくに学習が困難であったのは、たとえば *en.ton.ces* である。この言葉は接頭辞・接尾辞のない 1 形態素語であり、その上、これを基底(base)とする派生語もないため孤立した単語である。英語ならば 1 音節 4 文字の *then* で済ませることができるのに、*entonces* は 3 音節 8 文字となっている。高頻度語であるのに短縮される様子はまったくないし、その完全形は今でも健在である。しかし、*entonces* のような孤立した長い単語は例外的で、一般の基底は比較的短く、長い単語は語形成(接頭辞・接尾辞など)による派生形態であることが多い。Real Academia Española の *Diccionario de la lengua española*. (DLE, 23a ed.)によればスペイン語の最長の語は *electroencefalografista* 「脳波検査技師」(23 文字)であるが、この語は *electro.encefalo.graf.ista* のように分析され、*encéfalo* 「頭部・脳」や *cefalo* という合成要素(DLE: *elemeneto compositivo*)を知ればそこから派生される多くの語(*cefalalgia* 「(医)頭痛」、*cefalitis* 「(医)脳炎」、*cefalópodo* 「(生)頭足類」、*hidrocefalia* 「(医)脳水腫」など)が困難なく学習される。*electroencefalografista*(23 文字)に続く *anti.con.stitu.cion.al.idad* (22 文字)、*anti.norte.americ.an.ismo* (21 文字)も分解可能である。

以下では、語長を直線的・表面的に計測するだけでなく、語の構造(派生・合成)も視点に入れて、語長と使用頻度・語形成・文体の関係について計量的に分析し考察する。

あり、異なり語数は 3 である。本論で扱う「使用頻度」は述べ語数を指し、「見出し語の語数」は異なり語数を指す。

2.2. 短縮の法則

2.2.1. 先行研究

語長を計測するときに使う単位は文字・音節・音素が考えられる(Herdan 1956: 176-197, Gómez Guinovart 1999: 97-113)。文字数で計測する語長は、スペイン語で2文字で1音を表したり(*chico*, *que*, *llorar*, *perro*), 読まれない文字があつたり(*hombre*, *psicología*, *mnemotecnia*), 1文字が1音または2音を表したり(*extremo*, *examen*)するので正確でない。そこで、多くの研究では音節を単位として使用している(Navarro Tomás 1966: 54-55, 樺島 1968: 16, Miller 1979: 108)。しかし、たとえば *a*, *de*, *dos*, *tres* はどれも1音節語であり、*ele*, *paso*, *vista*, *puesto*, *muestra*, *monstruo* はどれも2音節語であるが、それらの語長は大きく違う。よって最も正確な単位は音素である: *a* (1音素), *de* (2), *dos* (3), *tres* (4); *ele* (3), *paso* (4), *vista* (5), *puesto* (6), *muestra* (7), *monstruo* (8)。Saporta (1963: 69)は形態素を構成する音素の数を計測した。Wierzbicka (1965, 2011 :211-213)は音素数で語長を計測しているが、個別の単語の頻度を比較するのではなく、使用頻度順位を3グループに分け(I:ランク 45-54, II: 295-304, III: 2495-2504), グループ間の比較をしている。

言語を横断した一般言語学の著作の中で語の長さと言語の使用頻度の関係が論じられ、一致して「語長が長くなると使用頻度が下降する」と述べられている(Zipf 1936: 3-39, Zipf 1949: 63-66, Wierzbicka 1965 [2011]: 211-226, Whatmough 1956 [1966]: 69, Martinet 1970 [1972]: 258-263, Miller 1979: 107-109, 樺島 1968: 15-19, 安本 2009: 255, 田中 2021: 168-174)。その要因として「最小努力の法則」や「経済」などが挙げられている。つまり、頻度の高い語の語長が短かければ使用における負担が少ないために最小努力の法則に基づき経済的である。そして頻度の高い語の語長が長ければ使用における負担が大きいため短縮される(「短縮の法則」 'Law of Abbreviation', Zipf 1949: 66, *ing. telephone > phone*, *gasoline > gas*, *omnibus > bus*; Martinet 1970 [1972]: 261, *fr. chemin de fer métropolitain > métro*)。逆に、語長が長い語はその負担が高いため使用頻度が低くなる、という(Zipf, Whatmough, Martinet)。

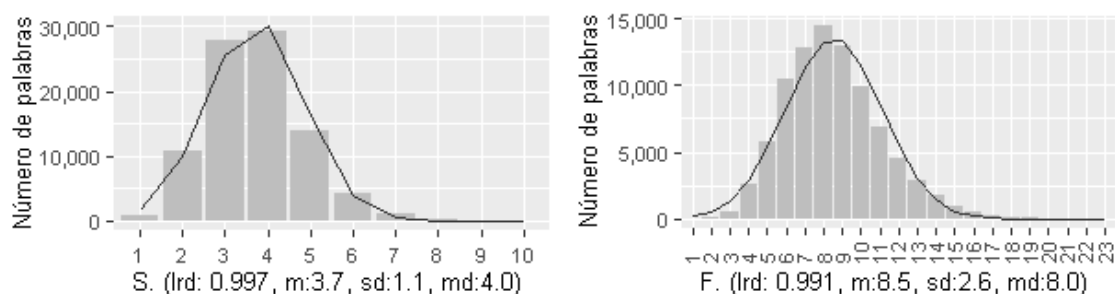
国立国語研究所報告「現代雑誌九十種の用語用字」の資料を分析した樺島(1968: 15-19)は「人間がことばを使った長い歴史の間に言語行動の労力を軽減しようとする力、すなわち、よく使われる語としては語形が短い語が生き残るという淘汰作用」が働いた一方、「新しい概念、事物、制度、あるいは複雑な概念などはいくつかの語構成要素が組み合わされて」「構成的に長いもの」になる、と述べている。安本(2009: 255)は「一般的に言って、英語の場合、使用頻度の大きい言葉は、単語の長さが短い傾向がある」、「音節数の長い言葉は、専門用語であつたり、難しい言葉であつたりすることが多い」

と述べ、「reconsideration」などのように、接頭辞や接尾辞がついて音節数の多い語」について触れている。一方 Bybee (2007: 260)によれば高頻度語に見られる「調音動作の短縮」(reduction of articulatory gesture)は「運動活動の正常な自動化」(normal automation of motor activity)に起因する。そして Lindquist (2009 [2016]: 10-11)は「going to のような高頻度の語の組み合わせは音の結合を促し(gonna), さらには新しい意味を獲得する(ここでは「未来」)傾向がある」と述べている。このようにどちらも高頻度使用が語形短縮の要因としている。しかし、田中(2021: 172)は、語長と使用頻度の関係が「人間の経済性を表すかどうかは疑問である」と述べ、むしろ単語は言語的理由(単語の合成的な構成)によって長くなることを指摘している。Divjak (2019: 33-36)はこれまで多くの言語学者が Zipf の「経済と最小努力の法則」(Principle of Economy and Least Effort)に疑義を呈してきたことを扱っている。

以下ではスペイン語の資料を使って「短縮の法則」の妥当性を検討し、単語の語長を決定する真の言語的要因を探る。

2.2.2. 辞書の見出し語

はじめにスペイン語単語の語長の全体像を見るために、Real Academia Española が発行した *Diccionario de la lengua española*. (DLE, 23a ed.) に収められた全見出し語(レマ lema: 87,729 語)⁸の音節表記と音素表記を使用し⁹, それぞれの語の音節数・音素数に対応する語の数を数えた。次がその結果である。



【図-1】辞書(DLE)の見出し語の音節数(S)・音素数(F)に対する語数

この左の棒グラフは音節数(S.sílaba)と対応する見出し語数(Número de

⁸ Domínguez [2021/08/26]の資料を使用する。ただし、b, c, nnなどの文字名や略語は除外した。psicología や mnemotecnia などの特殊な文字列は一般の発音(/s/, /n/)にした。一部の転記ミスを修正した。

⁹ 音素表記と音素数の計算には私たちの次のサイトを利用した。

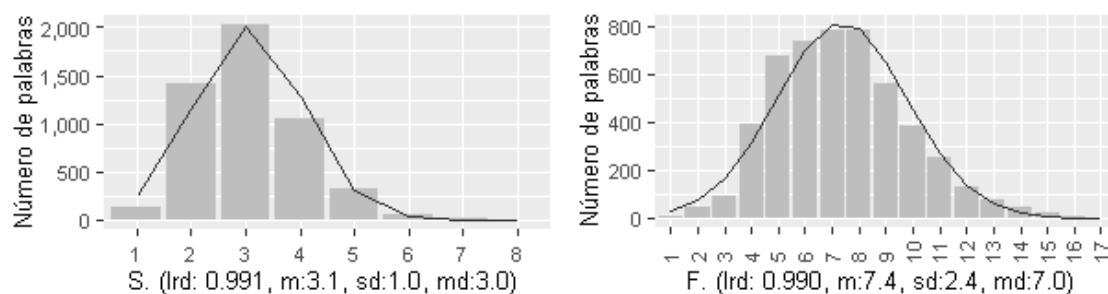
<https://h-ueda.sakura.ne.jp/lyneal/>

palabras)を示し、右の棒グラフは音素数(F.fonema)と対応する見出し語数を示す。折れ線グラフは理論的な正規分布を示す¹⁰。音節数-見出し語数の分布も音素数-見出し語数の分布もほぼ正規分布に近似している。その近似の程度は見出し語数と対応する正規分布確率の間の規定距離を測り、次のように1の補数によって近似の程度(「逆規定距離」inverse regular distance: ird)とした(上田 2016→プログラム-R)。*** Normality

$$\text{ird} = 1 - \text{規定距離}$$

見出し語数と対応する正規分布確率が完全に一致するとき、両者の規定距離はゼロ(0)になるので、逆規定距離(ird)は最大値1となる。音節数のグラフの逆規定距離は0.997を示しているので、近似の程度は非常に高い。音素数のグラフのirdは0.991であり、これも高い正規性を示している。その音節数の平均(m)は3.7であり、音素数の平均は8.5である。ばらつきを示す標準偏差はそれぞれ1.1, 2.6である。全データの中央値(median, md)はそれぞれ4.0, 8.0である。中央値が平均値に近いことは正規分布の特徴である。全体的に右に向かって裾が伸びていることが完全な正規分布と異なるが、その語数は僅少である。

言語の実態を知るために、Justicia (1995: 小学生の自由作文の使用語彙, 計 528,481), Juilland and Chang-Rodríguez (1964: 各種の出版物の使用語彙, 計 468,895), Martínez y Ueda (2021: Santander, España で行ったインタビュー録音資料, 計 115,095)に収録された見出し語を集めた。この3資料に重なって現れる語は1語とした「異なり語数」を見出し語の長さ(音節数・音素数)とクロスさせて計算して次のようなグラフを作成した。



【図-2】 X軸:見出し語の音節数(S)・音素数(F), Y軸: 語数(異なり語数)

先の辞書の資料と比べると見出し語数はかなり小さいので語数の規模が小さくなっている。しかし、全体的に正規分布によく近似している(ird: 0.993, 0.990)。

¹⁰ 正規分布の特徴は平均値の頻度数が最大でその左右に広がって頻度数が減少し、全体として釣鐘型(bell-shape)を示すことである。

このように、辞書(DLE)と資料(「8種資料」)の見出し語の長さと言語数の関係は正規分布に強く近似することを確認した。正規分布では平均値の頻度が非常に高く、その左右で頻度が急に下降し、左右の裾が低頻度で長く伸びる。音節数・音素数と言出し語数の関係が正規分布に近似する理由は次のように考えられる。語長の最小値(1音節)の近傍で最大になると、多くの語が1音節となって語の識別が困難になる。また、多くの語が1音素になることは現実的ではない。逆に語長の最大値(8音節, 18音素)の近傍で最大になると、生産に多大の負担がかかる。よって、両方の極端な最小値・最大値の集中を回避して語長の平均値で語数を最大とする平衡したシステムが理想的である(平均値=中央値=最頻値:3音節, 7音素)。

Greenberg (1963: 69)はスペイン語形態素資料(Saporta 1959)を使って形態素の音素数(1-14)と対応する形態素数(計 1679)の一覧を載せ、スペイン語形態素の平均音素数は4.4個であり、それを中心に「単峰性の曲線」(one-peak curve)を示す(音素数の最頻値4:形態素数387)、と述べている。この曲線も形態素の延べ数ではなく異なり数を示しており、その結果正規分布に近似している。

しかし、平均値が最頻値と一致してもその左右の頻度頻度が不規則に上下することもありうる。少数のサンプルで乱数を発生させるとそのようなことが起こる。しかし、スペイン語の最大規模の見出し語を収録したDLEや、多数の語形(468,895+528,481+115,095=1,112,471語)を集計した「8種資料」の資料が示す頻度分布は不規則に上下することはない。語長の平均値からの距離が近いほど平均値の最大語数に近づいて語数が多く、平均値から遠くなるほど、つまり両端の最小値・最大値に近づくほど語数が減少し、終端の最小値・最大値では語数が、平均値の最大語数から最も離れた最小になる。よって、見出し語の語長の使用頻度分布は語長が平均値の語数が最大頻度となり、その左右の方向に向かって次第に語数が減少し両端で最小となる、いわゆる釣鐘型の正規分布に近似する。以上が、辞書(DLE)と頻度辞典(J-Ch)の見出し語の語長の使用頻度分布は正規分布に近似する統計的理由である。

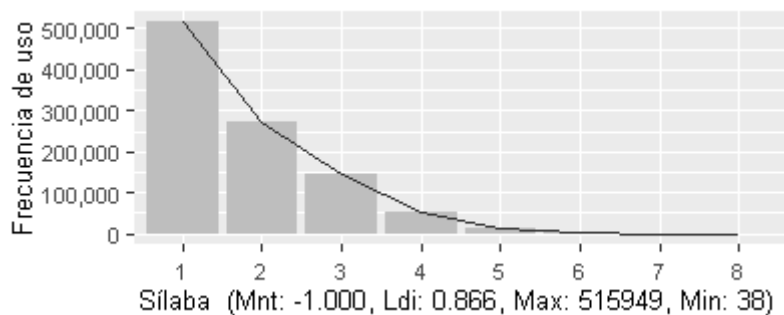
統計的理由は必ずしも言語の変異を完全に説明しない。中国語の単語は語長は最小の1音節であるが、その上に音調が加重されて弁別機能が維持される。一方、日本語(和語)ではさまざまな要素が次々に接合(膠着)されるので一般に語長が長くなる。スペイン語の単語の平均的語長(3-4音節, 8音素)は両者の中間になるが、多くの接頭辞・接尾辞による派生や一部の合成によってかなり延長することがある(前出の *electroencefalografista*, *anti.con.stitucion.al.idad*, *anti.norte.americ.an.ismo*, etc.)。派生は新語を形成するので、歴史を通じて次第に語長が延長してきたことが考えられる。一方、長い語は負担になるので短縮も起きるが、その数は少なく、使用範囲には制限がある

(*bicicleta*>*bici*, *colegio*>*cole*, etc.). 他にもスペイン語の語形はさまざまな条件(ラテン語形, 音韻変化, 形態変化, 借用語形)によって語形の長さが決定されてきた。このように多くの独立した要因が重なって起こる現象の頻度は一般に平均値を最大頻度とする正規分布に近似することが知られている。そして, 極端に短い1音素の語は母音音素の総数(=5)を超えることができない。一方極端に音素数が多い語もそれだけの語形成が加重されていなければならないので, 少数である。この両極端から平均に向かって語数が増えていることは自然である。

2.2.3. 使用頻度

見出し語の集合は使用可能な語の全リストを示している(「異なり語数」(type)と呼ばれる)。一方, 実際に使われている語の使用実態はそれと大きく異なる(「延べ語数」(token)と呼ばれる)。たとえば, 辞書の語数の調査では *a*, *de*, *en*, *con*, *por* などの前置詞や *que*, *y* などの接続詞はそれぞれ1回として数えられるが, 各種の頻度辞典が示すこれらの語の使用頻度は非常に多い。「8種資料」にはすべての見出し語の使用回数が示されているので, それを利用して語長(音節数・音素数)と使用頻度(Frecuencia)を総計した結果を以下に示す。

Ldi***



【図-3】 X軸:音節数 : Y軸:使用頻度(延べ語数).

このグラフを見ると, 音節数(S)が増すごとに使用頻度が減少していることがわかる。これは一見して, 従来言われてきたように「語長が長くなると使用頻度が小さくなる」「使用頻度が大きくなると語長は短縮する」という「短縮の法則」を示すようにも見える。

このグラフでは使用頻度が単調に下降する(上下しない)ことを示すために棒グラフの頂点を結んで線グラフとした。単調下降の程度を数量化して「単調性指数」(Mnt: Monotony)を定義した。単調性指数の範囲は[-1, 1]であり, -1が単調下降, 1が単調上昇を示す(上田 2016: 4章→5.プログラム-R)。この

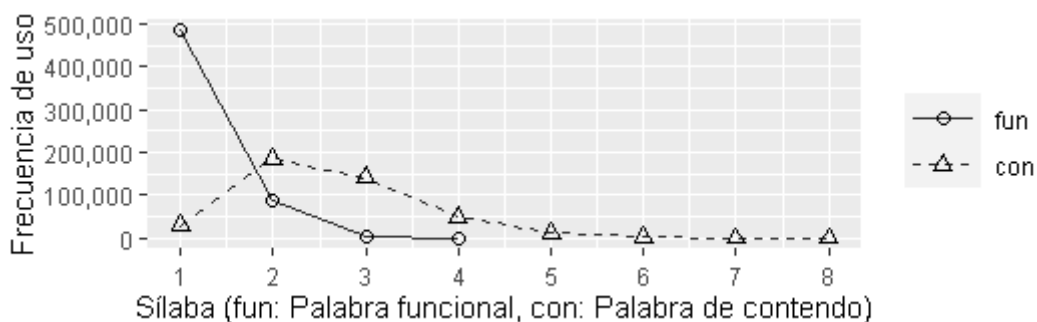
グラフのように第 1 項目の頻度が最も高く、以下急に頻度が下降するような分布は「L 字型分布」と呼ばれている。私たちの「L 字型分布指数」(Ldi: L-shaped distribution index: Ldi)は L 字型分布の強度を示す指数である(範囲[0, 1]) (上田 2016: 4 章→5.プログラム-R)。

このように語の音節数とその使用頻度が L 字型分布をする理由を探るために、それぞれの音節数をもつ語の内容を観察する¹¹。

¹¹ 品詞 : adj: adjetivo, adv: adverbio, art: artículo, cif: cifra, conj: conjunción, dem: demostrativo, extr: extranjerismo, interj: interjección, interrog: interrogación, num: numeral, pos: posesivo, prep: preposición, pro: pronombre, prop: nombre propio, rel: reletivo, sn: sí-no, sus: sustantivo, vb: verbo.

- (8) 1 音節語の高頻度語(使用頻度降順): el.art 118997, de.prep 76116, que.conj 36827, y.conj 32648, a.prep 30109, en.prep 28910, un.art 22875, ser.v 21015, se.pro 17597, no.sn 13365, con.prep 11419, su.pos 10826, por.prep 10656, lo.pro 10348, más.ind 5307, me.pro 5206, ir.v 5123, o.conj 4305, él.pro 3541, si.conj 2857...
- (9) 2 音節語の高頻度語(使用頻度降順): para.prep 7918, haber.v 7382, este.dem 6284, estar.v 5950, tener.v 5901, como.conj 5796, hacer.v 4741, todo.ind 4643, poder.v 4511, ese.dem 4472, pero.conj 4297, decir.v 4197, otro.ind 3249, mucho.ind 2257, porque.conj 2163, año.n 2158, cuando.conj 1996, solo.a 1995, también.adv 1955, saber.v 1954...
- (10) 3 音節語の高頻度語(使用頻度降順): alguno.ind 1622, primero.a 1537, ahora.adv 1111, parecer.v 1011, entonces.adv 953, encontrar.v 936, persona.n 842, conocer.v 802, momento.n 760, último.a 724, durante.prep 697, además.adv 679, trabajo.n 667, problema.n 611, esperar.v 551, ejemplo.n 543, amigo.n 538, nacional.a 527, público.a 523, nosotros.pro 521...

上のリストを見ると、1 音節語の中で最も高頻度である語のほとんどが冠詞・代名詞・前置詞・接続詞などの機能語であることがわかる。それらとくに上位の語は極めて頻度が高い。よって、機能語の頻度が1 音節語全体の頻度を高めていることが考えられる。しかし、低頻度語の内容語であっても、その数が多大であれば、その多数の内容語も1 音節語全体の頻度を高めていることになる。そこで、機能語と内容語の語長による頻度分布の全体を示すグラフを見なければならぬ。



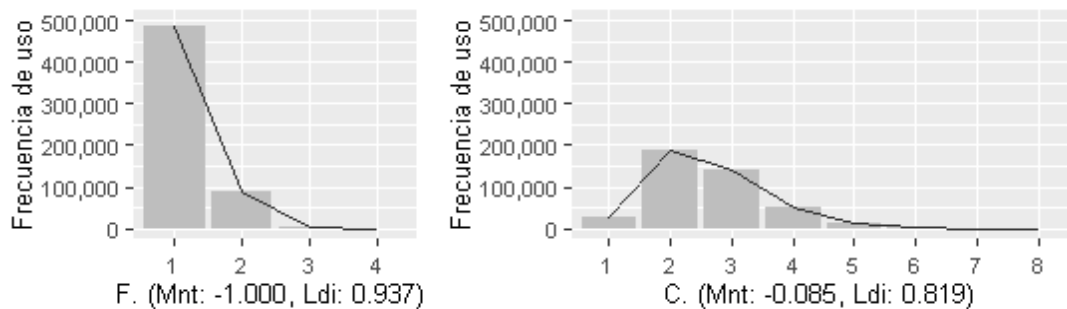
【図-4】機能語(fun)と内容語(con)の語長(音節)による頻度分布

上のグラフの1 音節の部分を見ると、圧倒的多数が機能語(fun)であることがわかる。2-3 音節では逆転して内容語(con)のほうが多くなり、4 音素以上では内容語のみとなる。先のリストを見ると、確かに1 音節語(8)は機能語ばかりであり高頻度語の中に内容語は見つからず、2 音節語(9)の中を見ると機

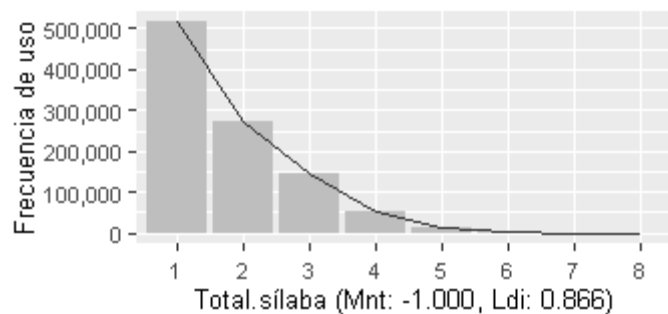
能語(para, como, haber, pero, ...)と内容語(vida, hombre, también, siempre)が混じり、3音節語(10)のほとんどは内容語である。よって、頻度分布は機能語(少ない音節数)→内容語(多い音節数)という順番になる。このことが頻度分布が音節数の増大に伴って全体的に単調下降する要因となっている、と考えられる。

さらに、L字型分布の特徴は単調下降であるだけでなく、それが急下降であることである。語の音節数に従って使用頻度が急下降する理由は機能語と内容語の極度の頻度差に因る、と考えられる。

このように、語の音節数に従って使用頻度がL字型の分布を成す理由は機能語と内容語の大きな頻度差にあるようだ。そのことを確かめるために、資料全体を機能語と内容語に分けて、それぞれの頻度分布を観察しよう。



【図-5】 X軸:音節数 : Y軸:使用頻度(延べ語数). F:機能語 / C:内容語



【図-6】 X軸:音節数 : Y軸:使用頻度(延べ語数) T:全体

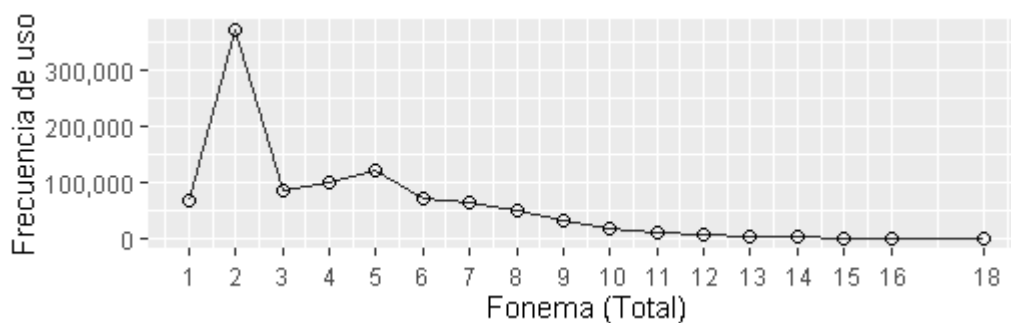
上の3つのグラフはそれぞれ、F:機能語、C:内容語、T:全体の頻度分布を示している。F:機能語は1-3音節の語の頻度分布が完全に単調な下降を示している(mnt: -1)。しかし、C:内容語の頻度分布はL字型分布ではなく、むしろ正規分布に近似している。F:機能語とC:内容語を合体させたT:全体の分布は再び単調下降になる。ここでT:全体の1音節語の頻度は、F:機能語の頻度が反映していることが確認できる。

英語の語長を扱った多くの研究では、音節を使って語長を計測しているが、

音節数よりも音素数を使ったほうが語長を正確に計算できるので(→2.2.1), 以下では音素数で8種資料の語長を計算する。



【図-7】 語長(音素)による頻度分布. 機能語(fun)と内容語(con)



【図-8】 語長(音素)による頻度分布. Total:全体

ここで、音節数で計算した語長では見えなかった事実が、音素数で計算すると明らかになる。それは、問題の頻度の下降が最初に1音素語から観察されず(T:全体【図-8】), 1音素語から2音素語に向けて頻度が上昇することである。2音素語の使用頻度が特出し、近傍の他の音素数の語の使用頻度にはあまり大きな差が見られない。このように一部が特出した分布を「I字型分布」と呼ぶことにする。これは正規分布とも大きく異なる。音素数で計測した語長と使用頻度の関係がI字型分布となる理由を以下に考察する。

I字型分布となる理由は、その特出した音素数=2を見ればわかる。次は音素数1-2の語の頻度を示すリストである(使用頻度を降順で並べ替えた)。

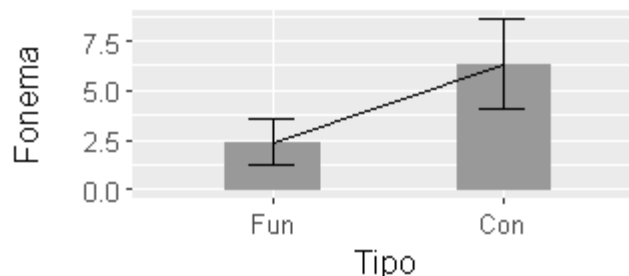
(11) 1音素語の高頻度語(使用頻度降順): y.conj 32648, a.prep 30109, o.conj 4305, eh.interjec 563, ah.interjec 197, oh.interjec 44, m.interjec 27, hm.interjec 25, e.sus 13.

(12) 2音素語の高頻度語(使用頻度降順): el.art 118997, de.prep 76116, que.conj 36827, en.prep 28910, un.art 22875, se.pro 17597, no.sn 13365, su.pos 10826, lo.pro 10348, me.pro 5206, ir.v 5123, él.pro 3541, si.conj 2857, mi.pos 2806,

ya.adv 2631, yo.pro 2631, qué.interrog 2067, te.pro 2017, sí.sn 2001, tu.pos 1243.

このように、高頻度の2音素語のリスト(12)には冠詞(art), 前置詞(pre), 接続詞(conj), 代名詞(pro)などの機能語が多い(→1.4)。一方、名詞(n)・形容詞(a)・動詞(v)などの内容語は見つからない。この2音素語の高頻度の機能語が2音素語全体の頻度を上昇させ、全体の頻度分布をI型にしている。よって、音節ではなく、音素で語長を計測すれば、「語長が長くなると使用頻度が小さくなる」「使用頻度が大きくなると語長は短縮する」という「短縮の法則」は成立しない。音節で語長を計測すると、1音素と2音素の区別をなくして1音節にまとめられてしまい、その結果、全体の頻度分布が単調に下降するように見えるので「短縮の法則」を認めていたことになる。従って音節よりも精密な音素による語長の計測によって「短縮の法則」は否定される。

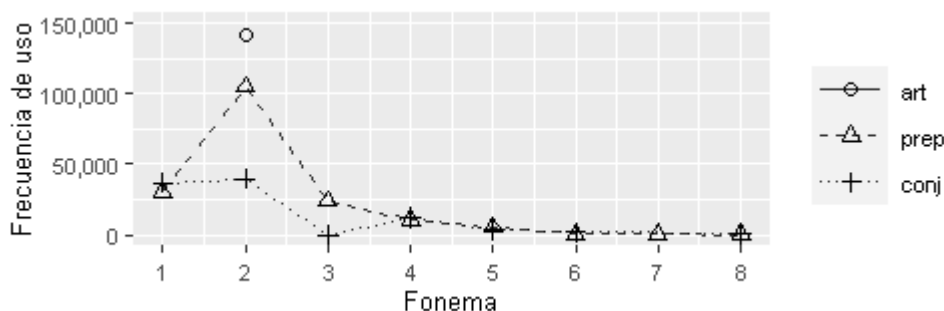
次は資料全体(「8種資料」)の機能語(Fun.)と内容語(Con.)の音素数の平均(m)・標準偏差(sd)を示すグラフである(Fun: m=2.39, sd=1.14; Con: m=6.32, sd=2.29)。



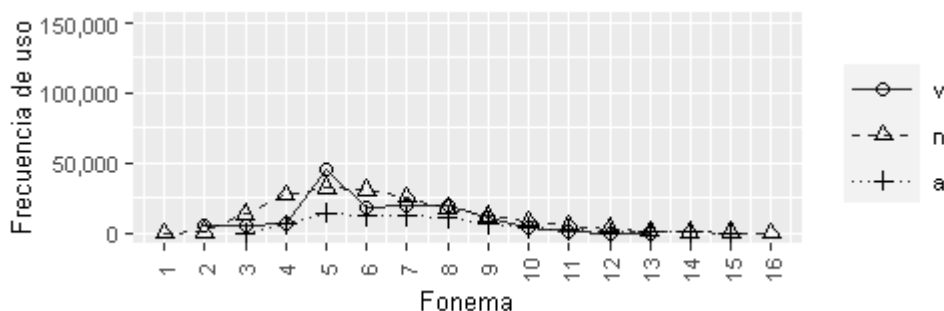
【図-9】機能語(Fun.)と内容語(Con.)の音素数の平均・標準偏差

このように、機能語の平均音素数は少なく、内容語の平均音素数はその2倍以上(2.7倍)である。また、標準偏差も内容語のほうが大きい。このことは意味を担う内容語の語長の変異が大きく、一方、語長が短い機能語は語長の平均値に集中して多用されることを示している。

次に、機能語と内容語の中を見るために、それぞれの主な品詞(機能語:冠詞・前置詞・接続詞・指示詞・所有詞; 内容語:名詞・動詞・形容詞・副詞)の音素数~使用頻度のグラフを観察する。



【図-10】機能語:冠詞(art)・前置詞(pre)・接続詞(conj)



【図-11】内容語:名詞(n)・形容詞(a)・動詞(v)

このように機能語の中で特に冠詞・前置詞が音素数=2の語の使用頻度を高くしている(図-10a)。先のリスト(14)を見ると音素数=2の冠詞・前置詞は *el, de, en, un* であることがわかる。よって、頻度が高くなるにつれ語長が短くなる、という「短縮の法則」は使用頻度を語長短縮化の要因としているが、実は語長が短い機能語(とくに冠詞・前置詞の *el, de, en, un*)の使用頻度が大きく影響しているのである。つまり、使用頻度が要因となって語長を短縮しているのではなく、逆に機能語の語長が短いことが要因となって、使用頻度を高くしているのである。よって、高頻度(要因)→語形短縮(結果)の関係よりも、逆に、短縮機能語(要因)→高頻度(結果)という関係を見るべきである。以上が機能語・内容語の区別から見た「短縮の法則」の解釈である。

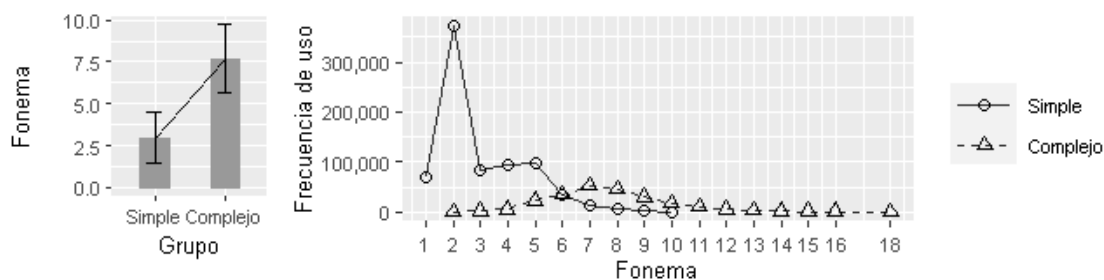
次に「単一語」(*palabra simple*)と「複合語」(*palabra compleja*)の区別から「短縮の法則」を解釈する。ここで「単一語」は語形が単一の形態素から成る語である(例：*noche, grande, tarde, cantar*)。一方、「複合語」は単一語に次のような語形成過程が加わった語を指す。

(13) 派生語: *re + considerar* → *reconsiderar*, *admirar - ble* → *admirable*.

(14) 合成語: *alto + voz* → *altavoz*, *sacar + corcho* → *sacacorchos*.

次のグラフはスペイン語の内容語の中の単一語(Simple)と複合語(Complejo)の音素数の平均・標準偏差(Simple: $m=2.99$, $sd=1.56$; Format.:

m=7.77, sd=1.99))と頻度分布を示している。



【図-12】(左)単一語(Simple)と複合語(Complejo)の音素数の平均・標準偏差
(右)音素数~使用頻度

このように、単一語と複合語の語長による使用頻度はどちらも平均値を中心とした正規分布に近似する。当然であるが、単一語に比べて複合語の分布のほうが音素数は多いので折れ線は右に寄っている。そして、単一語の使用頻度の曲線が複合語よりも上に伸びている。単一語に比べ複合語の使用頻度が低いのは、単一語に接頭辞・接尾辞が接合した派生語では意味が特殊化して有標となるためであると考えられる。たとえば Juilland and Chang Rodríguez (1964)によれば *corto* (使用頻度: 46) > *acortar* (9), *grande* (795) > *engrandecer* (5), *posible* (144) > *imposible* (42). このように、無標の単一語よりも有標の派生語のほうが使用頻度は少ないことは言語普遍(language universal)と合致する(Greenberg, 1966 [2005]: 56)。さらに、単一語の音素数の範囲が狭く[1, 10]であるのに対し、複合語の音素数の範囲は広く[2, 18]となっている。このことは派生・合成によって語彙の長さのバリエーションが豊富になっていることを示している。

以上で、機能語・内容語と単一語・複合語のそれぞれの区別の中で語長と使用頻度の関係を見た。それぞれの区別の左項(機能語, 単一語)の方が右項(内容語, 複合語)よりも語長が短く使用頻度が高いことを観察した。頻度分布の折れ線が完全に二分されていないので、一部重なる部分がある。その重なった部分では右項(内容語, 複合語)の方が左項(機能語, 単一語)よりも使用頻度が高いことがあるが、全体から見れば比較的少数である。よって、全体的な傾向としては、機能語, 単一語の方が内容語, 複合語よりも語長が短く頻度が高いことは否めない。

しかし、このことは「短い語長ならば高頻度」という因果関係も、その逆の「高頻度ならば短い語長」という因果関係(「短縮の法則」)も示していない。短い語長と高頻度に相関関係(因果関係ではない)があるように見えるのは、両者に共通して機能語・単一語という要因があるためである。同様に「長い語長=低頻度」という誤った因果関係については、内容語・複合語が共通

要因として考えられる。

2.3. 変異

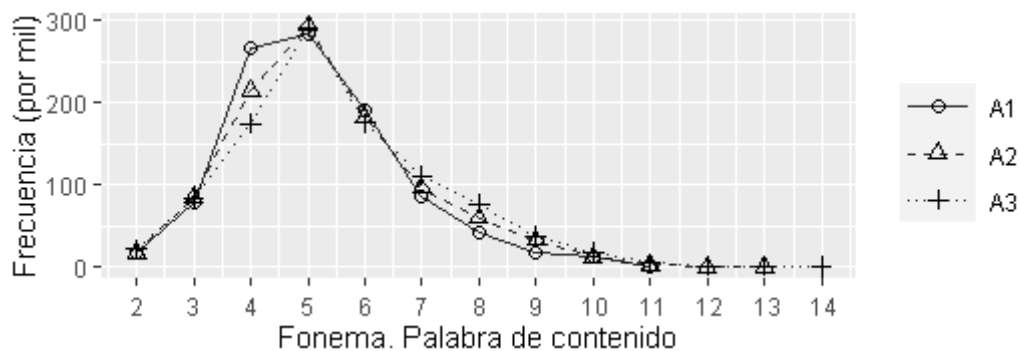
2.3.1. 先行研究

語長の長い語がふつう比較的稀に使われるために、語長が長い語が多く使用された文章は高度な文体である、という印象を与える。それに対して、使用頻度が高く短い語形が多用されている文章は平俗な文体となる。文学作品を対象とした文体論研究では作家に固有な特徴の一つとして語長が取り上げられている(安本 1960: 218-223, 安本 1977: 417, 波多野 1988: 95, 伊藤 2017: 125-126, Frías Delgado 2009)。言語学的文体論研究では文学作品に限らず様々な分野の出版物の比較分析において単語の語長を測定することがある(吉岡 1996: 200-201)。応用研究においてテキストの「読みやすさ」「平易さ」(legibilidad, 'readability')を数値化するための変数の1つとして利用されている(Crawford 1985, Alliende 1987, Gómez Guinovart 1999b, Ferrer et al. 2009, 安本 2009: 255-256, Ríos Hernández 2017)。

この章では音素数で語長を計算し、それと語の使用頻度との関係を観察する。このとき単語の語長とその使用頻度の逆関係、そしてその理由として挙げられる「短縮の法則」(使用頻度が高いと語長が短縮される)の一般化の可否を、語種と語形成(接辞・合成・逆成・短縮)の視点から改めて問題にしたい。続けて、語種と語形成と種々の文体との相関を観察する。

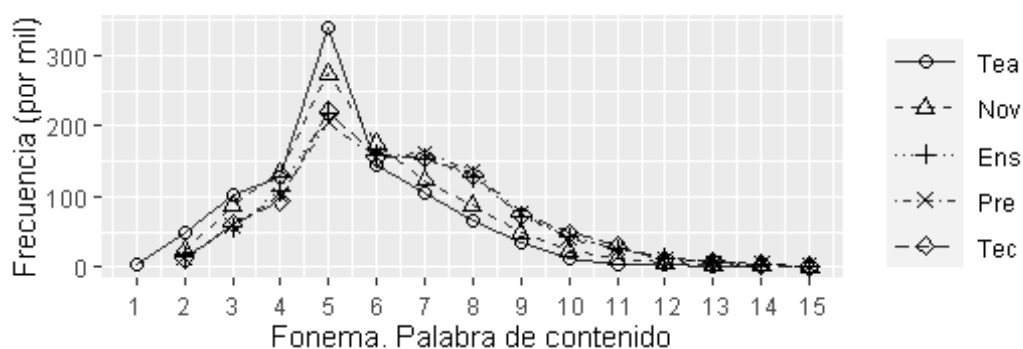
2.3.2. 内容語

以下では3種の資料を使って学年差・出版物・性差・年齢差・学歴差を示す各グループの内容語(名詞・形容詞・副詞・動詞など)の語長を比較する。一方、機能語についてはどのグループ内でも大きな差は見られなかった。



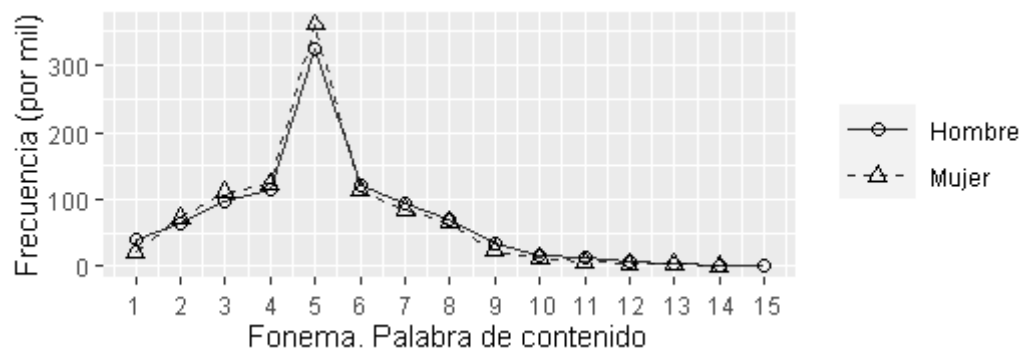
【図-13】 学年差 : A.1:低学年; A.2:中学年;A.3:高学年

全体的に低学年は最頻値より短い語が多用され、学年が上がるにつれて最頻値より長い語が多用されていく様子がわかる。



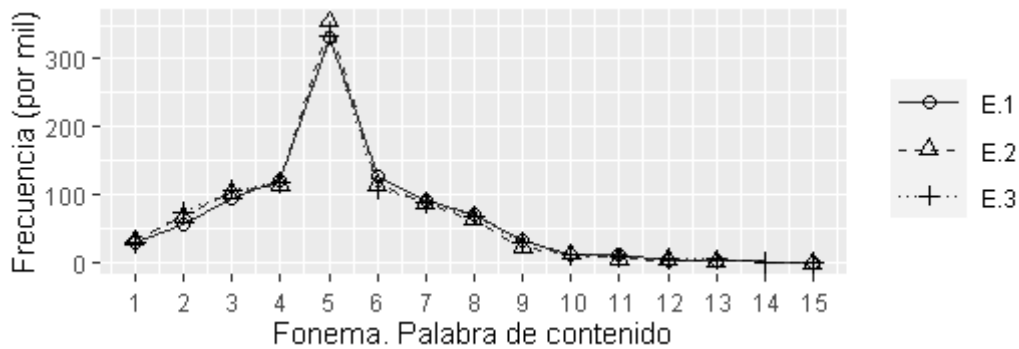
【図-14】 出版物 : Tea:演劇; Nov:小説; Ens:随筆; Pre:新聞雑誌; Tec:科学技術

内容語の語長は分野によって大きく変わる。演劇と小説の使用語の語長が短く、とくに演劇の語長が短い。随筆・新聞雑誌・科学技術文の使用語がかなり長い。



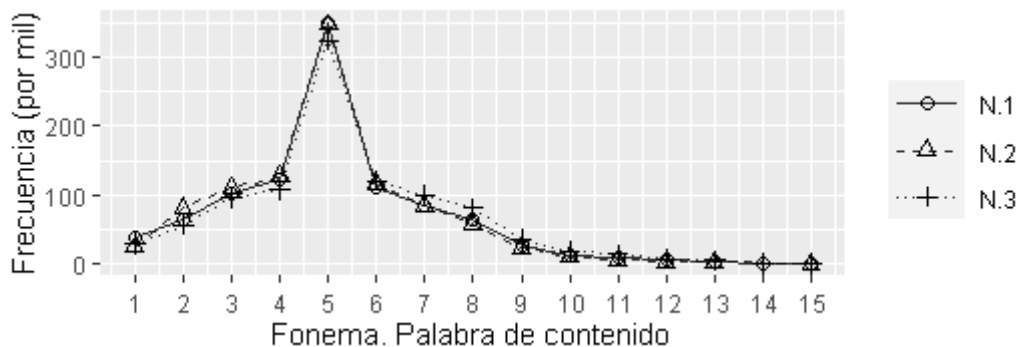
【図-15】 性差 : Hombre:男性; Mujer:女性

男性の使用語のほうがやや長い。女性は最頻値以下の長さの語をやや多く使い、男性は最頻値を超える長さの語をやや多く使う。



【図-16】 年齢差：E-1:若年；E-2:中年；E-3:高齢

年齢差はほとんど見られない。高齢者の使用語彙の語長が少し長い。



【図-17】 学歴差：N-1:低学歴；N-2:中学歴；N-3:高学歴

高学歴者の使用語彙の語長がやや長い。

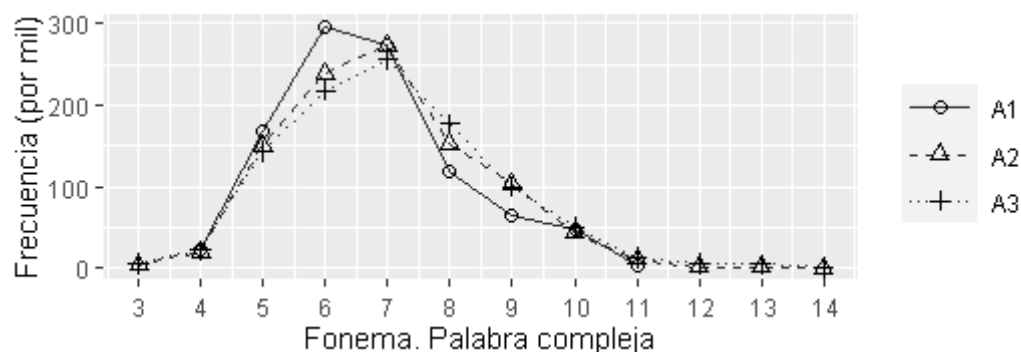
全体的に機能語はほとんど頻度分布に差がなかったのでグラフを作成しなかった。機能語は言語体系の中心としてどのような文体でもほとんど同じ比率・頻度で使われるのに対して、内容語は言語的・話題的変数に応じて自由に選択されるために変異差が生まれる、と考えられる。

内容語は小学生の学年別と出版物の種類別に高位のレベル(高学年・教養文体)の使用語彙の語長が長い。しかし、口語スペイン語のインタビュー録音資料(図-14, 15, 16)では内容語でもあまり変異が見られない。書き言葉の方が安定して変異が少ないことが予想されたが、事実逆であった。話し言葉は対話者との同調があるために、むしろ均一化すると考えられる。一方、書き言葉は個人の言語活動であるために、固有の文体が現れるのであろう。

2.3.3. 複合語

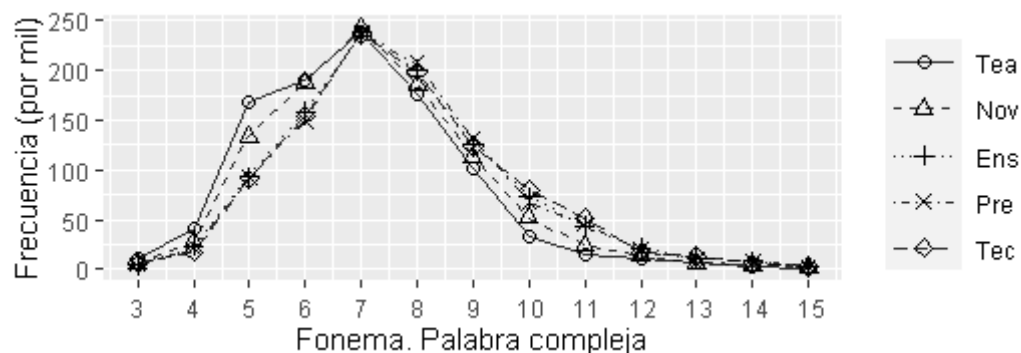
以下では3種資料を使って学年差・出版物・性差・年齢差・学歴差について

て各グループの使用語彙の語長をすべて内容語の複合語(派生語・合成語)に限って分析した。単一語については大きな頻度差が見られなかったのでグラフを作成しなかった。



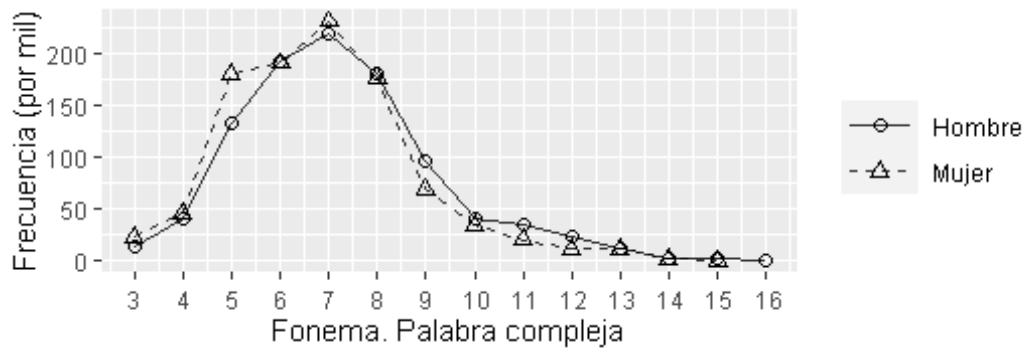
【図-18】 学年差 : A.1:低学年; A.2:中学年; A.3:高学年

最頻値の左の音素数が比較的低い領域では $A1 > A2 > A3$ の順になり、頂点の右の音素数が比較的高い領域では頻度数が逆転する。これは高学年になると音素数が増加することを示しているが、複合語の場合なので、単純に音素数の増加を要因とすることはできない。当然この逆転現象の要因は語形成の発達であると考えられる。



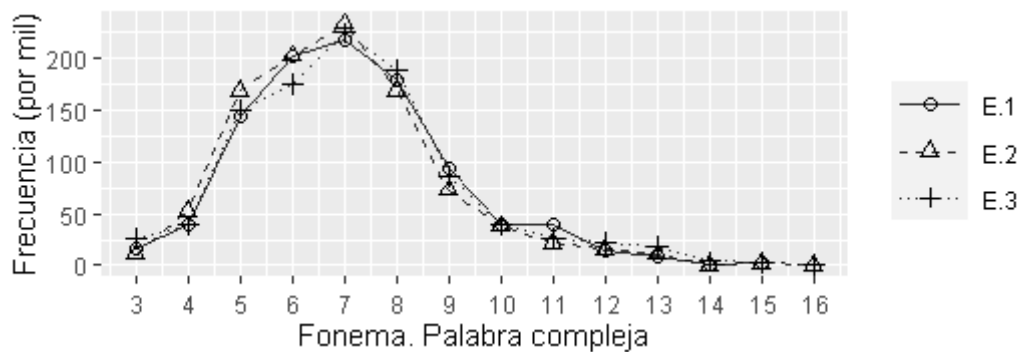
【図-19】 出版物 : Tea:演劇; Nov:小説; Ens:随筆; Pre:新聞雑誌; Tec:科学技術

分布の左側(音素数が少ない)では演劇の使用頻度が高く、右側(音素数が多い)では随筆・新聞雑誌・科学技術文の使用頻度が高くなっている。小説は両者の中間にある。



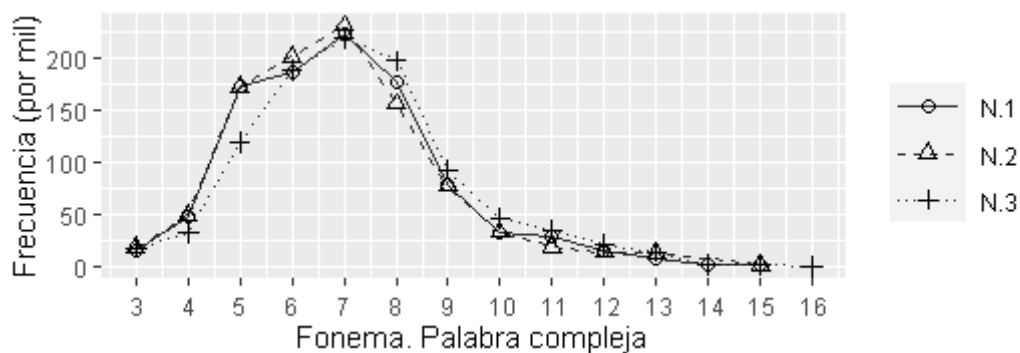
【図-20】 性差：Hombre:男性; Mujer:女性.

分布の左側で女性の使用語彙の頻度が高く，分布の右側で逆転して男性の使用語彙の頻度が高い。語形成の要因が考えられる。



【図-21】 年齢差：E-1:若年; E-2:中年; E-3:高齢.

年齢差の特徴がつかみにくい。中年の複合語の語長が短い。



【図-22】 学歴差：N-1:低学歴; N-2:中学歴; N-3:高学歴

分布の左側の音素数が少ない部分で高学歴の人の使用語彙の頻度が比較的少なく，逆に分布の右側では高学歴の人の使用語彙の頻度が比較的多くなる。よって，語形成の要因が考えられる。

全体を通して、もっとも頻度差がよく現れるのは、出版物で特に演劇と随筆・新聞雑誌・科学技術文の差が大きい。他のデータでは単一語では有意な差が見えず、差の多くは複合語で検知されるので、やはり語長の長さが頻度と逆関係になる要因は語形成にある、と考えられる。

2.4. まとめ

本章では、はじめに辞書と頻度辞典の見出し語の音節数・音素数と異なり語数を計測し、全体的に正規分布によく近似していることを観測し、その理由を統計的・言語的に考察した。次に、述べ語数について一般に言われる「短縮の法則」(「語長が長くなると使用頻度が小さくなる」「使用頻度が大きくなると語長は短縮する」)が成立するか否かを検証した。

たしかに、音節で語長を計測すれば使用頻度が L 字型分布を示すので「短縮の法則」は成り立つように見えるが、それは圧倒的高頻度の語長が短い機能語(冠詞・代名詞・前置詞・接続詞など)と比較的低頻度の語長が長い内容語(名詞・形容詞・副詞・動詞など)を分析データに混ぜたことによって生じたことであって、「語長が長くなると使用頻度が小さくなる」ことを証明してはいない。音節数の増加が要因となって頻度分布が単調に下降するわけではないからである。それぞれの音節数の語彙の実態を見ると、圧倒的高頻度の機能語の語長が短いために分布の左側に位置し、それに続いて比較的低頻度の内容語の語長が長いことが多いために頻度分布の右側に位置し、分布曲線は単調に下降しているように見えるのである。よって音節数に対する使用頻度分布の単調下降曲線は「短縮の法則」に起因するものではない。

このことはデータ全体を音節数を使って機能語と内容語に分離して分析すると、内容語の使用頻度は L 字型分布ではなく正規分布を示すこと、そして音節数でなく音素数を使って機能語と内容語に分離して分析すると、全体は L 字型分布ではなく I 型分布を示し、内容語が正規分布に近似することで確認することができた。音節数を使って語長を計測すると、頻度分布の最初の項である 1 音節語が 1 音素語と 2 音素語を含めたグループとなって、L 字型分布の重要な第 1 項が大雑把な内容になってしまう。音節のような大きな計測単位を使用すれば、大まかな観察となって正確な実態を示さない。

データ全体を見渡して「使用頻度が大きくなると語長は短縮する」という「短縮の法則」が成り立つ例はわずかである(*bicicleta* > *bici*, *colegio* > *cole*, *película* > *pele*, *zoológico* > *zoo*)¹²。もし「短縮の法則」が語形の頻度分布を決定するほどの一般的法則であるならば現代スペイン語で多数の例が見つか

¹² 他に語彙化して定着した例が初級レベルの参考書に載せられている：*fotografía* > *foto*, *motocicleta* > *moto*。

るはずである。

歴史的に見てもスペイン語の語尾は英語と比べると語尾の脱落・語形の短縮は少なかった。むしろ、逆に多用されている語が長くなった例も多い(lat. *edere* > *cum+edere* > esp. *comer*; lat. *auris* > *auris + cula* > *auricula* > esp. *oreja*, lat. *afflare* > esp. *hallar* → *encontrar* (後述 4.8), etc.)。このような語形の延長の要因の一つとして、短い語形が通信に不便なため一見不要に見える要素を付加することによって語形を補強されたことが考えられる。そして、一見「短縮の法則」を指示するような *trans* > *tras* > *tra*, *sub* > *so*, *illu(m)* > *elo* > *el* のような歴史的変化は、人間の行動一般に見られる多数回の行動を簡略化する最小努力の原則によって引き起こされたのではなく、機能語が無強勢であったため(内容語のように意味を表示するための強勢が不要であったため)、隣接する強勢語(内容語)の支えを受けてそれと一体化して音韻的に弱化したためである。このことは *ipse* > *ese*, *primero* > *primer* などの強勢のある機能語についても同様である。よって、スペイン語に関しては一般に(高頻度使用による)「短縮の法則」を認めることはできない。

一般言語学の著作では「短縮の法則」は語形の物理的な長さ(文字数・音節数・音素数)を使って機械的・数量的に論じられてきた。しかし、言語学的視点に立てば語形の長さは物理的にではなく、言語的・定質的に観察すべきものである。語形の言語的長さは派生形態論の単位(接頭辞・接尾辞など)を使って計測されるべきものである。

言語普遍の観点から見ればたしかに単一語と派生語の間に無標・有標の対立があるが(Greenberg, 1966 [2005]: 56), 使用頻度を考慮に入れるならば、無標・有標の二項対立ではなく、[0, 1]の範囲で有標性が漸進的に評価される。つまり、使用頻度の高い接辞は無標に近く、逆に使用頻度の低い接辞は有標性が高くなり、連続的な文体的価値が生まれる。このような分析を可能にするのは接辞の頻度調査である。実際的な計量文体論分析においては言語の理論的な対立だけでなく、実際的な使用に基づいた対立も考慮に入れなければならない。

たしかに物理的語長は重要な文体特徴の1つであり、物理的語長を使った研究成果の価値は否めない。よって、私たちの計量文体論でも音素を単位とした物理的語長を使用するが、それに限らず、さらに理論的・实际的形態論の見地からも文体的観察を続ける。語長を計量文体論分析の対象とするとき、単位は音素を使い、語は機能語と内容語を弁別すべきである。短い語が多用され、語が長くなるほど使用頻度が減少するのは機能語の語長は常に短く、非常に高頻度で使用されるためである。

機能語の語長が一般に短い理由は、機能語が内容語のように「意味」を担う要素ではなく、「文法的関係」を示す要素であるためである。意味を担う

ならばその意味に対応する相当な形式が必要であるから、自然に音素数が多くなる。さらに、意味を特定するための接辞を伴う派生語や意味を足し合わせる合成語の音素数も必然的に増加する。そして、内容語には意味単位を明示するために強勢が置かれるので、比較的強固に語形が保持される。一方、機能語の本質は文法的関係を示すことなので、その標識さえあれば十分である。強勢すら不要であることが多い。多くの機能語に強勢がない理由は、機能語は単独で出現することではなく、強勢のある内容語の前または後に付随して内容語(または内容語の連続)にその「働き」を指示するからである。そして、その結果、機能語は必然的に短い語形になる。*mediante, durante, unos, según*のように比較的長い機能語の使用頻度は、*el, la, a, de, en, con, por, que*などのよう単音節の機能語の使用頻度より低い。また、強勢がなければ音韻の弱화가起こり、その結果短縮されることも多い(lat. *illos > los, illud > lo, illam > ela > la, et > y, aut > o, etc.*)。

次に、同じ内容語の中で比較するならば、たしかに語が長くなるほど使用頻度が減少するが、語が長くなるということは、語が直線的に延長するのではなく、接頭辞・接尾辞・合成などによる語形成的構造変化によるものである。語形成によって語形が長くなる時、接辞や合成要素の意味が加重されて特殊化する。特殊な意味を持つ新語は一般的な意味を担う単一語よりも頻度が低くなることは当然である。「語が長くなるほど使用頻度が減少する」ということはこのような言語形態の形成過程を無視した表面的な観察である。たしかに、最小努力の法則・経済・コストなどの説明原理は否定できないが、それは言語の変化・変異の結果的効果であり、本質的要因ではない。言語現象の本質的要因は言語構造・体系の中で見出される。

2.3で全体の語長と社会言語学的変異の関係を見た。はじめに、機能語(冠詞・代名詞・前置詞・接続詞など)と内容語(名詞・形容詞・副詞・動詞など)を比較した。学年・出版物・性・年齢・学歴の変数を通じて、機能語の頻度分布はI字型を示し、内容語の頻度分布は正規性を示す。全体の頻度分布の変異は少ない。とくに機能語の変異は小さい。この理由は機能語が高頻度で使用され、その大部分が単一語(無標)であるためであろう。一方内容語は高頻度の単一語と比較的低頻度の複合語を含むため、機能語と比べて有標性が高く安定性に欠く。

次に、内容語を単一語(単一形)と複合語(複合形)に分離すると、学年・出版物・性・年齢・学歴の変異を通じて、単一語の頻度分布に大きな変異がないのに対し、複合語にはかなりの音素数に対する使用頻度に大きな変異が見られる。単一語に変異が少ない理由は単一語が語彙の基礎を構成する部分であるためであろう。最も分布が安定している機能語は言語体系の機能を担う部分であり、その大部分が単一語(無標)であるため極めて高頻度であり頻度の

変異が小さい。一方、内容語は高頻度の単一語(無標)と比較的低頻度である内容語(有標)を含むため、比較的低頻度であり頻度の変異が大きいと考えられる。

よって、頻度分布の変異を対象とする計量文体論では、内容語の中から複合語(formativo:有標)を抽出し、単一語(simple:無標)と比較しながら分析する、という方法を採用する。

3. 語形成

前章で、語形の延長の要因は低頻度使用ではなく、機能語(無標)よりも内容語(有標)のほうが長く、単一語(無標)よりも複合語(有標)が長くなる、という言語的理由によるものであることを確認したので、本章では内容語の中で複合語を取り上げ、その有標性の特徴を強く示す接頭辞・接尾辞・合成語を分析する。

語形成の観点から特徴語の集中分布を観察する。個体(派生語の接辞・合成語)と変数(学年・出版物・性・年齢・学歴)の組み合わせによって、分布が明確に区分される場合と分布が複雑に重なる場合がある。どちらの場合も変数の順番を固定して、個体をその数値の分布状態に従って並べ変えると分布の全体像が明らかになる。

3.1. 接頭辞

次は全資料(「8種資料」)の中で接頭辞を持つ語の見出し語数のリストである(数値は頻度降順)。

(15) 接頭辞のリスト：{co(n)-}, 245, conocer; {a(d)-}, 219, así; {i(n)-, en-}, 218, entonces; {re-}, 184, recibir; {ex-, e(s)-}, 130, existir; {des-, dis- [neg]}, 123, desde; {de-, di-}, 91, dentro; {pro-}, 70, producir; {i(n)- [neg]}, 63, enfermedad; {per(i)-}, 41, permitir; {pre-}, 41, presentar; {so(n)-, su(b)-}, 37, suponer; {o(b)-}, 33, ocurrir; {tra(n)s-, tra-}, 28, través; {inter-}, 21, interés; {ab(s)-}, 9, avanzar; {super-}, 6, superior; {di(a)-}, 5, diálogo; {se-}, 5, separar; {ant(e)-}, 4, antecedente; {entre-}, 4, entrevista; {extr(a.o)-}, 4, extraño; {sin-}, 4, síntoma; {para-}, 3, parámetro; {proto-}, 3, protagonista; {sobre-}, 3, sobrevivir; {a(n)- [neg]}, 2, anónimo; {cata-}, 2, catástrofe; {contra-}, 2, contradicción; {indo-}, 2, industria; {intro-}, 2, introducir; {per-}, 2, prestigio; {por-}, 2, porcentaje; {auto-}, 1, automáticamente; {bi-}, 1, bicicleta; {con-}, 1, contagio; {de-}, 1, delantero; {en-}, 1, enmarcar; {epi-}, 1, epidemia; {infra-}, 1, infraestructura; {mono-}, 1, monarca; {pan-}, 1, pandemia; {pen-}, 1, penumbra; {retro-}, 1, retroceder; {semi-}, 1, semifinal; {tra(n)s-}, 1, tranvía; {vice-}, 1, vicepresidente.

次は派生語の中の接頭辞のリストとその頻度、および次の式を適用した「対数最大値比」(Lmr: Log maximum ratio)である。

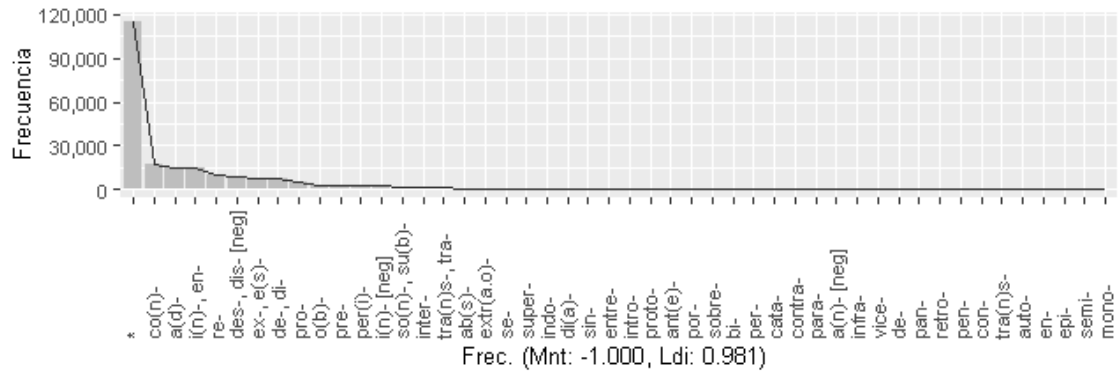
$$\text{Lmr} = \log(\text{頻度}) / \log(\text{頻度の最大値: base})$$

たとえば $Lmr('a(d)-')$ は $\log(25625) / \log(211800) = 0.828$ となる。対数をとることによって、頻度の小さな値の最大値比を求める感度が高くなる。たとえば、最初の接頭辞 a の最大値比 $25625/211800=0.121$ と比較すると対数を使う利点が見える。対数最大値比(Lmr)は頻度が単一語の頻度に近いほど 1 に近づくので、これで頻度の大きさを共通のスケールで計算することができる。

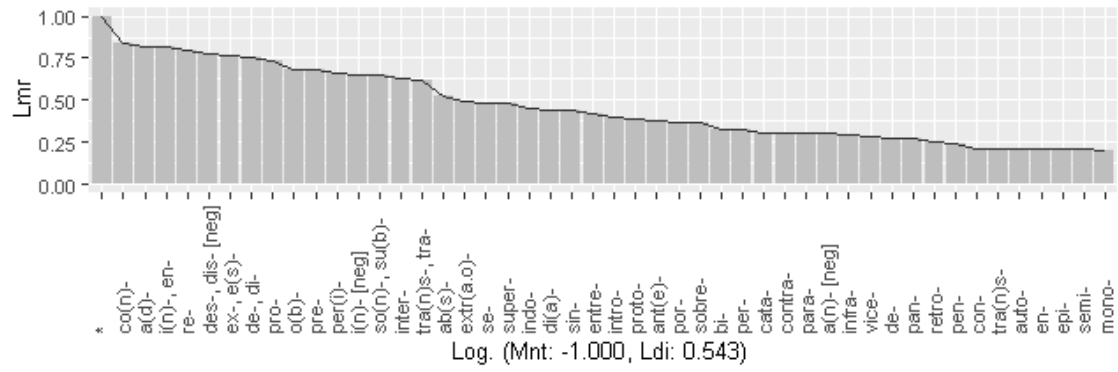
(16) 接頭辞の使用頻度(Freq)と対数最大値比(Lmr)

Ini	Freq	Lmr	Ini	Freq	Lmr
*	115,512	1.000	intro-	107	0.401
co(n)-	16,840	0.835	proto-	94	0.390
a(d)-	14,576	0.822	ant(e)-	86	0.382
i(n)-, en-	14,540	0.822	por-	76	0.372
re-	10,482	0.794	sobre-	71	0.366
des-, dis- [neg]	8,238	0.773	bi-	45	0.327
ex-, e(s)-	7,308	0.763	per-	42	0.321
de-, di-	7,052	0.760	cata-	37	0.310
pro-	5,197	0.734	contra-	37	0.310
o(b)-	3,001	0.687	para-	37	0.310
pre-	2,840	0.682	a(n)- [neg]	35	0.305
per(i)-	2,139	0.658	infra-	33	0.300
i(n)- [neg]	2,082	0.655	vice-	27	0.283
so(n)-, su(b)-	1,877	0.647	de-	26	0.279
inter-	1,616	0.634	pan-	23	0.269
tra(n)s-, tra-	1,351	0.618	retro-	19	0.253
ab(s)-	471	0.528	pen-	17	0.243
extr(a.o)-	302	0.490	con-	12	0.213
se-	273	0.481	tra(n)s-	12	0.213
super-	270	0.480	auto-	11	0.206
indo-	203	0.456	en-	11	0.206
di(a)-	182	0.446	epi-	11	0.206
sin-	164	0.437	semi-	11	0.206
entre-	137	0.422	mono-	10	0.198

上のリストを見ると、頻度が高い接頭辞の種類は少なく、頻度が低い接頭辞の種類が多いことがわかる。使用頻度は典型的な L 字型分布を示し、対数最大値比(Lmr)はその頻度を比較する際の不都合を緩和する。



【図-23】接頭辞の使用頻度



【図-24】接頭辞の対数最大値比

次に、接頭辞と各変数(学年・出版物・性・年齢・学歴)との関係を観察する。

Prefijo. Año

	A1	A2	A3
bi-	124	83	60
entre-	12	4	6
a(n)- [neg]	0	3	1
tra(n)s-	0	1	2
cata-	0	1	2
di(a)-	0	1	2
sin-	0	0	2
intro-	0	0	4
indo-	0	9	7
sobre-	0	2	6
super-	0	1	6
se-	0	3	8
de-	0	1	8
extr(a.o)-	6	8	15
per(i)-	51	42	59
tra(n)s-, tra-	4	10	27
inter-	4	12	29
ab(s)-	8	21	42
so(n)-, su(b)-	8	22	57
i(n)- [neg]	22	55	76
pre-	43	49	83
o(b)-	18	32	86
pro-	59	143	262
ex-, e(s)-	113	200	306
de-, di-	365	344	470
re-	182	272	403
des-, dis- [neg]	231	387	581
i(n)-, en-	547	652	842
a(d)-	419	660	900
co(n)-	1159	1098	1456

Prefijo. Publicación

	Tea	Nov	Ens	Pre	Tec
a(n)- [neg]	1	28	4	4	3
entre-	13	7	7	5	3
proto-	1	8	8	1	0
cata-	3	7	12	3	3
retro-	2	4	1	3	0
sobre-	0	3	1	1	1
con-	1	2	0	3	0
pen-	0	2	2	3	0
tra(n)s-	3	4	1	7	0
di(a)-	1	6	3	2	4
contra-	0	0	8	5	5
per-	1	3	2	7	5
sin-	2	7	15	9	14
ab(s)-	12	45	46	44	28
intro-	3	1	3	2	14
indo-	0	2	13	27	22
se-	16	17	15	9	29
super-	6	22	45	30	37
tra(n)s-, tra-	21	58	119	99	111
inter-	58	73	139	162	150
i(n)- [neg]	121	245	260	240	222
pre-	159	205	238	314	235
pro-	127	241	445	489	477
o(b)-	203	188	299	299	485
des-, dis- [neg]	500	753	679	774	821
re-	500	722	1111	1212	883
ex-, e(s)-	281	428	699	666	796
de-, di-	337	484	617	709	924
i(n)-, en-	740	990	1120	1126	1232
co(n)-	951	1090	1576	1651	1614

【図-25】接頭辞 (Prefijo) : 学年 (Año) / 出版物 (Obra)

学年(Año): 低学年(A.1)に集中している接頭辞は少ない(bi-, entre-)。全体的に学年と相関して使用頻度が上昇している。

出版物(Obra): 全体の接頭辞の頻度は作品の配列順(Tea - Nov - Ens - Pre - Tec)に沿っている。この作品の順番は日常文体→教養文体の推移を示すように思われる。学年(Año)の傾向と一部類似している(re-, ex-, de-, di-, i(n)-, en-, co(n)-)。

Prefijo. Sexo

	H	M
a(d)-	1041	806
ex-, e(s)-	356	161
re-	387	296
co(n)-	922	895
de-, di-	325	279
pro-	136	85
per(i)-	108	77
pre-	108	81
i(n)- [neg]	89	55
so(n)-, su(b)-	71	28
ab(s)-	43	23
tra(n)s-, tra-	33	25
super-	14	8
extr(a.o)-	7	2
por-	5	0
se-	9	8
cata-	3	0
di(a)-	3	0
auto-	2	0
indo-	3	4
sin-	2	4
contra-	0	6
proto-	0	6
entre-	5	9
bi-	7	13
o(b)-	60	62
inter-	33	49
des-, dis- [neg]	384	400
i(n)-, en-	1001	1008

Prefijo. Edad

	E-1	E-2	E-3
i(n)-, en-	1144	849	1051
de-, di-	327	315	274
o(b)-	112	35	47
des-, dis- [neg]	416	360	403
inter-	74	38	17
so(n)-, su(b)-	74	30	52
pre-	109	81	100
i(n)- [neg]	74	83	62
ab(s)-	42	25	35
entre-	16	8	0
tra(n)s-, tra-	29	35	22
super-	13	15	5
cata-	6	0	0
bi-	10	18	2
auto-	3	0	0
indo-	3	5	2
contra-	0	8	0
di(a)-	0	5	0
por-	3	0	5
proto-	3	0	5
sin-	0	3	5
se-	6	8	10
extr(a.o)-	0	3	10
pro-	106	93	134
per(i)-	70	93	112
ex-, e(s)-	263	189	336
re-	314	317	394
co(n)-	839	917	956
a(d)-	906	801	1073

Prefijo. Nivel

	N-1	N-2	N-3
de-, di-	367	185	362
ab(s)-	40	32	28
indo-	5	5	0
se-	8	11	6
auto-	3	0	0
di(a)-	3	3	0
bi-	3	26	0
contra-	3	3	3
extr(a.o)-	3	5	6
entre-	5	8	8
por-	3	0	6
cata-	0	0	6
super-	11	8	14
proto-	0	0	8
sin-	0	0	8
tra(n)s-, tra-	34	16	37
so(n)-, su(b)-	50	48	54
per(i)-	95	77	110
o(b)-	61	40	85
i(n)- [neg]	34	93	93
inter-	21	21	82
pre-	76	77	136
pro-	71	71	198
des-, dis- [neg]	367	365	447
re-	338	275	424
ex-, e(s)-	248	182	365
co(n)-	883	835	1016
i(n)-, en-	947	975	1098
a(d)-	910	801	1087

【図-26】 接頭辞 (Prefijo) : 性 (Sexo) / 年齢 (Edad) / 学歴 (Nivel)

接頭辞と性・年齢・学歴の相関はほとんど見えない。高齢者(E3)は super をほとんど使わない¹³。

3.2. 接尾辞

次は全資料(「8 種資料」)の中で接尾辞を持つ語・語数・使用頻度のリス

¹³ 現在では接尾辞 -ísimo に代わって接頭辞 super-, hiper-, mega-が多用されるようになったようである(García Mouton 2003: 85)。

トである¹⁴。

(17) 接尾辞のリスト : {-ar [vb]}, 346, gustar; {-(c.s.t.x)ión}, 334, relación; {-a(l,r)}, 169, social; {-(a.e.ie)nte}, 120, durante; {-o [na]}, 115, trabajo; {(d.t)ad}, 103, ciudad; {-mente}, 101, realmente; {-ico.ac}, 93, político; {-(a.e)ncia}, 82, experiencia; {-to}, 80, punto; {-(d.s.t)or}, 73, sector; {-a [na]}, 53, cuenta; {-ario}, 53, necesario; {-ivo}, 50, objetivo; {-ble}, 48, posible; {-ero}, 42, primero; {-oso}, 40, religioso; {-miento}, 36, movimiento; {-ura}, 33, cultura; {-ia}, 32, historia; {-ía}, 31, mayoría; {-ado}, 28, estado; {-e [na]}, 28, informe; {-ano}, 27, humano; {-or}, 26, valor; {-io}, 24, medio; {-ón}, 24, millón; {-ecer [vb]}, 23, parecer; {-mento}, 22, momento; {-ez(a)}, 20, naturaleza; {-izar [vb]}, 20, realizar; {-ada}, 18, mirada; {-ista}, 17, artista; {-so}, 16, proceso; {-aje}, 14, viaje; {-ino}, 14, vecino; {-tud}, 14, actitud; {-ica.ac}, 13, práctica; {-ida}, 13, medida; {-illa}, 13, rodilla; {-ta}, 13, vista; {-eta}, 12, peseta; {-ir [vb]}, 12, partir; {-ismo}, 12, organismo; {-ear [vb]}, 11, desear; {-era}, 11, carrera; {-ido.ac}, 11, rápido; {-ificar [vb]}, 11, significar; {-illo}, 11, sencillo; {-ina}, 11, cocina; {-ito.ac}, 11, éxito; {-culo.ac}, 10, artículo; {-imo.ac}, 10, último; {-ior}, 10, mejor; {-orio}, 10, territorio; {-anza}, 9, esperanza; {-eo.ac}, 9, contemporáneo; {-il.ac}, 9, difícil; {-(e.i)nt(a.e) [num]}, 8, veinte; {-ato}, 8, candidato; {-ería}, 8, galería; {-iano}, 7, cristiano; {-ido}, 7, sentido; {-iva}, 7, iniciativa; {-sis.ac}, 7, análisis; {-eno}, 6, terreno; {-eo}, 6, europeo; {-és}, 6, francés; {-ío}, 6, frío; {-itar}, 6, necesitar; {-ito.dim}, 6, bonito; {-men}, 6, régimen; {-ce}, 5, doce; {-erio}, 5, ministerio; {-ete}, 5, paquete; {-il}, 5, civil; {-uo}, 5, antiguo; {-án}, 4, alemán; {-ar [sus]}, 4, centenar; {-el}, 4, cartel; {-icar}, 4, publicar; {-ín}, 4, calcetín; {-aña}, 3, campaña; {-az}, 3, capaz; {-eño}, 3, brasileño; {-ienda}, 3, vivienda; {-ito}, 3, propósito; {-monio}, 3, matrimonio; {-o}, 3, uruguayo; {-ula.ac}, 3, célula; {-undo}, 3, segundo; {-ana}, 2, ventana; {-azgo}, 2, hallazgo; {-do}, 2, contenido; {-edia}, 2, comedia; {-enda}, 2, leyenda; {-endo}, 2, tremendo; {-ense}, 2, estadounidense; {-icia}, 2, justicia; {-icio}, 2, servicio; {-ie}, 2, nadie; {-ísimo}, 2, muchísimo; {-logía}, 2, metodología; {-no}, 2, materno; {-ón.pers}, 2, campeón; {-una}, 2, fortuna; {-(d)umbre}, 1, incertidumbre; {-0 [na]}, 1, perdón; {-al}, 1, cordial; {-ancio}, 1, cansancio; {-anda}, 1, propaganda; {-ante}, 1, adelante; {-aria}, 1, maquinaria; {-avo}, 1, octavo; {-aza}, 1, terraza; {-cia}, 1, farmacia; {-

¹⁴ ac は接尾辞に先行する強勢, .aum は増大辞, dim は縮小辞, na は動詞を短縮した動作名詞(sumar>suma), pers は「人」を示す。

cilio}, 1, domicilio; {-cipio}, 1, principio; {-cula. ac}, 1, molécula; {-cula.ac}, 1, partícula; {-dad}, 1, efectividad; {-dote}, 1, sacerdote; {-eco}, 1, muñeco; {-eda}, 1, vereda; {-ego}, 1, gallego; {-ela}, 1, novela; {-ella}, 1, botella; {-ello}, 1, cabello; {-elo}, 1, caramelo; {-ena}, 1, docena; {-ento}, 1, violento; {-eo. ac}, 1, sanguíneo; {-esco}, 1, gigantesco; {-estre}, 1, terrestre; {-í}, 1, israelí; {-ico. ac}, 1, quirúrgico; {-ida.ac}, 1, pérdida; {-iño}, 1, cariño; {-ira}, 1, mentira; {-ive}, 1, inclusive; {-ja}, 1, aguja; {-logo.ac}, 1, diálogo; {-metro.ac}, 1, parámetro; {-mienta}, 1, herramienta; {-mo}, 1, supremo; {-ol}, 1, español; {-ondo}, 1, redondo; {-oro}, 1, sonoro; {-ota}, 1, pelota; {-plice.ac}, 1, cómplice; {-pulo.ac}, 1, discípulo; {-ua}, 1, estatua; {-uelo}, 1, pañuelo; {-ulo.ac}, 1, capítulo; {-uno}, 1, alguno; {-urno}, 1, nocturno; {-uto}, 1, diminuto; {-za}, 1, fuerza.

次は派生語の中の接尾辞のリストとその頻度, および対数最大値比を示す。

(18) 接尾辞の頻度(Freq)と対数最大値比(Lmr)¹⁵

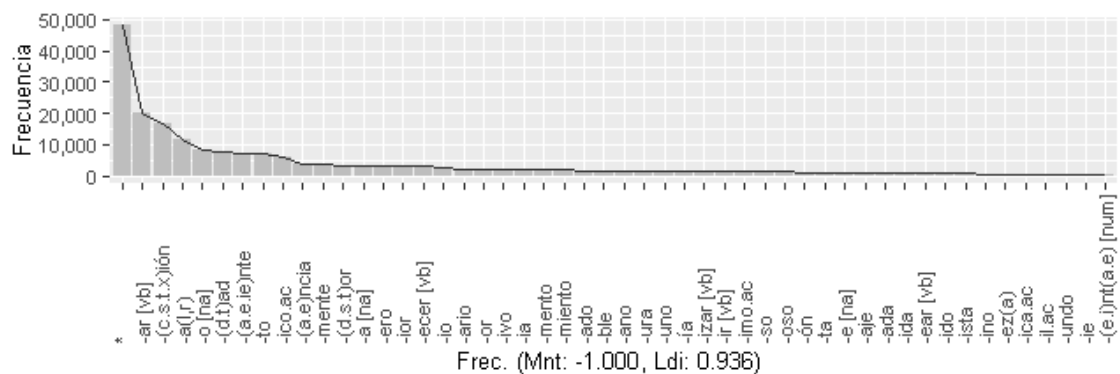
	Fin	Freq	Lmr	Fin	Freq	Lmr
	*	48,220	1.000	-il	258	0.515
-ar [vb]	19,847	0.918	-iano	257	0.515	
-(c.s.t.x)ión	16,524	0.901	-az	244	0.510	
-a(l,r)	11,688	0.869	-iva	242	0.509	
-o [na]	8,227	0.836	-eo.ac	237	0.507	
-(d.t)ad	7,432	0.827	-án	230	0.504	
-(a.e.ie)nte	7,118	0.823	-aña	221	0.501	
-to	7,000	0.821	-ana	202	0.492	
-ico.ac	6,223	0.810	-ula.ac	161	0.471	
-(a.e)ncia	3,878	0.766	-ito	151	0.465	
-mente	3,567	0.759	-monio	144	0.461	
-(d.s.t)or	3,360	0.753	-ería	143	0.460	
-a [na]	3,298	0.751	-ela	135	0.455	
-ero	3,203	0.749	-ienda	135	0.455	
-ior	3,148	0.747	-ete	133	0.454	
-ecer [vb]	2,956	0.741	-icia	132	0.453	
-io	2,552	0.727	-ante	122	0.445	
-ario	2,330	0.719	-do	109	0.435	
-or	2,309	0.718	-ón.pers	108	0.434	

¹⁵ * : 接尾辞のない見出し語。

-ivo	2,134	0.711	-ida.ac	91	0.418
-ia	2,018	0.706	-ense	87	0.414
-mento	1,939	0.702	-ín	82	0.409
-miento	1,925	0.701	-ulo.ac	81	0.408
-ado	1,723	0.691	-el	80	0.406
-ble	1,702	0.690	-eño	70	0.394
-ano	1,663	0.688	-ar [sus]	69	0.393
-ura	1,635	0.686	-edia	68	0.391
-uno	1,622	0.685	-logo.ac	68	0.391
-ía	1,620	0.685	-enda	66	0.389
-izar [vb]	1,598	0.684	-mienta	65	0.387
-ir [vb]	1,369	0.670	-ísimo	64	0.386
-imo.ac	1,346	0.668	-endo	62	0.383
-so	1,346	0.668	-ota	62	0.383
-oso	1,284	0.664	-ella	54	0.370
-ón	1,155	0.654	-una	51	0.365
-ta	974	0.638	-o	50	0.363
-e [na]	944	0.635	-eco	49	0.361
-aje	917	0.633	-ira	49	0.361
-ada	822	0.622	-iño	48	0.359
-ida	762	0.615	-ello	46	0.355
-ear [vb]	761	0.615	-ento	45	0.353
-ido	722	0.610	-mo	45	0.353
-ista	678	0.605	-azgo	44	0.351
-ino	653	0.601	-dote	43	0.349
-ez(a)	638	0.599	-0 [na]	42	0.347
-ica.ac	635	0.598	-logía	33	0.324
-il.ac	635	0.598	-no	33	0.324
-undo	623	0.597	-ondo	32	0.321
-ie	604	0.594	-cilio	29	0.312
-(e.i)nt(a.e) [num]	594	0.592	-estre	28	0.309
-culo.ac	575	0.589	-urno	28	0.309
-ito.ac	547	0.585	-ancio	26	0.302
-era	542	0.584	-cia	26	0.302
-ificar [vb]	540	0.583	-cula.ac	25	0.298
-ol	502	0.577	-anda	24	0.295
-ina	498	0.576	-avo	24	0.295

-icio	481	0.573	-aza	24	0.295
-eta	474	0.571	-uelo	24	0.295
-itar	469	0.570	-ua	23	0.291
-tud	455	0.568	-(d)umbre	22	0.287
-illo	443	0.565	-esco	21	0.282
-és	438	0.564	-oro	21	0.282
-ido.ac	404	0.557	-ive	20	0.278
-sis.ac	404	0.557	-ja	20	0.278
-eo	397	0.555	-ena	19	0.273
-ismo	388	0.553	-pulo.ac	19	0.273
-anza	383	0.552	-aria	18	0.268
-icar	372	0.549	-elo	17	0.263
-ito.dim	364	0.547	-ego	15	0.251
-orio	329	0.537	-cula. ac	14	0.245
-ío	313	0.533	-ico. ac	14	0.245
-men	310	0.532	-metro.ac	14	0.245
-erio	309	0.532	-uto	14	0.245
-za	306	0.531	-eda	13	0.238
-ce	297	0.528	-eo. ac	13	0.238
-uo	296	0.528	-plice.ac	13	0.238
-ato	295	0.527	-í	12	0.230
-cipio	287	0.525	-dad	11	0.222
-illa	281	0.523	-al	10	0.214
-eno	265	0.517			

接頭辞と同様に接尾辞でも L 字型の頻度分布を示している。接尾辞は多数なので、高頻度の上位 50 までをグラフに含める：



【図-27】接尾辞の使用頻度

次に、接尾辞の相対頻度と各種変数の関係を見る。

Sufijo. Año				Sufijo. Publicación					
	A1	A2	A3		Tea	Nov	Ens	Pre	Tec
-eta	1045	676	326	-ito.dim	150	50	9	16	2
-ana	567	402	205	-ira	32	6	4	0	1
-eco	328	100	43	-ana	5	44	21	8	2
-ota	188	75	67	-o [na]	14	7	0	2	0
-ete	132	52	35	-ela	19	43	82	14	16
-eo	130	71	40	-ella	3	9	0	0	0
-iño	0	0	8	-anda	1	1	11	3	0
-ato	0	4	11	-enda	7	1	33	8	9
-cipio	0	2	12	-ienda	5	3	3	22	1
-avo	0	7	14	-eño	5	9	3	24	1
-anza	0	11	18	-eta	18	17	11	47	6
-ismo	0	7	17	-avo	0	0	25	14	21
-uo	0	8	23	-azgo	0	2	9	0	19
-ido.ac	0	22	35	-ato	5	4	9	52	3
-és	12	14	43	-il	3	30	26	66	15
-icar	0	7	38	-ito.ac	25	25	51	85	15
-izar [vb]	4	14	44	-men	6	17	71	38	50
-miento	2	18	48	-ista	25	38	100	125	11
-il.ac	2	26	51	-orio	1	12	33	25	57
-imo.ac	8	21	56	-icar	14	13	50	92	80
-(a.e)ncia	10	18	57	-ula.ac	2	10	12	17	69
-so	10	53	72	-ulo.ac	2	0	11	4	68
-ano	18	52	93	-ol	32	56	189	175	117
-mente	12	24	90	-ivo	37	34	170	170	226
-ía	36	73	120	-ura	63	123	262	130	332
-ir [vb]	24	50	150	-imo.ac	49	110	173	226	341
-(a.e.ie)nte	134	283	388	-ico.ac	157	252	718	657	618
-(c.s.t.x)ión	150	335	645	-(a.e)ncia	135	236	494	420	592
-a(l,r)	269	622	743	-a(l,r)	283	465	1073	1101	1088
-(d.t)ad	130	391	721	-(c.s.t.x)ión	430	756	1675	1722	1994

【図-28】接尾辞 (Sufijo) : 学年 (Año) / 出版物 (Obra)

低学年に多い接尾辞は-eta, -ana, -eco, -ota, -ete. これらは日常接尾辞である。教養接尾辞である-izar, -miento, -imo.ac, -(a.e)ncia, -mente, -(a.e.ie)nte, (c.s.t.x)ión, (d.t)adは学年とともに使用頻度が上昇している。

Tea, Nov に多い接尾辞は-ito.dim, -iraであり, Ens, Pre, Tec に多い接尾辞は-ista, -ismo, -torio, -ivo, -imo.ac, -ico.ac, -(c.s.t.x)iónである。これらの多くは教養接尾辞である。学年と出版物の推移傾向が類似している。

Sufijo. Sexo	Sexo		Sufijo. Edad	Edad			Sufijo. Nivel	Nivel		
	H	M		E-1	E-2	E-3		N-1	N-2	N-3
-(c.s.t.x)ión	606	327	-(a.e.ie)nte	522	237	319	-ce	90	114	28
-mente	633	374	-ón	131	58	25	-il.ac	87	11	68
-uno	246	138	-il.ac	77	73	20	-0 [na]	24	5	6
-(a.e)ncia	108	57	-e [na]	61	10	20	-sis.ac	26	0	14
-ble	76	25	-ano	54	13	40	-ón.pers	18	0	0
-ario	76	30	-miento	58	10	50	-ante	18	3	8
-ivo	65	26	-és	35	38	5	-ondo	13	0	0
-miento	55	19	-eta	22	18	2	-cilio	8	0	0
-ta	45	15	-ida.ac	10	3	0	-ela	5	0	0
-és	41	8	-ísimo	32	76	30	-ive	5	0	0
-itar	45	19	-ela	6	0	0	-és	11	58	6
-e [na]	38	17	-ive	6	0	0	-ella	0	11	0
-0 [na]	21	2	-ulo.ac	6	0	0	-az	0	8	0
-ienda	21	2	-undo	10	20	2	-án	0	5	0
-illo	22	8	-ismo	0	18	0	-monio	0	13	3
-za	15	2	-az	0	8	0	-ato	0	11	3
-ante	15	4	-án	0	5	0	-al	0	0	6
-ana	9	0	-ificar [vb]	0	10	5	-ento	0	0	6
-ato	9	0	-monio	0	10	5	-iano	3	0	8
-ondo	9	0	-al	0	0	5	-ito.ac	8	0	14
-ella	7	0	-ar [sus]	0	0	5	-culo.ac	0	5	14
-ete	7	0	-ento	0	0	5	-ina	34	95	40
-il	7	0	-ina	35	93	37	-iva	11	0	20
-erio	0	6	-ato	3	0	10	-ista	5	11	25
-monio	2	9	-illa	0	0	10	-men	0	3	28
-aña	2	13	-ón.pers	0	5	12	-miento	45	11	59
-il.ac	33	79	-anza	3	15	25	-izar [vb]	13	50	59
-ísimo	21	76	-endo	3	13	27	-ano	13	32	59
-ce	48	111	-ica.ac	35	10	47	-ear [vb]	37	32	93
-ido	41	147	-ería	42	18	67	-ia	69	66	153

【図-29】接尾辞 (Sufijo) : 性 (Sexo) / 年齢 (Edad) / 学歴 (Nivel)

- (c.s.t.x)ión, -mente は男性(H)に多く, ísimo は女性(M)に多い。-miento, -zar, -ear は特に N3 に多い。

3.3. 合成語

Comp. Año

	A1	A2	A3
televisión.n	253	95	70
sacapuntas.n	233	141	39
teléfono.n	57	30	18
mariposa.n	36	42	12
cumpleaños.n	28	14	3
triángulo.n	16	8	2
doscientos.num	2	8	2
agricultor.n	2	20	6
psicólogo.n	0	0	3
diecisiete.num	0	3	4
superficie.n	0	2	4
veinticuatro.num	0	2	4
diecinueve.num	0	2	4
agricultura.n	0	5	5
beneficio.n	0	1	4
geografía.n	0	1	4
trescientos.num	0	3	5
uniforme.a	0	5	6
siquiera.adv	0	1	5
universo.n	0	0	5
mantener.v	0	4	7
jamás.adv	0	3	7
veinticinco.num	2	4	8
novecientos.num	2	4	8
atmósfera.n	0	3	8
oxígeno.n	0	4	9
delante.adv	0	9	11
mamífero.a	4	12	14
tampoco.adv	8	6	16
edificio.n	8	13	21
cualquiera.ind	0	7	22
ninguno.ind	12	22	35
autobús.n	14	46	42
baloncesto.n	2	16	33
vosotros.pro	2	18	62
ahora.adv	22	38	80
fútbol.n	32	39	97
aunque.conj	6	22	101
bolígrafo.n	120	387	253
nosotros.pro	18	96	191

Comp. Publicación

	Tea	Nov	Ens	Pre	Tec
ahora.adv	235	95	118	91	50
asimismo.adv	49	3	10	1	3
maldito.a	21	2	0	2	0
vosotros.pro	19	2	10	2	0
delante.adv	18	20	5	7	5
bendito.a	13	4	1	2	0
teléfono.n	12	2	4	1	0
veintidós.num	11	1	2	1	3
automóvil.a	5	15	7	4	0
fotografía.n	1	7	1	4	0
universitario.a	2	4	11	4	0
autónomo.a	0	0	5	2	1
equivaler.v	0	1	8	3	2
norteamericano.a	0	1	2	5	0
filosófico.a	1	1	12	7	1
psicología.n	0	0	9	2	3
municipio.n	0	0	5	8	0
filosofía.n	4	3	23	10	6
agricultura.n	0	0	1	8	3
vanguardia.n	0	0	3	10	0
democrático.a	0	1	8	13	1
dieciocho.num	1	2	3	13	3
mediodía.n	0	4	0	11	5
beneficio.n	2	1	3	15	1
democracia.n	0	0	5	15	0
psicológico.a	0	1	8	1	8
filósofo.n	3	3	9	21	2
biblioteca.n	1	7	16	2	12
edificio.n	0	7	12	26	6
hipótesis.n	1	2	4	13	11
diecisiete.num	2	0	2	8	11
universal.a	0	2	23	7	14
diecinueve.num	1	0	2	9	12
universidad.n	6	6	40	31	29
análogo.a	1	0	11	9	23
superficie.n	1	4	3	2	41
geográfico.a	0	0	2	4	49
microscopio.n	0	2	0	1	77
geografía.n	1	0	3	1	89
península.n	1	1	10	11	128

【図-30】 合成語 (Comp) : 学年 (Año) / 出版物 (Obra)

学年を見ると、*televisión, sacapuntas, teléfono, mariposa, cumpleaños* は学年が上がるにつれ次第に頻度が減少し、逆に *baloncesto, vosotros, ahora, fútbol, aunque, bolígrafo, nosotros* は次第に頻度が上昇する。この中で話題語(名詞)を除いて *aunque, ahora, nosotros* が高学年(A3)の文体特徴であると思われる。

出版物では、演劇: *ahora, asimismo, maldito, vosotros*, 小説: *automóvil*, 随筆: *filosofía, filosófico*, 新聞雑誌: *vanguardia, democrático, agricultura, beneficio, democracia, biblioteca*, 科学技術文: *geográfico, microscopio,*

geografía, península に使用頻度が集中している。この大部分は話題語である。

Comp. Sexo		Comp. Edad			Comp. Nivel					
	H	M	E-1	E-2	E-3	N-1	N-2	N-3		
todavía.adv	60	30	ninguno.ind	115	116	52	autobús.n	37	40	0
edificio.n	21	2	aunque.conj	67	35	27	dieciséis.num	21	5	0
fútbol.n	26	13	cualquiera.ind	48	33	15	autónomo.a	13	0	0
universidad.n	14	8	edificio.n	35	3	2	atmósfera.n	11	0	0
televisión.n	12	6	tampoco.adv	83	161	77	delante.adv	13	5	3
primavera.n	9	0	autobús.n	38	10	32	diecisiete.num	13	5	3
centímetro.n	9	0	universidad.n	22	3	10	primavera.n	11	0	3
delante.adv	10	4	psicología.n	16	0	0	fotografía.n	8	0	0
atmósfera.n	7	0	centímetro.n	13	0	2	vosotros.pro	8	0	0
veinticinco.num	7	0	psicológico.a	10	0	0	biblioteca.n	8	0	0
autónomo.a	7	2	veintidós.num	13	3	7	dieciocho.num	8	11	0
siquiera.adv	5	0	veinticinco.num	10	3	0	superficie.n	5	0	0
fotografía.n	5	0	arquitecto.n	6	0	0	municipio.n	5	0	0
psicológico.a	5	0	biología.sus	6	0	0	uniforme.a	5	0	0
península.n	5	0	siquiera.adv	6	0	2	arquitecto.n	5	0	0
baloncesto.n	5	2	biblioteca.n	6	0	2	siquiera.adv	5	3	0
superficie.n	3	0	aeropuerto.n	6	0	2	teléfono.n	5	0	3
arquitecto.n	3	0	baloncesto.n	6	0	5	simple.a	3	0	0
psicólogo.n	3	0	equivocar.v	3	5	0	beneficio.n	3	0	0
veinticuatro.num	3	0	península.n	3	5	0	aeropuerto.n	0	8	0
novcientos.num	3	0	tecnología.n	3	8	0	filosofía.n	0	3	0
autopista.n	3	0	década.n	0	5	0	península.n	0	5	3
biología.sus	3	0	superficie.n	0	5	0	bolígrafo.n	0	0	3
beneficio.n	2	0	uniforme.a	0	5	0	universitario.a	0	0	3
simple.a	0	2	autopista.n	0	5	0	equilibrio.n	0	0	3
bolígrafo.n	0	2	atmósfera.n	0	8	2	baloncesto.n	0	5	6
década.n	0	4	dieciocho.num	0	13	5	mediodía.n	3	0	6
municipio.n	0	4	sacrificio.n	0	0	5	década.n	0	0	6
uniforme.a	0	4	psicólogo.n	0	0	5	sacrificio.n	0	0	6
sacrificio.n	0	4	novcientos.num	0	0	5	psicólogo.n	0	0	6
tecnología.n	2	6	teléfono.n	0	0	7	biología.sus	0	0	6
teléfono.n	0	6	fotografía.n	0	0	7	tecnología.n	3	0	8
democracia.n	0	6	democracia.n	0	0	7	democracia.n	0	0	8
diecisiete.num	5	9	mediodía.n	0	0	7	psicológico.a	0	0	8
veintidós.num	5	9	diecinueve.num	0	5	10	centímetro.n	3	0	11
psicología.n	0	9	delante.adv	3	5	12	psicología.n	0	3	11
autobús.n	17	36	primavera.n	0	0	12	veinticinco.num	0	0	11
aunque.conj	33	51	jamás.adv	0	3	15	edificio.n	5	11	20
nosotros.pro	67	93	kilómetro.n	0	0	17	universidad.n	3	11	20
también.adv	227	387	nosotros.pro	26	103	97	mantener.v	16	8	31

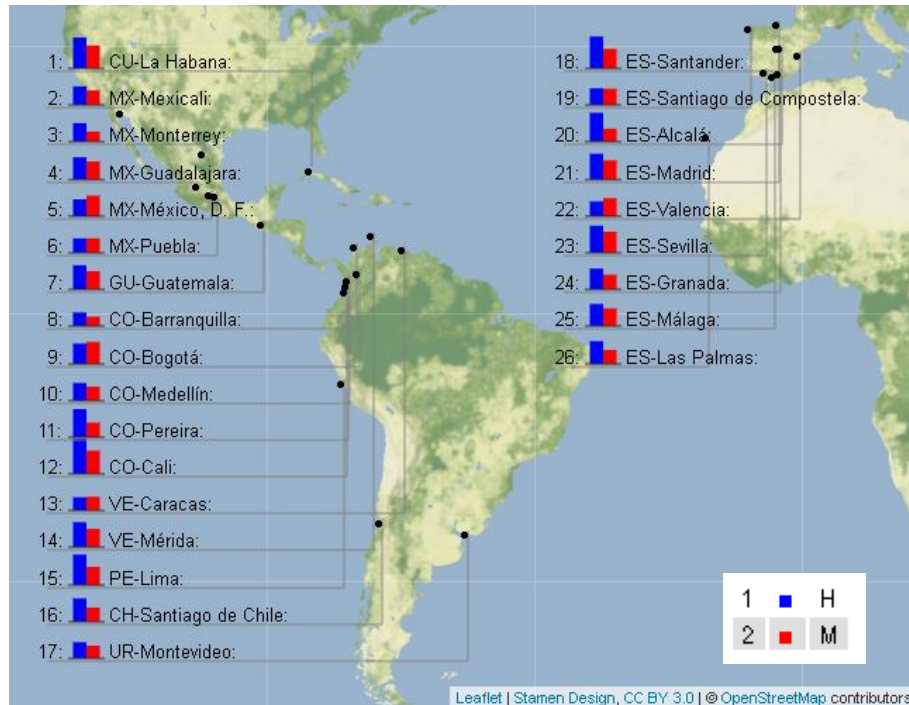
【図-31】 合成語 (Comp) : 性 (Sexo) / 年齢 (Edad) / 学歴 (Nivel)

jamás は頻度は少ないが、高齢者(E3)に多いのでその文体特徴でありうる。

3.4. 地域差

調査範囲を拡大するためには、たとえば Santander に見られた集中分布がスペイン・ラテンアメリカの各地でも同様に観察されるか調査する方法が考えられる。以下の図は私たちの PRESEEA en LYNEAL サイトを使って、使

用頻度を地図上にプロットしたグラフである¹⁶。先の観察によれば(3.3), *-mente* は男性に多く, *-ísimo* は女性(M)に多い。*-miento* は特に N3 に多い。*jamás.adv* は高齢者(E3)に多い。これらの変異をスペイン語圏全体の言語地図の中で確認する¹⁷。

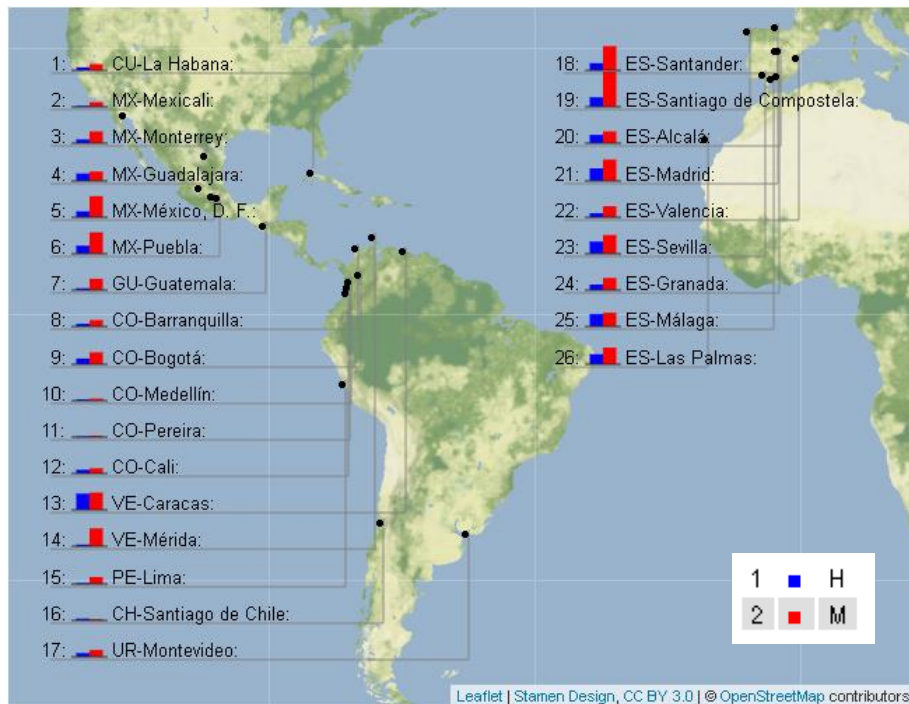


【図-32】 *mente* (計:13,078) Sexo: H ~ M

このように, *-mente* (H)は, ほぼスペイン語圏全体で確認される。

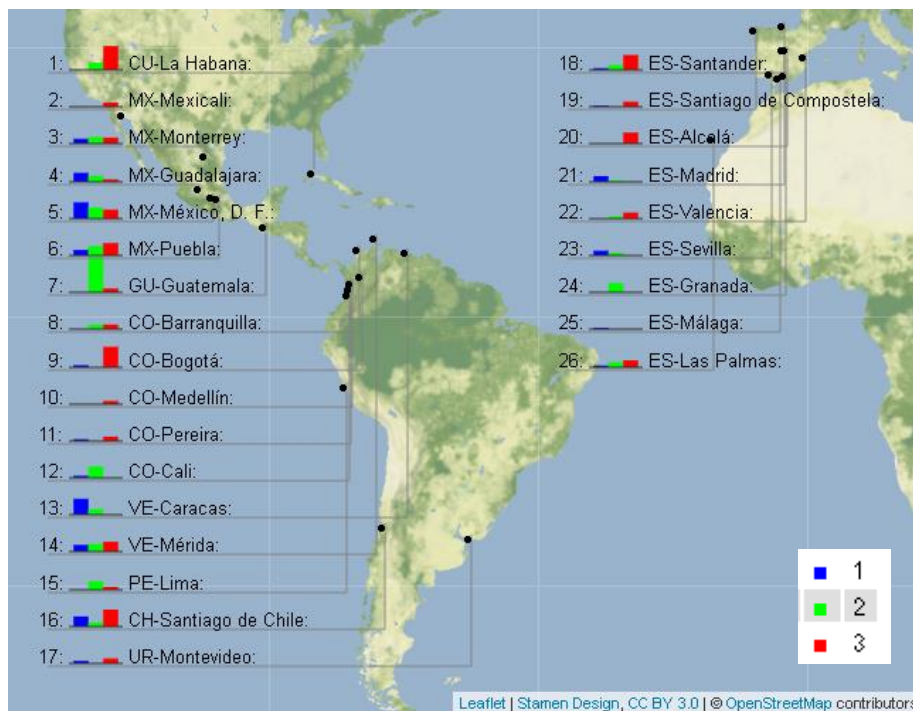
¹⁶ <https://h-ueda.sakura.ne.jp/lyneal/preseea.htm>

¹⁷ 地図のプロットは数値を比較をするため 10 万語による正規化頻度とした。



【図-33】 -ísimo[oa](s) (計:2,333) Sexo: H ~ M

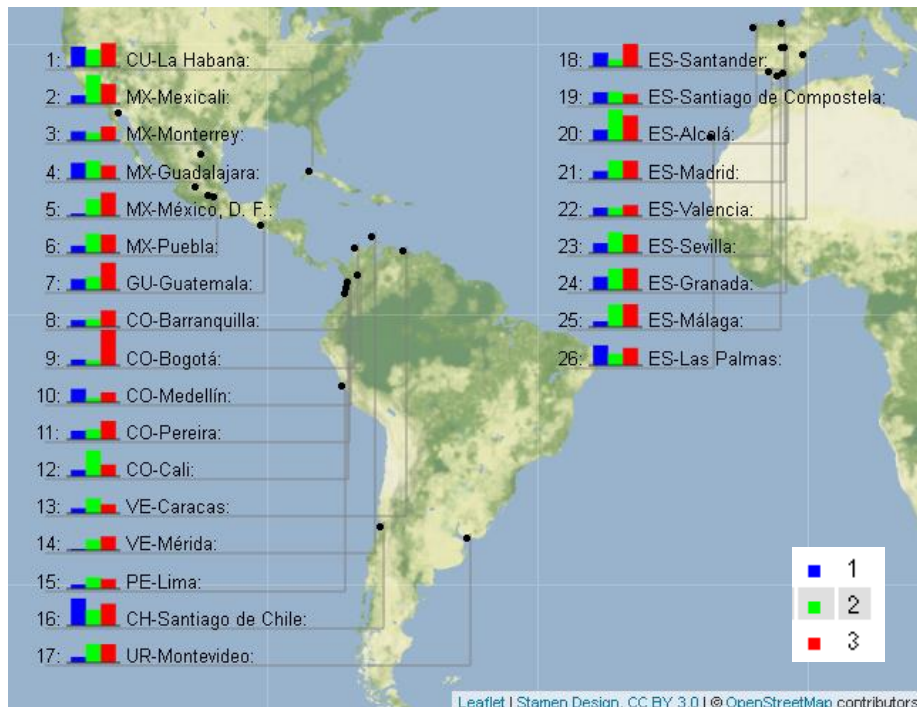
-ísimo (M)は、ほぼスペイン語圏全体で確認される。強調語の使用は女性の話言葉の特徴であるように思われる。



【図-34】 jamás (計:212) Edad: E1 ~ E3

jamás (E3)は北スペインの文体特徴であってラ米では E1-E2 にも分布して

いるので、衰退の通時的変化が地理的分布に反映しているように見える。3種の資料をまとめた「8種資料」が示す使用頻度によれば *nunca* (587) に対し *jamás* (90) であってかなり低い。*jamás* はフォーマルな文体で使用される有標的な語である。



【図-35】 -miento(s) (計:2,343) Nivel: N1 ~ N3

-miento (N2-N3)はほぼスペイン語圏全域に見られるが、SantanderではN3が顕著である。この接尾辞は学年の推移で[A1:6 ~ A2:23 ~ A3:56]という学習過程を示している。よって歴史的には二重母音化という民衆的文体の特徴を帯びているが、現代のスペイン語では教養的接尾辞として動作名詞の形成のために使用されている(*alejar* > *alejamiento*, *vencer* > *vencimiento*)。これに対し、同語源の*-mento*はラテン語から直接移入され、語彙化された事物を示す教養的名詞を形成する(*apartamento*, *departamento*: Lang 1992: 172)。動詞的な*-miento*は文体特徴となるが、*-mento*は具体的な事物を示すので話題語となり特徴語とはならない。

3.5. 年代差

先に資料の説明で見たように、私たちの資料(「8種資料」)では、1. García Hoz (1953), 2. Juilland and Chang-Rodríguez (1964), 3. 上田 (1987), 4. Justicia (1995), 5. Davies (2006), 6. Martínez y Ueda (2021a, 2013), 7. Real Academia Española (2023)の頻度分布を照合してある。その中の1. García Hoz (1953)と

2. Juilland and Chang-Rodríguez (1964) は 20 世紀前半のスペインの書き言葉を扱い、5. Davies (2006) は 20 世紀後半 (1970-2000) のスペイン・ラテンアメリカの話し言葉・書き言葉を収集し、7. Real Academia Española (2023) は 21 世紀 (2001-2023) のスペイン・ラテンアメリカの話し言葉・書き言葉を含めている。このように、それぞれの収集範囲は異なるが、とくに年代差が顕著である。ここで、3 つの年代を区別して 4 つの資料を比較して 1 つの通時的変遷を辿ることにする。現在では年代ごとの語彙(見出し語)の総体を、地域・年代・文体を揃えた同じ基盤で厳密に比較するための資料はできていないので、これは試験的・探索的な試みに過ぎない。

次の左図は接頭辞の変遷を示し、右図は接尾辞の変遷を示している(数値は 10 万語あたりの正規化頻度)。

Prefijo. s.20-21

	GJ	Davies	CORPES
extr(a.o)-	46	34	29
a(n)- [neg]	10	2	3
mono-	3	0	1
tra(n)s-	3	2	1
pen-	1	3	1
epi-	1	0	2
en-	0	0	2
semi-	0	0	2
bi-	1	1	3
pan-	0	0	4
de-	1	0	5
vice-	0	0	5
infra-	0	0	6
sobre-	2	8	9
para-	0	0	7
intro-	4	13	12
proto-	3	4	16
entre-	7	10	19
di(a)-	4	10	22
se-	21	26	34
ab(s)-	32	57	48
inter-	136	152	189
so(n)-, su(b)-	155	195	209
tra(n)s-, tra-	89	125	166
pre-	271	240	337
pro-	432	472	606
ex-, e(s)-	639	693	831
re-	969	993	1199
co(n)-	1525	1538	1868
i(n)-, en-	1129	1362	1643

Sufijo. s.20-21

	GJ	Davies	CORPES
-ior	451	308	292
-ol	112	41	43
-ica.ac	89	110	33
-tud	81	53	40
-és	79	52	36
-eta	61	23	28
-erio	53	21	36
-iano	47	23	25
-ce	45	28	20
-ito.dim	40	26	15
-ísimo	13	9	0
-iño	12	4	4
-dote	11	5	3
-avo	7	0	3
-anda	7	2	2
-pulo.ac	6	2	1
-al	4	1	1
-cula.ac	0	5	2
-eño	9	2	10
-logía	0	0	6
-o	2	2	8
-mienta	1	3	10
-do	5	4	17
-ense	0	0	17
-e [na]	54	78	124
-izar [vb]	85	131	200
-ivo	154	162	277
-(d.s.t)or	290	241	423
-ico.ac	514	499	768
-a(l,r)	1070	924	1405

【図-36】接頭辞の変遷 / 接尾辞の変遷

この図では接頭辞・接尾辞の使用について、とくに大きな変化の傾向を見出すことはできないが、形容詞の接尾辞*-ior*, *-ol*, *-ica.ac*, *-tud*, *-és*の使用が次第に減少し、逆に、*-izar*, *-ivo*, *-(d.s.t)or*, *-ico.ac*, *-a(l,r)*の使用が増大していることが注目される。前者は生産的ではなく語彙的に定着して安定しているのに対し、後者は生産的・活動的な接尾辞であるようだ。

3.6. 文体差

Moreno (2005: C-ORAL-ROM)と Martínez y Ueda (2021a, 2013)は 20 世紀後半(1990-)の話し言葉の録音資料であり、García Hoz (1953)と Juilland and

Chang-Rodríguez (1964)は 20 世紀前半の書き言葉の語彙の集計である(→1.3)。このように、厳密には半世紀以上の年代差があるが、(年代差を含めた)文体差を接頭辞と接尾辞で比較すると次のようになる。

Prefijo. Estilo

	L.hablada	L.escrita
ab(s)-	42	32
di(a)-	12	4
contra-	5	2
bi-	4	1
infra-	2	0
para-	2	0
mono-	0	3
tra(n)s-	0	3
cata-	2	5
per-	2	5
por-	3	7
ant(e)-	4	10
sin-	3	10
a(n)- [neg]	2	10
indo-	8	29
super-	7	36
extr(a.o)-	6	46
inter-	97	136
so(n)-, su(b)-	98	155
per(i)-	119	200
i(n)- [neg]	111	210
pre-	141	271
o(b)-	130	296
pro-	236	432
ex-, e(s)-	353	639
des-, dis- [neg]	488	739
de-, di-	410	715
a(d)-	1129	1341
re-	496	969
co(n)-	1068	1525

Sufijo. Estilo

	L.hablada	L.escrita
-(e.i)nt(a.e) [num]	195	80
-mo	0	8
-dote	1	11
-ela	2	20
-icia	3	22
-ato	9	36
-za	11	37
-án	8	39
-men	11	42
-uo	14	46
-iano	13	47
-erio	8	53
-icar	12	56
-ino	28	73
-ez(a)	21	70
-és	26	79
-culo.ac	27	82
-tud	21	81
-ol	29	112
-imo.ac	71	169
-ario	94	197
-oso	68	195
-ado	67	197
-ia	99	253
-or	101	255
-(d.s.t)or	134	290
-ero	200	387
-(a.e)ncia	168	374
-o [na]	400	726
-(c.s.t.x)ión	815	1553

【図-37】 文体差. 接頭辞 / 接尾辞

この相対頻度の集計によれば、接頭辞では re-と co(n)-が書き言葉に多い。先の【図-35】によれば、20 世紀前半(GJ)と後半(Davies)において re-, co(n)-の頻度差はほとんど検知されなかった。よって、かなりの信頼度を持って re-と co(n)-が書き言葉に多い、と言える(ただし GJ/Davies の地域差を考慮していない)。

接尾辞については-(a.e)ncia, -o [na], -(c.s.t.x)ión が書き言葉に多い。これらの接尾辞は 3.5 の分析では地域差を示さなかった(図に表示されていない)。よって、これらも書き言葉の特徴であると考えられる。

4. 語彙

先述のように(1.4, 1:5, 2.3.2, 2.3.3), 語彙は大きく文法的観点から機能語と内容語に区分され, 語形成的観点から単一語と複合語に区分される。それぞれ対照的に分布・種類数・頻度が大きく異なる。

4.1. 機能語

Fun. Año

	A1	A2	A3
mi.pos	1384	615	553
yo.pro	976	331	403
mí.pro	527	214	180
qué.interrog	689	1016	342
tal.ind	0	5	16
mediante.prep	0	3	16
demasiado.ind	0	7	18
mas.conj	0	0	19
cualquiera.ind	0	7	22
durante.prep	0	10	23
vario.ind	0	22	33
cuál.interrog	4	11	35
bastante.ind	0	12	37
demás.ind	0	20	47
menos.ind	6	35	60
mismo.ind	14	29	63
tras.prep	4	10	58
vosotros.pro	2	18	62
según.prep	0	10	69
bajo.prep	14	26	80
sino.conj	4	23	80
algo.ind	2	27	83
entre.prep	12	26	88
aquel.dem	10	54	100
aunque.conj	6	22	101
pues.conj	24	53	120
contra.prep	0	15	110
cada.ind	12	44	130
desde.prep	18	35	138
nuestro.pos	24	77	156
hasta.prep	36	78	165
poco.ind	30	89	171
dónde.interrog	51	112	184
nosotros.pro	18	96	191
sin.prep	16	73	214
o.conj	26	171	289
como.conj	142	345	453
nos.pro	132	476	614
todo.ind	217	500	767
que.conj	519	767	2430

Fun. Publicación

	Tea	Nov	Ens	Pre	Tec
no.sn	3354	1140	1164	891	916
lo.pro	2779	1674	992	974	860
me.pro	1962	563	243	67	107
usted.pro	1354	262	38	36	2
yo.pro	1338	324	165	47	52
te.pro	1100	148	24	20	13
ese.dem	932	375	421	284	173
estar.v	991	606	368	383	362
mi.pos	751	323	211	47	95
tú.pro	540	77	13	2	1
pero.conj	775	329	321	223	293
sí.sn	533	79	25	40	53
si.conj	600	248	195	190	204
tu.pos	301	61	8	11	1
pues.conj	339	61	86	75	125
cómo.interrog	318	101	57	121	38
porque.conj	359	152	157	118	115
nada.ind	304	107	84	54	38
quién.interrog	161	32	16	10	3
nadie.ind	135	52	32	19	18
cuando.conj	106	92	1	0	1
os.pro	53	6	16	13	3
cuánto.interrog	25	14	2	3	0
vosotros.pro	19	2	10	2	0
alguien.ind	15	17	9	1	1
vos.pro	7	0	0	0	0
tras.prep	4	18	11	5	1
mediante.prep	0	1	3	3	10
contra.prep	20	30	58	70	21
ambos.ind	0	6	27	33	39
hacia.prep	20	74	41	22	64
durante.prep	9	34	41	64	62
aquel.dem	66	339	161	167	184
sino.conj	53	56	168	82	109
cuyo.rel	4	33	62	91	116
según.prep	18	27	45	54	115
cuál.interrog	35	40	122	79	172
nuestro.pos	96	73	317	252	266
o.conj	173	181	423	242	613
de.prep	4295	8386	9769	10879	10767

【図-38】機能語 (Fun) : 学年 (Año) / 出版物 (Obra)

低学年の機能語の使用語彙の特徴として 1 人称単数形が高頻度であることが挙げられる(*mi, yo, mí*)。高学年になると, 1 人称複数形の頻度が高くなる(*nos, nosotros, nuestro*)。否定の前置詞 *sin*, 接続詞 *aunque, o, como que* の使用頻度も高い。これらは話題依存性が低いので, 文体特徴であると考えられる。

出版物については，1~2 人称の語(*me, usted, tú, te, tu, yo, mi, os*)が演劇・小説・随筆に多く，新聞雑誌・科学技術文では少ない。全体的にほとんどの機能語の特徴語は演劇・小説・随筆に集中するが，わずかに *o.conj* だけが科学技術文に多い。これらは話題語ではなく，文体特徴であると考えられる。

Fun. Sexo			Fun. Edad			Fun. Nivel				
	H	M		E-1	E-2	E-3		N-1	N-2	N-3
un.art	2597	1750	o.conj	1185	1216	767	o.conj	1189	1137	792
para.prep	664	532	más.ind	955	632	705	te.pro	818	922	589
otro.ind	454	276	poco.ind	468	292	229	tanto.ind	182	106	102
poco.ind	373	259	bastante.ind	247	53	47	cierto.ind	108	8	23
alguno.ind	246	138	algo.ind	227	106	174	uno.ind	134	69	79
algo.ind	194	132	cierto.ind	112	28	15	demás.ind	124	45	82
tú.pro	208	153	ninguno.ind	115	116	52	hasta.prep	134	161	93
bastante.ind	126	83	sin.prep	109	65	65	entre.prep	84	48	40
usted.pro	105	64	cada.ind	109	131	57	nadie.ind	40	40	20
cierto.ind	71	21	aunque.conj	67	35	27	alguien.ind	37	16	23
sobre.prep	76	47	tu.pos	83	91	47	donde.conj	29	11	8
aquel.dem	62	28	cualquiera.ind	48	33	15	cual.rel	32	16	20
entre.prep	67	47	durante.prep	32	3	2	cada.ind	121	56	119
nuestro.pos	55	36	alguien.ind	42	23	15	vario.ind	18	0	6
ti.pro	36	23	tampoco.adv	83	161	77	contra.prep	18	3	6
cual.rel	29	15	cuándo.interrog	10	0	2	quien.rel	18	3	8
hacia.prep	24	11	ambos.ind	3	0	0	bajo.prep	11	0	3
durante.prep	17	4	vario.ind	6	13	5	vosotros.pro	8	0	0
contra.prep	12	6	bajo.prep	3	8	2	cuándo.interrog	8	3	0
bajo.prep	7	2	según.prep	0	25	7	durante.prep	16	3	14
demasiado.ind	2	6	os.pro	0	20	10	os.pro	8	24	0
cuál.interrog	9	15	ti.pro	16	48	22	ambos.ind	3	0	0
os.pro	7	15	contra.prep	10	3	15	ante.prep	3	0	0
cuánto.interrog	12	23	cual.rel	13	33	20	tras.prep	3	0	0
quién.interrog	14	25	entre.prep	32	93	42	cuál.interrog	5	24	6
aunque.conj	33	51	hacia.prep	10	18	25	ti.pro	16	40	34
sin.prep	67	89	donde.conj	6	13	27	cuánto.interrog	11	13	28
nosotros.pro	67	93	cuanto.rel	10	10	30	quién.interrog	13	13	31
demás.ind	71	98	cuánto.interrog	6	10	32	cuanto.rel	11	8	34
desde.prep	89	117	nuestro.pos	29	33	72	aquel.dem	24	63	51
uno.ind	77	113	su.pos	115	229	147	nada.ind	235	555	241
mí.pro	193	255	nosotros.pro	26	103	97	nuestro.pos	29	32	79
también.adv	227	387	aquel.dem	3	23	102	bastante.ind	113	56	150
mucho.ind	507	697	dónde.interrog	45	58	137	tampoco.adv	69	127	133
mi.pos	659	863	qué.interrog	199	340	319	sobre.prep	32	34	125
o.conj	812	1301	tal.ind	54	224	269	él.pro	222	373	328
me.pro	1414	1750	también.adv	205	320	364	nos.pro	227	219	345
yo.pro	1161	1730	él.pro	237	274	394	este.dem	464	346	563
sí.sn	1947	2349	este.dem	327	360	650	pues.conj	1382	1118	1752
ser.v	3200	4056	ese.dem	570	879	1250	sí.sn	1466	2582	2386

【図-39】機能語 (Fun) : 性 (Sexo) / 年齢 (Edad) / 学歴 (Nivel)

機能語の使用について大きな性差は見つからない。年齢差については，指示語(*aquel, este, ese*)の使用が高齢者(E3)に多い。しばしば人名や特定の名詞がすぐに思い出せないときに指示語が使われている。このことは日本語でも

同様であるように思われる。高学歴者(N3)は *quién, cuánto, cuanto, sobre* を多く使用する。これらの機能語は話題語ではなく、文体特徴であると考えられる。

4.2. 内容語

Con. Año

	A1	A2	A3
estuche.n	429	269	67
mamá.n	340	58	23
muñeco.n	328	100	43
abuelo.n	318	91	47
ayer.n	273	19	16
papá.n	233	51	19
maceta.n	235	84	32
ratón.n	207	72	21
cuento.n	192	42	7
dibujo.n	211	79	32
lobo.n	199	82	14
pato.n	164	59	22
otoño.n	144	31	22
circo.n	75	23	8
regalar.v	59	3	14
pala.n	51	12	5
molino.n	47	8	1
ja.interjec	24	0	1
caso.n	0	0	18
contrario.a	0	0	18
practicar.v	0	1	22
clásico.a	0	0	23
asignatura.n	0	2	24
general.a	0	3	25
bajo.a	2	2	33
jugador.n	0	8	35
costumbre.n	0	2	35
equipo.n	2	8	39
barba.n	2	7	39
amistad.n	0	8	40
brazo.n	0	0	37
soler.v	0	12	47
contaminación.n	0	20	53
aire.n	0	1	53
uña.n	0	0	55
importante.a	0	16	64
contaminar.v	0	15	83
vida.n	6	36	107
deporte.n	12	24	111
etcétera.n	2	274	291

Con. Publicación

	Tea	Nov	Ens	Pre	Tec
señor.n	538	149	38	190	54
dios.n	230	38	29	27	10
ahí.adv	118	8	18	8	0
señorito.n	111	15	3	10	0
gustar.v	115	28	22	5	2
callar.v	107	31	8	9	5
ah.interjec	84	9	0	1	0
marido.n	84	20	5	7	1
ja.interjec	55	10	0	0	0
jurar.v	46	9	0	0	0
oh.interjec	46	3	8	1	0
mamá.n	40	2	0	1	2
ay.interjec	1	102	4	0	1
sindicato.n	1	0	4	33	0
toro.n	7	3	3	59	2
ministro.n	6	8	13	86	3
apertura.n	0	0	0	4	30
longitud.n	0	0	2	1	32
mencionar.v	1	2	2	3	40
célula.n	0	0	2	2	39
matemático.a	1	0	3	4	41
gobierno.n	11	8	47	132	2
curva.n	0	8	5	0	47
matemática.n	1	1	4	0	48
geográfico.a	0	0	2	4	49
lente.n	1	2	0	1	50
reino.n	3	1	5	7	59
investigación.n	0	2	10	9	63
capítulo.n	2	0	11	4	68
objetivo.a	2	0	9	4	79
método.n	1	0	8	14	83
microscopio.n	0	2	0	1	77
región.n	0	4	24	17	96
árabe.a	0	3	1	12	89
geografía.n	1	0	3	1	89
descripción.n	0	1	7	5	107
península.n	1	1	10	11	128
grupo.n	2	19	27	24	161
cultura.n	0	3	50	11	162
vaso.n	1	4	2	0	170

【図-40】 内容語 (Con) : 学年 (Año) / 出版物 (Obra)

内容語の学年差は顕著である。低学年(A1) : *estuche, mamá, muñeco, abuelo*. 高学年(A3) : *etcétera, deporte, vida, contaminar*. 出版物でも次のように明確

に区別される。演劇：*señor, dios, ahí, señorit(o,a)*。小説：*ay*。随筆：なし。新聞雑誌：*sindicato, toro, ministro*。科学技術文：*curva, matemática, ..., cultura, vaso*。特徴的な内容語は、ほとんどが話題性のある名詞だが、唯一、間投詞 *ay* (Nov)が文体特徴として認められる。

Con. Sexo	H M		Con. Edad	E-1 E-2 E-3			Con. Nivel	N-1 N-2 N-3		
	H	M		E-1	E-2	E-3		N-1	N-2	N-3
eh.interjec	-1424	599	leer.v	-173	20	12	mil.num	-74	8	11
normalmente.adv	-107	25	libro.n	-141	8	20	crío.n	-55	11	0
caso.n	-84	21	dinero.n	-112	18	32	marchar.v	-47	5	3
exactamente.adv	-67	13	realmente.adv	-90	10	22	genial.a	-34	3	0
intentar.v	-55	8	intentar.v	-77	23	7	joder.v	-34	3	3
cerca.adv	-50	8	película.n	-48	15	0	paso.n	-29	0	3
marchar.v	-36	0	edificio.n	-35	3	2	negocio.n	-26	0	0
inglés.a	-29	4	avión.n	-32	0	0	patrón.n	-18	0	0
temperatura.n	-29	4	básicamente.adv	-32	0	0	jo.interjec	-45	296	6
diferencia.n	-26	2	apunte.n	-29	0	0	copa.n	-16	0	0
absoluto.a	-22	0	matemático.a	-29	3	0	coste.n	-16	0	0
claramente.adv	-22	0	loco.a	-22	0	2	pff.interjec	-32	116	17
regalo.n	-22	0	cantar.v	-19	0	0	bicicleta.n	-3	26	0
escuela.n	-19	0	coste.n	-19	0	0	inglés.a	-3	45	3
plaza.n	-17	0	deuda.n	-19	0	0	fruta.n	-0	19	0
salsa.n	-17	0	italiano.a	-19	0	0	ti.interjec	-0	19	0
apunte.n	-15	0	actor.a	-16	0	0	bautizar.v	-0	16	0
etcétera.n	-15	0	psicología.n	-16	0	0	arroz.n	-0	53	6
humedad.n	-15	0	sábana.n	-16	0	0	premio.n	-0	0	17
marcar.v	-15	0	tradición.n	-16	0	0	procesión.n	-0	0	17
base.n	-14	0	final.a	-38	174	25	turístico.a	-0	0	20
madera.n	-14	0	primo.a	-6	30	0	celebrar.v	-0	0	23
ocasión.n	-14	0	carne.n	-0	25	0	posiblemente.adv	-0	0	23
posiblemente.adv	-14	0	ti.interjec	-0	18	0	religioso.a	-0	0	23
favorito.a	-0	17	genial.a	-0	33	2	matemático.a	-0	3	25
móvil.a	-0	21	jo.interjec	-22	277	35	apunte.n	-0	0	25
genial.a	-2	25	administrador.n	-0	0	17	primaria.n	-0	0	25
huevo.n	-2	25	consciente.a	-0	0	17	interesante.a	-0	5	31
cocer.v	-2	26	kilómetro.n	-0	0	17	humano.a	-3	0	31
precioso.a	-2	26	otoño.n	-0	0	20	accidente.n	-0	0	31
santo.a	-2	30	cultural.a	-0	5	37	participar.v	-0	0	34
serie.n	-7	42	santo.a	-0	3	40	procurar.v	-0	8	40
tío.n	-15	74	nacer.v	-6	5	45	dedicar.v	-0	3	48
ay.interjec	-14	77	maestro.a	-0	0	42	maestro.a	-0	0	48
libro.n	-10	93	efectivamente.adv	-0	18	57	santo.a	-0	0	48
leer.v	-22	102	morir.v	-13	3	60	efectivamente.adv	-3	16	65
encantar.v	-31	123	campo.n	-0	0	60	iglesia.n	-5	11	71
final.a	-40	128	señor.n	-3	18	70	relación.n	-13	3	74
marido.n	-7	123	fijar.v	-3	35	107	nivel.n	-34	8	125
jo.interjec	-50	193	ay.interjec	-16	3	107	uhum.interjec	-21	11	127

【図-41】内容語 (Con) : 性 (Sexo) / 年齢 (Edad) / 学歴 (Nivel)

内容語の使用頻度の性差・年齢差・学歴差は明確である。とくにたとえば H: *marchar, absoluto*, M: *favorito, jolín* などのような排他的選択が顕著なものが多く見つかる。排他的でなくても *encantar* (H:31, M:123), *jo* (H:50, M:193)

のように使用頻度の差が大きい例が多数ある。これらの語は名詞ではないので、文体特徴と考えられる。

4.3. 単一語

Simple. Año

	A1	A2	A3
yo.pro	976	331	403
estuche.n	429	269	67
mamá.n	340	58	23
abuelo.n	318	91	47
ayer.n	273	19	16
papá.n	233	51	19
lobo.n	199	82	14
pato.n	164	59	22
otoño.n	144	31	22
queso.n	79	25	9
circo.n	75	23	8
pala.n	51	12	5
molino.n	47	8	1
ja.interjec	24	0	1
etapa.n	0	0	14
caso.n	0	0	18
mas.conj	0	0	19
crear.v	0	3	22
bajo.a	2	2	33
costumbre.n	0	2	35
equipo.n	2	8	39
barba.n	2	7	39
paz.n	6	6	42
brazo.n	0	0	37
soler.v	0	12	47
tras.prep	4	10	58
aire.n	0	1	53
uña.n	0	0	55
según.prep	0	10	69
sino.conj	4	23	80
algo.ind	2	27	83
entre.prep	12	26	88
vida.n	6	36	107
deporte.n	12	24	111
contra.prep	0	15	110
cada.ind	12	44	130
sin.prep	16	73	214
etcétera.n	2	274	291
o.conj	26	171	289
que.conj	519	767	2430

Simple. Publicación

	Tea	Nov	Ens	Pre	Tec
no.sn	3354	1140	1164	891	916
me.pro	1962	563	243	67	107
usted.pro	1354	262	38	36	2
yo.pro	1338	324	165	47	52
te.pro	1100	148	24	20	13
mi.pos	751	323	211	47	95
tú.pro	540	77	13	2	1
sí.sn	533	79	25	40	53
querer.v	581	232	119	115	70
venir.v	359	116	62	67	41
tu.pos	301	61	8	11	1
aquí.adv	353	91	94	55	58
hijo.n	274	104	44	35	28
díos.n	230	38	29	27	10
oír.v	189	91	34	31	6
quién.interrog	161	32	16	10	3
mirar.v	177	121	28	15	14
ahí.adv	118	8	18	8	0
callar.v	107	31	8	9	5
cuando.conj	106	92	1	0	1
ah.interjec	84	9	0	1	0
ja.interjec	55	10	0	0	0
matar.v	62	20	4	7	1
chico.a	57	22	2	0	0
jurar.v	46	9	0	0	0
oh.interjec	46	3	8	1	0
mamá.n	40	2	0	1	2
ay.interjec	1	102	4	0	1
toro.n	7	3	3	59	2
ministro.n	6	8	13	86	3
eje.n	1	3	7	1	40
gobierno.n	11	8	47	132	2
curva.n	0	8	5	0	47
norte.n	2	7	2	11	53
lente.n	1	2	0	1	50
serie.n	1	1	9	16	65
método.n	1	0	8	14	83
árabe.a	0	3	1	12	89
grupo.n	2	19	27	24	161
vaso.n	1	4	2	0	170

【図-42】単一語 (Base) : 学年 (Año) / 出版物 (Obra)

低学年は身近な人や生活基本語を多く用いる (*váter*, *ayer*, *tito*, *molino*, *mamá*)。 *según*, *contra* はほとんど高学年に集中して用いられる。 *que* の使用が学年が上がるほど増加する。学年に応じて語彙力が上昇していることがわか

る。話者と相手を示す語(*me, usted, yo, te, mi, tú*)が演劇・小説に多く用いられる。*toro, ministro, gobierno* が新聞に多く, *curva, norte, ..., grupo, vaso* がとくに科学技術文に多い。これらの集中分布を示す語の中で名詞以外は文体特徴になる。名詞は話題語であると思われる。

Simple. Sexo			Simple. Edad			Simple. Nivel				
	H	M		E-1	E-2	E-3		N-1	N-2	N-3
eh.interjec	1424	599	leer.v	173	20	12	cierto.ind	108	8	23
hoy.n	89	26	libro.n	141	8	20	libro.n	98	21	28
caso.n	84	21	cierto.ind	112	28	15	mil.num	74	8	11
cierto.ind	71	21	dinero.n	112	18	32	crío.n	55	11	0
cerca.adv	50	8	tren.n	51	3	17	marchar.v	47	5	3
tienda.n	52	13	película.n	48	15	0	costumbre.n	45	3	11
tranquilo.a	46	13	loco.a	22	0	2	joder.v	34	3	3
quizás.adv	43	11	deuda.n	19	0	0	jugar.v	66	204	45
marchar.v	36	0	freír.v	19	0	2	suegro.n	42	77	3
regalo.n	22	0	moto.n	19	0	2	norte.n	24	0	6
escuela.n	19	0	ropa.n	19	0	2	jo.interjec	45	296	6
plaza.n	17	0	sábana.n	16	0	0	copa.n	16	0	0
salsa.n	17	0	alcalde.n	16	0	2	cueva.n	13	0	0
etcétera.n	15	0	lista.n	13	0	0	renta.n	13	0	0
base.n	14	0	mojar.v	13	0	0	pff.interjec	32	116	17
madera.n	14	0	negro.a	13	0	0	tren.n	16	48	0
vino.n	12	0	pimienta.n	13	0	0	chico.a	21	93	17
obra.n	10	0	uhum.interjec	29	101	20	idea.n	5	29	3
pie.n	10	0	hola.interjec	13	48	5	fruta.n	0	19	0
planta.n	10	0	primo.a	6	30	0	ti.interjec	0	19	0
recto.a	10	0	carne.n	0	25	0	planta.n	0	16	0
salvo.a	10	0	ti.interjec	0	18	0	sala.n	0	16	0
caja.n	0	11	bodega.n	0	13	0	feliz.a	3	29	3
dulce.a	0	11	piel.n	0	13	0	bodega.n	0	13	0
prado.n	0	11	vaca.n	0	20	2	don.n	0	13	0
miércoles.n	0	13	norte.n	0	23	5	lleno.a	0	13	0
rey.n	0	13	jo.interjec	22	277	35	arroz.n	0	53	6
rollo.n	0	13	cueva.n	0	0	12	sábado.n	8	50	11
ayer.n	2	17	don.n	0	0	12	huelga.n	0	0	14
durar.v	2	17	huelga.n	0	0	12	premio.n	0	0	17
huevo.n	2	25	pleno.a	0	0	12	espacio.n	5	0	23
cocer.v	2	26	otoño.n	0	0	20	etcétera.n	3	0	23
sábado.n	9	40	espacio.n	0	3	22	nota.n	5	3	31
serie.n	7	42	crío.n	10	8	47	preguntar.v	0	13	37
tío.n	15	74	nacer.v	6	5	45	maestro.a	0	0	48
ay.interjec	14	77	maestro.a	0	0	42	época.n	11	29	82
libro.n	10	93	morir.v	13	3	60	sobre.prep	32	34	125
leer.v	22	102	campo.n	0	0	60	nivel.n	34	8	125
hm.interjec	52	134	aquel.dem	3	23	102	uhum.interjec	21	11	127
jo.interjec	50	193	ay.interjec	16	3	107	pueblo.n	45	48	161

【図-43】単一語 (Base) : 性 (Sexo) / 年齢 (Edad) / 学歴 (Nivel)

marchar, etcétera, joder, eh は男性に多く, *ay, uy, jo, rollo, cole* は女性に多い。*ehh* は若年, *jo* は中年, *sss, santo, aquel, ay* は老年に多い。*pff, jo* は中学歴, *uhum* は高学歴に多く使われている。これらのうち, 名詞(*rollo, cole, santo*)は

話題語であり，それ以外は文体特徴になる。

4.4. 複合語

Complejo. Año

	A1	A2	A3
muñeco.n	328	100	43
televisión.n	253	95	70
maceta.n	235	84	32
ratón.n	207	72	21
cuento.n	192	42	7
dibujo.n	211	79	32
juguete.n	126	44	26
paseo.n	79	19	14
regalar.v	59	3	14
señorita.n	61	37	0
perseguir.v	34	7	3
cambio.n	0	0	12
educación.n	0	1	15
libertad.n	0	2	17
mayoría.n	0	2	17
realizar.v	0	2	18
recordar.v	0	3	19
contrario.a	0	0	18
observar.v	0	2	19
practicar.v	0	1	22
clásico.a	0	0	23
asignatura.n	0	2	24
general.a	0	3	25
alrededor.adv	0	7	27
destruir.v	2	5	28
jugador.n	0	8	35
bastante.ind	0	12	37
amistad.n	0	8	40
demás.ind	0	20	47
celebrar.v	2	15	48
contaminación.n	0	20	53
alegrar.v	10	11	55
partir.v	10	14	60
vosotros.pro	2	18	62
importante.a	0	16	64
contaminar.v	0	15	83
aunque.conj	6	22	101
así.adv	6	68	115
desde.prep	18	35	138
nosotros.pro	18	96	191

Complejo. Publicación

	Tea	Nov	Ens	Pre	Tec
señor.n	538	149	38	190	54
señorita.n	111	15	3	10	0
gustar.v	115	28	22	5	2
marido.n	84	20	5	7	1
casar.v	88	32	7	8	3
asimismo.adv	49	3	10	1	3
adiós.interjec	29	6	0	0	0
átomo.n	1	27	1	0	1
literario.a	5	4	77	21	9
literatura.n	0	3	60	10	12
información.n	0	1	2	27	2
sindicato.n	1	0	4	33	0
socialista.a	0	2	14	46	0
civil.a	1	9	5	50	3
manejo.n	2	0	1	0	24
integral.a	0	0	2	2	27
apertura.n	0	0	0	4	30
longitud.n	0	0	2	1	32
mencionar.v	1	2	2	3	40
superficie.n	1	4	3	2	41
célula.n	0	0	2	2	39
matemático.a	1	0	3	4	41
nervioso.a	7	6	2	5	53
matemática.n	1	1	4	0	48
instrumento.n	1	2	11	5	53
geográfico.a	0	0	2	4	49
describir.v	1	1	12	10	57
reino.n	3	1	5	7	59
investigación.n	0	2	10	9	63
capítulo.n	2	0	11	4	68
objetivo.a	2	0	9	4	79
microscopio.n	0	2	0	1	77
región.n	0	4	24	17	96
geografía.n	1	0	3	1	89
descripción.n	0	1	7	5	107
autor.n	3	8	59	36	142
ciencia.n	6	14	52	19	143
península.n	1	1	10	11	128
décimo.a	1	4	60	52	167
cultura.n	0	3	50	11	162

【図-44】複合語 (Format.) : 学年 (Año) / 出版物 (Obra)

低学年(A1)の使用語彙は学年が進むにつれて頻度が低くなり(例:*muñeco*, *televisión*, *maceta*, *ratón*)，逆に高学年(A3)の使用語彙は上級生になるほど頻度が上昇する(例:*contaminar*, *aunque*, *así*, *desde*, *nosotros*)。演劇(Tea)では日常的な *señor*, *señorita*, *gustar*, *marido*, *casar*, *asimismo*, *adiós* を多く使い，小

説(Nov)・随筆(Ens)・新聞雑誌(Pre)・科学技術文(Tec)では特定の話題語(話題に関連する語)が使われる：Tea: *átomo*, Ens: *literario*, *literatura*, Pre: *información*, *sindicato*, *socialista*, *civil*, Tec: *manejo*, *integral*, *apertura*, *longitud*, ..., *península*, *décimo*, *cultura*. 全体的に名詞・形容詞が多いことは、その多くが話題語であることを示している。少数の動詞も話題語と思われる語がある(*llenar*, *destruir*, *observar*, *practicar*, *mencionar*)。話題語であれば文体特徴とはならない。この資料は書き言葉なので書き手が意識的に吟味・選択した語彙であると考えられる。そのような意識的な文章では言語規範が強く作用するため自由な文体の特徴が現れにくい。

Complejo. Sexo			Complejo. Edad			Complejo. Nivel				
	H	M		E-1	E-2	E-3		N-1	N-2	N-3
normalmente.adv	107	25	bastante.ind	247	53	47	final.a	161	50	31
exactamente.adv	67	13	realmente.adv	90	10	22	empresa.n	40	8	3
intentar.v	55	8	intentar.v	77	23	7	genial.a	34	3	0
acostumbrar.v	38	8	dedicar.v	42	5	7	paso.n	29	0	3
inglés.a	29	4	edificio.n	35	3	2	claramente.adv	29	3	3
temperatura.n	29	4	avión.n	32	0	0	negocio.n	26	0	0
diferencia.n	26	2	básicamente.adv	32	0	0	construcción.n	24	3	0
situación.n	26	4	durante.prep	32	3	2	patrón.n	18	0	0
absoluto.a	22	0	apunte.n	29	0	0	suceder.v	16	0	0
claramente.adv	22	0	matemático.a	29	3	0	pensamiento.n	16	0	0
existir.v	21	2	construcción.n	26	3	2	coste.n	16	0	0
edificio.n	21	2	cantar.v	19	0	0	casar.v	18	61	6
perdón.n	21	2	italiano.a	19	0	0	alegría.n	5	26	0
circunstancia.n	19	2	coste.n	19	0	0	bicicleta.n	3	26	0
marcar.v	15	0	actor.a	16	0	0	inglés.a	3	45	3
humedad.n	15	0	tradición.n	16	0	0	bautizar.v	0	16	0
apunte.n	15	0	psicología.n	16	0	0	actor.a	0	0	14
producto.n	14	0	ocupar.v	13	0	0	comportamiento.n	0	0	14
ocasión.n	14	0	control.n	13	0	0	procesión.n	0	0	17
posiblemente.adv	14	0	danza.n	13	0	0	turístico.a	0	0	20
motivo.n	12	0	final.a	38	174	25	decisión.n	0	3	23
aprobar.v	12	0	montar.v	13	38	0	celebrar.v	0	0	23
generación.n	12	0	clásico.a	3	20	0	religioso.a	0	0	23
apartar.v	12	0	bautizar.v	0	15	0	posiblemente.adv	0	0	23
mejora.n	12	0	turismo.n	0	13	0	matemático.a	0	3	25
turístico.a	12	0	sexto.a	0	13	0	primaria.n	0	0	25
administrador.n	12	0	relativamente.adv	0	13	0	apunte.n	0	0	25
político.a	0	13	madrina.n	0	13	0	carrera.n	8	0	31
repartir.v	0	13	genial.a	0	33	2	interesante.a	0	5	31
favorito.a	0	17	procesión.n	0	0	15	humano.a	3	0	31
móvil.a	0	21	kilómetro.n	0	0	17	accidente.n	0	0	31
once.num	3	25	consciente.a	0	0	17	participar.v	0	0	34
genial.a	2	25	administrador.n	0	0	17	procurar.v	0	8	40
precioso.a	2	26	alternar.v	0	3	22	dedicar.v	0	3	48
santo.a	2	30	cultural.a	0	5	37	santo.a	0	0	48
muchísimo.a	21	76	santo.a	0	3	40	efectivamente.adv	3	16	65
encantar.v	31	123	ambiente.a	6	10	50	iglesia.n	5	11	71
final.a	40	128	efectivamente.adv	0	18	57	relación.n	13	3	74
marido.n	7	123	señor.n	3	18	70	además.adv	11	53	122
verdad.n	101	259	fijar.v	3	35	107	normalmente.adv	53	19	136

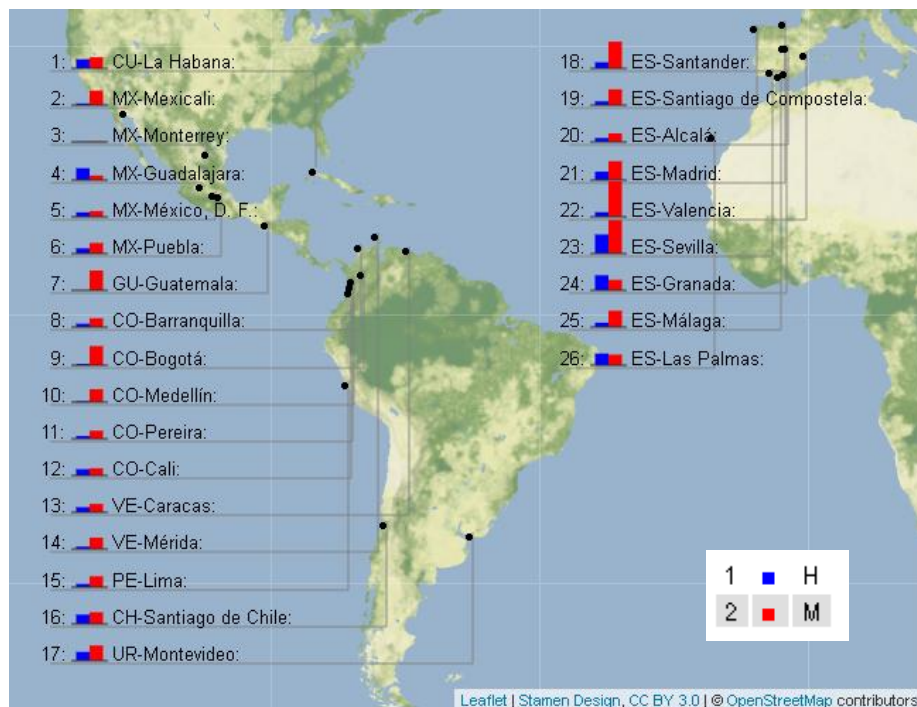
【図-45】複合語 (Complejo) : 性 (Sexo) / 年齢 (Edad) / 学歴 (Nivel)

男性(H)は *normalmente, exactamente, claramente, posiblemente* などの *-mente* の副詞を多く使用し、女性(M)は *favorito, jolín, muchísimo, encantar* を多く使用している。若年者(E1)は *bastante, realmente*、中年者(E2)は *final, genial*、老年者(E3)は *efectivamente* を多く使用する。低学歴者(N1)は *final, genial*、高学歴者(N3)は *efectivamente, normalmente* を多く使用している。これらは文体特徴として認められる。

先の小学生の作文・出版物の資料は規範が働く書き言葉であるため文体特徴が見つからなかったのに対し、この録音転写資料は自由な話し言葉であるので自然な文体特徴がよく現れている。

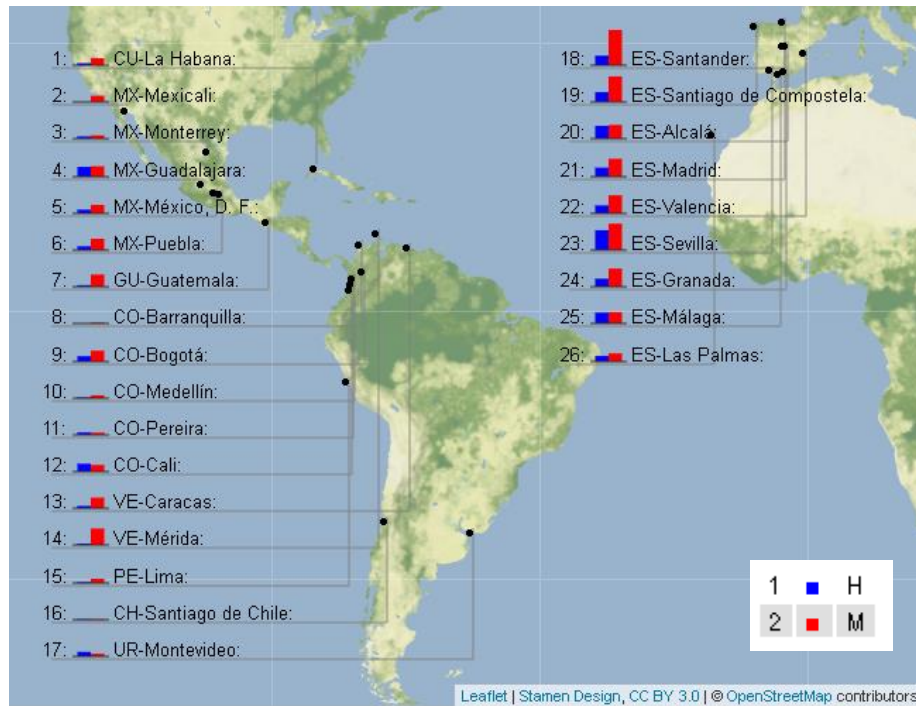
4.5. 地域差

この章では先に見た次の事項を取り出す(4.1, 4.2, 4.3) : *efectivamente* は男性と高学歴者に多く、*jolín, muchísimo, encantar* は女性に多い。以下ではそれぞれの特徴を PRESEEA 全体で広範囲に観察する。



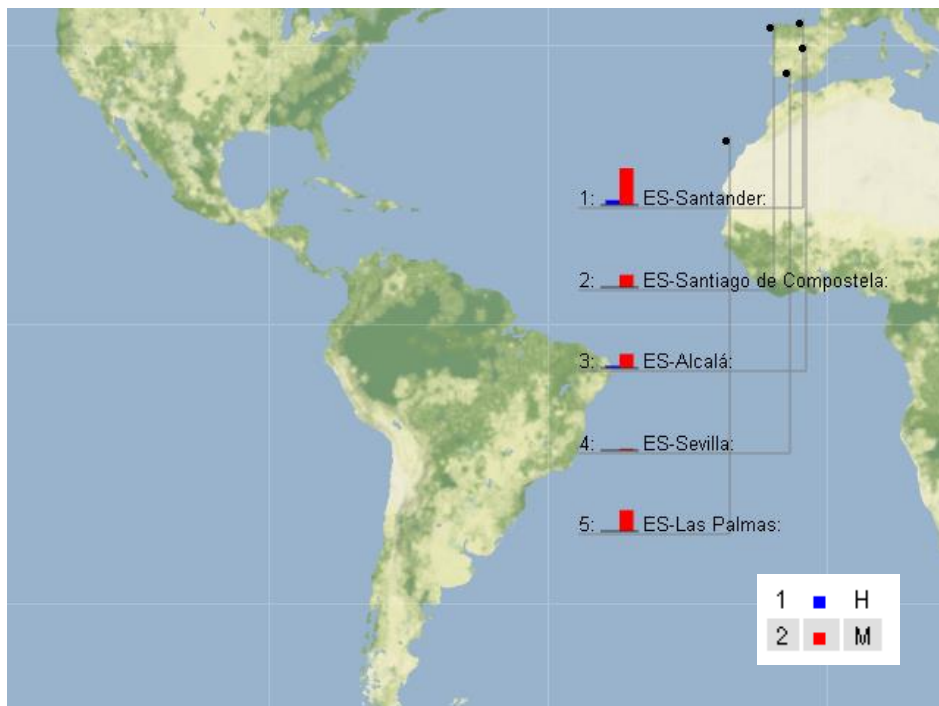
【図-46】 *encant{ar}* (計:1,264) Sexo: H ~ M

encantar (Mujer)は、ほぼスペイン語圏全体の文体特徴である。



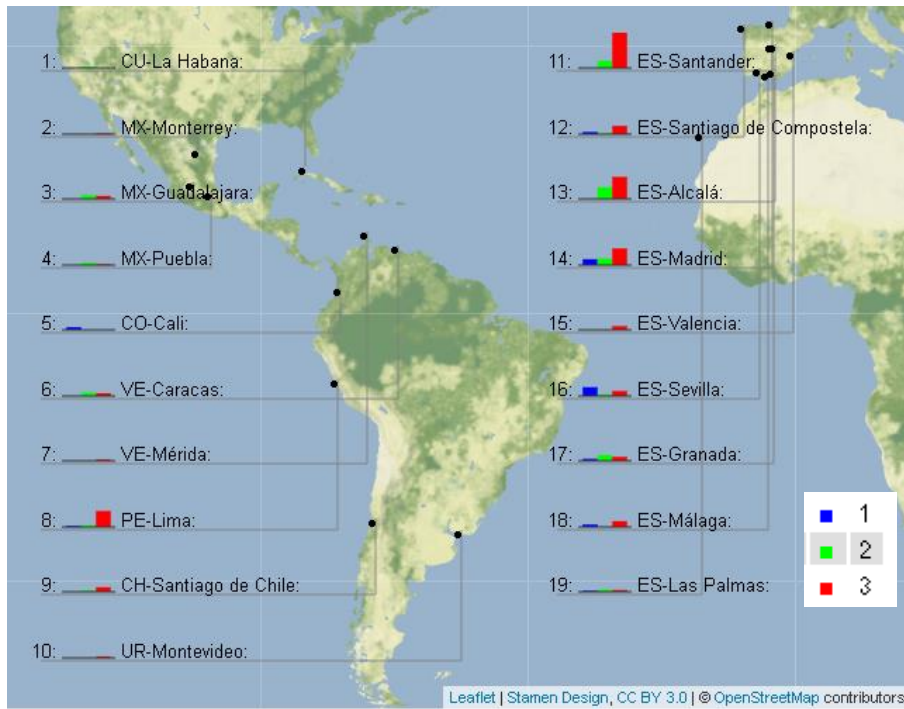
【図-47】 muchísimo[oa](s) (計:1,232) Sexo: H ~ M

muchísimo (M)はスペイン語圏全体の文体特徴である。



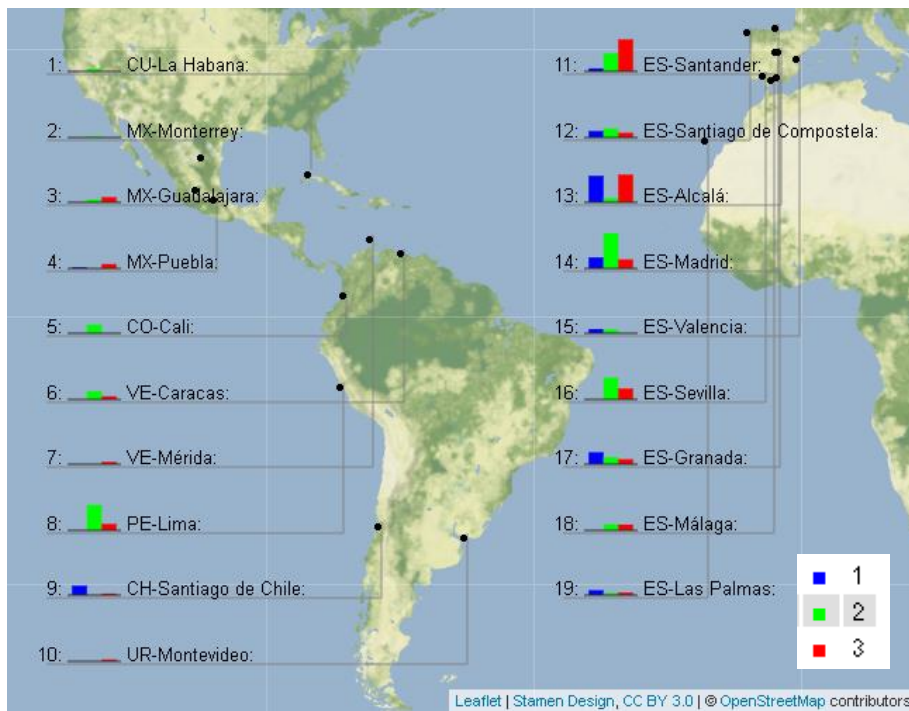
【図-48】 jolín (計:57) Sexo: H ~ M.

間投詞 *jolín* はスペイン・女性(M)の文体特徴である。



【図-49】 *efectivamente* (計:156) Edad: E1 ~ E3

efectivamente はスペインとペルーでは E3 の文体特徴として認められる。



【図-50】 *efectivamente* (計:156) Nivel: N1 ~ N3.

efectivamente (N3) は Santander に限られる。

4.6. 年代差

先に語形成の章で見たように(→3.4), García Hoz (1953)と Juilland and Chang-Rodríguez (1964) [GH.JCh] (20 世紀前半のスペインの書き言葉), 5. Davies (2006) (20 世紀後半のスペイン・ラテンアメリカの話し言葉・書き言葉), Real Academia Española (2023) [Colpes] (21 世紀のスペイン・ラテンアメリカの話し言葉・書き言葉)を使って, 機能語(Fun)と内容語 (Con)の年代的変遷を観察する(数値は 10 万語あたりの正規化頻度)。

Fun. Diacronía

	S.20.a	S.20.b	S.21
haber.v	1302	1137	251
lo.pro	1431	1412	674
mi.pos	387	286	221
usted.pro	239	113	36
yo.pro	286	315	139
te.pro	272	172	161
tanto.ind	241	228	102
nuestro.pos	208	119	109
tu.pos	177	68	66
cuándo.interrog	131	7	6
pues.conj	165	117	54
aquel.dem	158	119	83
sí.sn	154	195	93
tú.pro	103	54	31
dónde.interrog	78	24	19
porque.conj	174	257	152
cuál.interrog	63	23	16
os.pro	55	12	8
vuestro.pos	44	4	2
cuyo.rel	61	39	31
vosotros.pro	24	3	2
excepto.prep	1	3	0
mas.conj	4	12	4
vos.pro	1	18	5
sí.pro	0	0	3
y/o.conj	0	0	3
alguien.ind	6	23	24
bajo.prep	21	0	30
mí.pro	31	73	48
cualquiera.ind	27	55	47
hacia.prep	37	78	59
tras.prep	9	29	46
cual.rel	14	57	61
mientras.conj	29	60	70
donde.conj	21	109	110
qué.interrog	109	242	161
uno.ind	158	124	212
cuando.conj	58	229	200
o.conj	302	476	382
ese.dem	308	549	397

Con. Diacronía

	S.20.a	S.20.b	S.21
señor.n	207	82	30
don.n	187	58	15
dios.n	78	35	7
aun.adv	64	13	9
peseta.n	54	10	2
sindical.a	33	0	2
décimo.a	33	1	2
temer.v	36	8	7
obrero.a	29	4	4
abrazo.n	26	2	3
señorita.n	24	5	3
sindicato.n	21	3	4
marqués.n	17	0	1
península.n	17	0	2
nene.n	12	0	2
excelentísimo.a	11	0	0
conde.n	11	0	1
cortejo.n	8	0	0
música.n	12	25	0
eh.interjec	1	91	4
ambiental.a	0	0	9
perspectivo.a	0	0	9
torneo.n	0	0	9
vídeo.n	0	0	9
digital.a	0	0	10
gol.n	1	0	11
evento.n	0	2	12
pasarela.n	0	0	12
estadounidense.a	0	0	13
contenido.n	0	0	14
usuario.n	0	0	14
informe.a	0	0	15
internet.ex	0	0	15
listo.a	1	4	18
euro.n	0	0	17
tecnología.n	0	6	22
mundial.a	1	9	26
generar.v	0	10	27
músico.a	3	4	31
equipo.n	4	21	48

【図-51】年代差：機能語(Fun) / 内容語(Con)

機能語(Fun)と内容語(Con)のそれぞれの年代の変遷を比較すると、機能語ではとくに大きな傾向が見られないが、内容語では *señor, don, dios, aun, peseta* の使用頻度が大きく減少し、*euro, teconología, actor, mundial, generar, músico, equipo, partido* の使用頻度が大きく上昇していることがわかる。このように、機能語は文法的な役割を果たすので、社会や文化の変化を反映することはないが、内容語の使用や生成は社会や文化の変化に対応していく。

次に、内容語の中から、名詞(Sus)、形容詞(Adj)、動詞(Vb)を抽出して比較しよう(G: García Hoz (1953) y Juilland and Chang-Rodríguez (1964); D: Davies (2006); C: Real Academia Española (2023))。

Diacronía. Nombre.				Diacronía. Adjetivo.				Diacronía. Verbo.			
	D.1	D.2	D.3		D.1	D.2	D.3		D.1	D.2	D.3
señor.n	207	82	30	español.a	112	41	43	haber.v	1302	1137	251
don.n	187	58	15	santo.a	58	47	6	escribir.v	103	47	41
dios.n	78	35	7	sindical.a	33	0	2	desear.v	45	16	15
peseta.n	54	10	2	décimo.a	33	1	2	mandar.v	40	16	11
carta.n	55	16	15	obrero.a	29	4	4	hallar.v	38	12	9
abrazo.n	26	2	3	católico.a	22	6	6	temer.v	36	8	7
señorita.n	24	5	3	virgen.a	21	1	9	celebrar.v	39	11	16
sindicato.n	21	3	4	caballero.a	20	6	3	marchar.v	26	9	7
marqués.n	17	0	1	magnífico.a	16	4	2	amar.v	26	9	9
península.n	17	0	2	excelentísimo.a	11	0	0	figurar.v	23	5	5
obispo.n	16	3	2	culto.a	13	2	3	comunicar.v	21	6	8
política.n	20	27	0	análogo.a	8	0	1	alegrar.v	17	3	3
nene.n	12	0	2	céntrico.a	7	0	1	perdonar.v	18	6	5
conde.n	11	0	1	sevillano.a	7	0	1	conceder.v	16	5	5
camarada.n	9	0	1	muchísimo.a	2	9	0	crujir.v	6	0	1
marina.n	9	0	1	celular.a	0	0	6	coordinar.v	0	0	4
microscopio.n	9	0	1	puto.a	0	0	6	implementar.v	0	0	4
cortejo.n	8	0	0	variable.a	0	0	6	liderar.v	0	0	4
substancia.n	8	0	0	docente.a	1	0	7	protagonizar.v	0	0	4
música.n	12	25	0	carretero.a	0	0	7	reportar.v	0	0	5
virus.n	0	0	7	dispositivo.a	0	0	7	fortalecer.v	0	2	6
deudo.n	0	0	8	clínico.a	0	3	9	conectar.v	0	2	7
evaluación.n	0	0	8	federal.a	0	0	8	detectar.v	0	4	8
festival.n	0	0	8	venezolano.a	0	0	8	incrementar.v	0	4	8
promedio.n	0	0	8	móvil.a	1	1	9	vincular.v	0	4	8
liga.n	1	0	9	adulto.a	0	3	10	diseñar.v	0	3	8
torneo.n	0	0	9	tecnológico.a	0	3	10	involucrar.v	0	0	7
video.n	0	0	9	ambiental.a	0	0	9	evaluar.v	0	2	8
municipio.n	2	0	11	perspectivo.a	0	0	9	controlar.v	0	11	12
gol.n	1	0	11	cubano.a	0	5	11	recuperar.v	1	9	14
impacto.n	0	3	12	laboral.a	0	3	11	analizar.v	2	8	15
evento.n	0	2	12	digital.a	0	0	10	ubicar.v	0	5	15
pasarela.n	0	0	12	colombiano.a	0	2	11	enfrentar.v	1	10	17
estrategia.n	0	5	16	educativo.a	0	5	13	identificar.v	0	7	16
contenido.n	0	0	14	estadounidense.a	0	0	13	compartir.v	2	11	19
usuario.n	0	0	14	mexicano.a	1	9	17	agregar.v	5	9	20
euro.n	0	0	17	listo.a	1	4	18	participar.v	5	14	23
tecnología.n	0	6	22	paciente.a	1	12	22	generar.v	0	10	27
nivel.n	4	30	45	mundial.a	1	9	26	utilizar.v	9	36	38
equipo.n	4	21	48	músico.a	3	4	31	incluir.v	7	18	36

【図-52】年代差：名詞(Sus) / 形容詞(Adj) / 動詞(Vb)

上の3つのグラフを比較すると、年代差が顕著なのは名詞と形容詞であり、動詞では *haber* を除いてとくに大きな差が見えない。これは特定の名詞と形容詞の使用は話題の選択に左右され、動詞は話題に応じて特定の語が使われることはあまりなく、どのような話題であっても比較的共通の動詞が使われるためであると思われる。

資料 G (García Hoz 1953 / Juilland and Chang-Rodríguez 1964) はすべてスペイン(100%)、資料 D (Davies 2006) は 57%がラテンアメリカ、資料 C (Real Academia Española 2023) は 70%がラテンアメリカのデータを収集している。そこで、*haber* の使用頻度が 21 世紀の C (Real Academia Española) で非常に下降しているが、これは資料 C がラテンアメリカでの使用頻度を多く載せているためであると思われる。ラテンアメリカではスペインに比べて現在完了形の使用が限られていることが知られている。*peseta* も年代順に使用頻度が大きく減少しているが、これには年代差と地域差の両方が反映していると考えられる。形容詞・名詞 *español* の使用頻度についても同様に年代差よりも地域差が大きく反映しているのであろうか。

そこで、*haber* (活用形), *peseta* (単複), *español* (単複) の現在の使用頻度をスペインとラテンアメリカで比較してみよう(10万語あたりの相対頻度の平均)¹⁸。

(19) *haber, español, peseta* の使用頻度

FR.10 ⁵	# <i>haber</i> #	# <i>pesetas</i> ?#	# <i>español(es)</i> ?#
España	1312.2	18.6	7.6
América	999.7	0.0	8.4

この結果を見ると、たしかに、*haber* と *peseta* は両地域の頻度差が見られるが、*español* の地域差は大きくない(むしろラテンアメリカのほうが僅かに多い)。しかし、*español* の減少傾向を年代差に帰することは困難なので、むしろ資料 G (García Hoz 1953, Juilland and Chang-Rodríguez 1964) に多用された要因を探ると、とくに新聞・公文書(García Hoz)、随筆・新聞(Juilland and Chang-Rodríguez)に集中していることがわかった。よって、20世紀前半における *español* の多用は両資料が収集した当時の時局を反映するものであって、とくに年代差を示すものではないようだ。

このように、3.4 と 4.5 で扱った「年代差」はむしろ「地域差」や資料の特質を示している可能性がある。よって、*haber, peseta, español* については、大まかに見てもスペイン語圏全体の年代差を示していることにはならない。あくまでも、それぞれの資料の性質を念頭に置いて比較しなければならない。

¹⁸ <https://h-ueda.sakura.ne.jp/lyneal/preseea.htm>

一方, *don/doña, dios, carta(s)*の減少傾向についてはどうであろうか? 次の表はそれぞれの使用頻度(10万語あたりの相対頻度の平均)を示している。

(20) *don/doña, dios, carta(s)*の使用頻度

FR.10 ⁵	#(don doña)#	#dios#	#cartas?#
España	14.2	55.6	6.1
América	22.3	60.2	6.2

このようにスペインとラテンアメリカの間に大きな頻度差がないので、3つの資料が示す大きな減少傾向は地域差ではなく年代差を示している、と考えられる。しかし、上の頻度分布は現在(21世紀)の共時態のものであって、20世紀も同様であった、という保証はない。

4.7. 文体差

限られた資料の中で語形成の章(3.6)で行ったように、語彙についても文体差の有無を観察したい。先に説明したように書き言葉の資料として 20 世紀前半の手書きの書簡や各種出版物を収集した García Hoz (1953)と Juilland and Chang-Rodríguez (1964)を利用し、話し言葉の資料として 1990 年代の録音資料を集計した Moreno (2005: C-ORAL-ROM)と Martínez y Ueda (2021a, 2013)を利用する。はじめに機能語と内容語の変異を観察する。

Estilo. Fun.			Estilo. Con.		
	L. hablada	L. escrita		L. hablada	L. escrita
no. sn	2987	1236	ir. v	1146	380
ser. v	3378	2019	bueno. a	683	129
sí. sn	1142	154	saber. v	537	170
yo. pro	1027	286	eh. interjec	401	1
estar. v	1147	534	entonces. adv	336	49
pues. conj	950	165	claro. a	301	38
me. pro	1121	539	luego. adv	268	49
pero. conj	971	325	ahí. adv	217	28
ese. dem	923	308	gustar. v	168	34
o. conj	854	302	gente. n	157	32
te. pro	690	272	ah. interjec	142	15
porque. conj	637	174	mm. interjec	98	0
si. conj	570	231	aproximar. v	90	5
qué. interrog	501	109	valer. v	94	17
cuando. conj	310	58	coger. v	80	11
nada. ind	308	102	joder. v	52	0
tú. pro	227	103	jo. interjec	42	0
mí. pro	162	31	nevar. v	42	0
algo. ind	133	55	poquito. a	34	2
tampoco. adv	81	19	muchísimo. a	32	2
cual. rel	47	14	normalmente. adv	31	0
alguien. ind	25	6	hm. interjec	24	0
bajo. prep	5	21	señorita. n	0	24
vuestro. pos	5	44	décimo. a	1	33
según. prep	18	55	sindical. a	1	33
quien. rel	15	56	alma. n	1	34
vario. ind	19	59	espíritu. n	1	37
ante. prep	9	59	rey. n	3	38
cuyo. rel	6	61	hallar. v	1	38
contra. prep	19	68	ministro. n	2	41
cuándo. interrog	18	131	desear. v	5	45
sobre. prep	68	165	carta. n	6	55
tu. pos	83	177	santo. a	5	58
aquel. dem	41	158	nacional. a	11	65
sin. prep	83	203	aun. adv	1	64
nuestro. pos	89	208	recibir. v	11	86
tanto. ind	111	241	obra. n	13	96
usted. pro	64	239	escribir. v	17	103
su. pos	284	1519	don. n	14	187
de. prep	4388	8599	señor. n	59	207

【図-53】 文体差：機能語(Fun) / 内容語(Con)

話し言葉で多く使われる機能語は書き言葉でも使われるが、逆に書き言

葉で多く使われる機能語は話し言葉では *de* を除いてあまり使用されていない。とくに *bajo, segu/n, quien.rel, ante, cuyo, contra* は書き言葉の文体特徴であると考えられる。

内容語については、*eh, mm, valer, coger, joder, jo*などは話し言葉の特徴であり、*señorita, décimo, sindical, hallar, ..., don, señor* は書き言葉に多い。書き言葉の資料は話し言葉の資料よりも古いので、書き言葉に多い語は古い用法を示している可能性がある(→4.8)。

次の図は話し言葉と書き言葉の違いを品詞(名詞・形容詞・動詞)に分けて集計した結果を示している。

Estilo. Nombre.			Estilo. Adjetivo.			Estilo. Verbo.		
	L.h.	L.e.		L.h.	L.e.		L.h.	L.e.
gente.n	157	32	bueno.a	683	129	ir.v	1146	380
sitio.n	56	12	claro.a	301	38	estar.v	1147	534
nivel.n	38	4	final.a	68	17	tener.v	1117	498
chaval.n	19	0	poquito.a	34	2	decir.v	989	458
virus.n	18	0	normal.a	35	7	saber.v	537	170
bar.n	15	0	muchísimo.a	32	2	gustar.v	168	34
euro.n	15	0	ordenador.a	18	0	mirar.v	142	50
emoción.n	0	14	ejemplar.a	0	10	trabajar.v	101	25
virtud.n	0	14	sabio.a	0	10	aproximar.v	90	5
decreto.n	0	15	excelentísimo.a	0	11	valer.v	94	17
patria.n	0	16	comunista.a	1	13	empezar.v	97	31
marqués.n	0	17	árabe.a	1	14	comer.v	86	21
sentimiento.n	0	19	bello.a	0	15	coger.v	80	11
vaso.n	1	21	lector.a	0	15	meter.v	62	14
poeta.n	0	21	numeroso.a	2	17	joder.v	52	0
figura.n	2	23	alemán.a	2	18	nevar.v	42	0
nación.n	2	23	extraordinario.a	2	19	sobrar.v	26	4
señorita.n	0	24	enemigo.a	1	19	pillar.v	14	0
abrazo.n	1	26	caballero.a	1	20	liar.vb	13	0
ministerio.n	2	28	virgen.a	1	21	pronunciar.v	0	14
acto.n	3	30	católico.a	2	22	poseer.v	0	15
acción.n	5	32	preciso.a	2	23	conceder.v	1	16
corazón.n	5	32	superior.a	2	23	satisfacer.v	1	16
doctor.n	4	34	real.a	3	25	alcanzar.v	1	18
arte.n	6	36	verdadero.a	5	26	declarar.v	3	23
artículo.n	6	36	oficial.a	5	28	emplear.v	3	23
alma.n	1	34	obrero.a	3	29	citar.v	2	23
autor.n	3	35	francés.a	7	31	figurar.v	1	23
espíritu.n	1	37	maestro.a	6	31	constituir.v	3	25
rey.n	3	38	militar.a	4	31	detener.v	3	25
ministro.n	2	41	presente.a	5	32	enviar.v	3	28
guerra.n	5	43	décimo.a	1	33	publicar.v	3	28
ojo.n	9	45	sindical.a	1	33	comprender.v	5	31
modo.n	7	48	viejo.a	7	36	temer.v	4	36
carta.n	6	55	pronto.a	10	41	disponer.v	8	41
orden.n	11	60	santo.a	5	58	hallar.v	1	38
estado.n	17	76	nacional.a	11	65	desear.v	5	45
obra.n	13	96	general.a	29	85	recibir.v	11	86
don.n	14	187	español.a	29	112	escribir.v	17	103
señor.n	59	207	grande.a	62	179	deber.v	49	158

【図-54】 文体差：名詞・形容詞・動詞

それぞれの品詞で話し言葉(L.h.)に多用されている語は日常的な語彙であり、書き言葉(L.e.)で多用されている語は教養語である。名詞・形容詞の日常語は少なく、多くは書き言葉に見られる教養語である。これは話し言葉では少ない日常語が高頻度で使われ、逆に書き言葉では多くの教養語が低頻度で使われていることを示している。動詞でも同じ傾向が見られるが、とくに話し言葉における日常語の頻度は非常に高い(*ir, estar, tener, decir, saber, gustar, mirar, trabajar*)。

4.8. 年代・地域・文体

先に見たように(4.7)、動詞 *hallar* は書き言葉で多く使用されている(【図-52】)。そこで、ここでは *hallar* を類義語 *encontrar* と比較しながら文体差・年代差・地域差を観察し、多面的な分析を通して *hallar* が書き言葉で多く使用される理由を探る。

はじめに 21 世紀のスペイン語の状態を示す CORPES を使って *hallar* の正規化頻度(*por millón*)を調べると次の結果が得られた。

(21) <hallar>の使用頻度. 21 世紀. CORPES

CORPES	絶対頻度	正規化頻度
Oral	41	8.73
Escrito	30,103	76.98

よって、21 世紀のスペイン語では、*hallar* は話し言葉ではなく書き言葉の特徴語である、と考えられる。

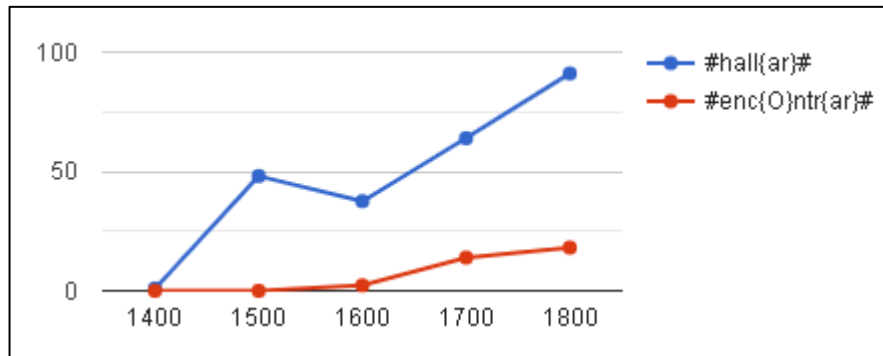
20 世紀のスペイン語での *hallar* の概況を CREA を使って次のように計算した。CREA はレマを検索できないので、*halla**で *hallazgo* を含めて検索した後、*hallazgo* の頻度を減じて絶対頻度とした。正規化頻度を計算できないので、*de* の頻度で割り、その千分率を使った。その結果、21 世紀と同様に 20 世紀のスペイン語でも *hallar* は書き言葉で多く使用されたことがわかった。

(22) *hallar* の使用頻度. 20 世紀. CREA

CREA	<hallar>	<de>	<hallar>/<de>*10 ³
Oral	437	451,724	0.97
Libros	11,775	4,816,867	2.44

次のグラフと表は公証文書における<hallar>と<encontrar>の正規化頻度

(FN por 100.000)の年代推移を示している¹⁹。



【図-55】年代差：<hallar>

Lema	FA					FN: 10 ⁵				
	1400	1500	1600	1700	1800	1400	1500	1600	1700	1800
#hall{ar}#	3	214	149	193	171	0.0	0.0	2.3	13.9	18.1
#enc{O}ntr{ar}#	0	0	9	42	34	1.0	48.1	37.5	63.9	91.1

このように歴史的には *hallar* (<lat. aflare)が先行し、*encontrar* の出現は遅れて 1600 年代であった(当初の意味は「出迎える」('salir al encuentro')であり、1600 年代に *hallar* と同義になった(Corominas 1976, s.v. *contra*))。よって、19 世紀までは<hallar>が大勢を占めていたが、20 世紀に逆転して *encontrar* が多用されるようになった、と言える。

次の表と地図は現在の話し言葉における地理的分布を示している。

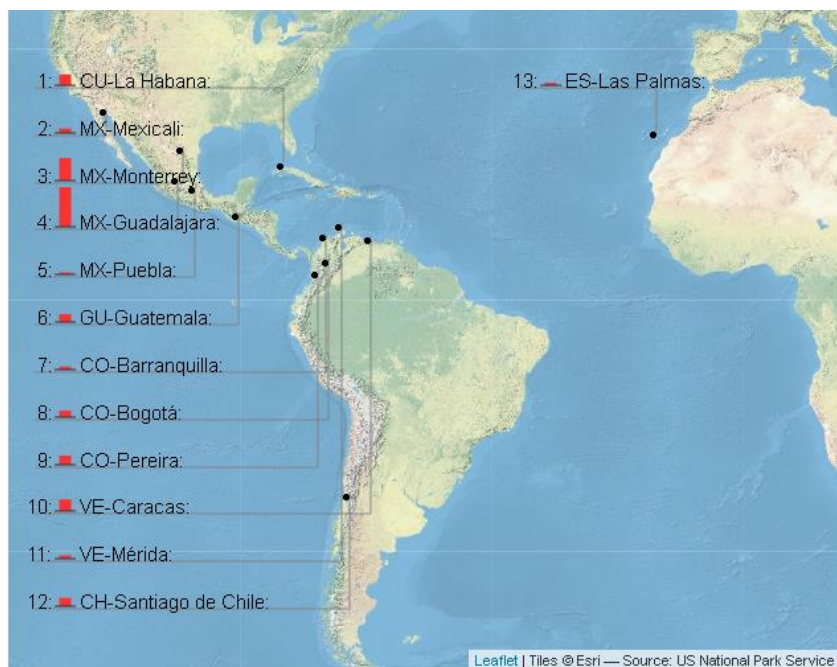
(23) <hallar> y <encontrar> en PRESEEA

(FA: frecuencia absoluta, FN: frecuencia normalizada por 100000)

<hallar>	FA	FN	<encontrar>	FA	FN
MX-Guadalajara	12	8.3	CH-Santiago de Chile	142	82.1
MX-Monterrey	10	4.7	CO-Bogotá	72	48.8
VE-Caracas	4	2.4	CU-La Habana	59	44.4
CU-La Habana	3	2.3	ES-Alcalá	53	41.5
CO-Pereira	2	1.9	ES-Madrid	67	41.2
CH-Santiago de Chile	3	1.7	MX-Puebla	93	37.2
GU-Guatemala	2	1.6	ES-Valencia	39	35.9
CO-Bogotá	2	1.4	CO-Cali	35	34.4
MX-Mexicali	1	0.9	ES-Granada	41	34.1
VE-Mérida	1	0.7	GU-Guatemala	41	33.5

¹⁹ <https://h-ueda.sakura.ne.jp/lyneal/codea.htm> [2023/11/15]

CO-Barranquilla	1	0.6	MX-Mexicali	35	31.8
ES-Las Palmas	1	0.5	MX-México, D. F.	66	31.4
MX-Puebla	1	0.4	PE-Lima	40	31.3
			CO-Barranquilla	49	30.7
			CO-Medellín	38	30.6
			VE-Mérida	43	29.4
			MX-Guadalajara	39	26.9
			ES-Sevilla	37	26.4
			ES-Santander	43	26.2
			ES-Málaga	41	24.6
			UR-Montevideo	29	22.7
			ES-Las Palmas	40	21.1
			CO-Pereira	18	17.0
			ES-Santiago de Compostela	32	14.8
			MX-Monterrey	31	14.5
			VE-Caracas	10	6.1



【図-56】 <hallar> en PRESEEA

現代では *hallar* はフォーマルな文語体として有標化し、一般には *encontrar* が使用される(無標)。Batchelor(2006: 276)によれば<hallar>はスペインよりもメキシコでかなり多く使用される(例: *¿Hallaste lo que buscabas en el súper?*), という。たしかに、上の表と地図で<hallar>の 20-21 世紀の地理的分布を見

ると特にメキシコに多く残存しているが、メキシコでもやはり<encontrar>のほうが多用されている。そして、上の地図を見ると少数ながら話し言葉で使用されている地域のほとんどがラテンアメリカである。これは古形が周辺地域に残存する、という言語地理学的傾向を示している。

このように、言語(語彙)は年代・地域・文体的要因によって多様に変異する。そのとき、古形-新形、ラテンアメリカ-スペイン、文語-口語のベクトルは一般に一致することが多い。このことを確認するためには現存する限られた資料を多面的に観察しなければならない。

現存する限られた資料の扱い方については、さまざまな地理的・歴史的・社会的要因によって変化する社会的現象の比較においても同じ問題が生じる。言語(使用語彙)も地理的・歴史的・社会的要因によって変化するのので、厳密に比較するには、各種の要因を変数として揃えたデータを使用しなければならないであろう。現在私たちが準備中の資料では地理的・歴史的・社会的要因を変数としているが、しかし歴史資料の中には欠損部も生じることは避けられない。社会科学で行われる組織的な層別アンケートや自然科学で行われる変数を揃えた実験のように網羅性を追求することができないためである。私たちができることは、言語資料を可能な限り幅広く多量に収集し、それでも限定されてしまう資料を使って、その限定されていることを意識しながら注意深く観察することである。

5. 結論

本研究ではスペイン語の語彙の形態・機能・意味の問題を語長・語種・変異の観点から分析し考察した。資料は異なる性質をもつ 8 種(1. García Hoz (1953), 2. Juilland and Chang-Rodríguez (1964), 3. 上田 (1987), 4. Justicia (1995), 5. Moreno et al. (2005), 6. Davies (2006), 7. Martínez y Ueda (2021a, 2013), 8. Real Academia Española (2023b))を使ったが、これらによって同質の資料が示す偏りを避けた。統計的方法を使って頻度の和・分散・分布を観察した。

一般に言われる「短縮の法則」(「語長が長くなると使用頻度が小さくなる」「使用頻度が大きくなると語長は短縮する」)が主張する使用頻度と語長の関係は、語彙を機能語(冠詞・代名詞・前置詞・接続詞など)と内容語(名詞・形容詞・副詞・動詞など)に峻別することによって明らかになる。機能語は語彙的意味を表示せず、文法的機能を果たすので短い弱勢語になることが多い。特に非常に高頻度の定冠詞・前置詞・接続詞はすべて短い弱勢語である。一方、内容語の意味表示に一定の音形が必要である。この機能語と内容語の対

照的な違いが短縮の法則の要因である。よって、短縮の法則は機能語と内容語の文法的・語彙的対立の結果であって、その要因ではない。つまり、高い使用頻度が語形を短縮するのではなく、逆に、語形が短い機能語・単一語が高頻度で使用されるのである。

統計学で扱う「疑似相関」(spurious correlation)の例として学童の a.勉強時間と b.成績が挙げられるだろう。両者(a, b)は直接の関係はないが、疑似的な相関が生まれることが多い。しかし、a.勉強時間を増やしても b.成績が上昇するとは限らない(下降することもある)。ここで x.「交絡因子」(confounding factor)として、a.勉強時間と b.成績のどちらに大きく関係する x.「教科への興味/好悪」が考えられる。x.教科への興味があれば(x.教科が好きであれば)、a.勉強時間は増え、b.成績も向上する。

本研究で扱ってきた語の a.使用頻度と b.語長(音素数)の交絡因子は x.語種の違いである(機能語・内容語、単一語・複合語)。機能語・単一語は語長が比較的短く高頻度であり、内容語・複合語は語長が比較的長く低頻度である。このことが a.使用頻度と b.語長(音素数)の疑似相関を生んでいる、と考えられる。

表面的な音形の増加は本質的に加重される形態素(接頭辞・接尾辞)の増加によるものである。そこで形態素が加重されることから有標性を帯び、その結果頻度が減少する。有標的形態素は各種の変数に従って様々な変異を生む。それらの変異は計量文体論の研究で定量的・定質的に分析される。

ここでは資料が示す学年・出版物・性・年齢・学歴という変数に依存する文体特徴を探ってきた。そして、言語地図に現れる「地域」という変数も一部で観察した。Guiraud (1957 [1959]: 58-59)によれば、さらに「時代差」も考慮しなければならない。たとえば Cervantes は *Don Quijote* に当時既に古くなった文体を使わせて独特の個性を与えている。また、先に見たように有標の *jamás* が北スペインの老齡者(E3)に使われていることは、時代的文体特徴を示している可能性がある。よって計量文体論は歴史的变化と地域的変異によって立体化される。本研究では現在使用可能な資料を使って、かなり大きな限界があるにも拘わらず、20-21 世紀間の通時的な変化を辿る試みを行った。

この研究では話題語(会話や文章の話題・テーマに関連する語、とくに名詞)と文体特徴を弁別した。話題語は話題が変われば当然変化するので、文体の特徴(学年・出版物・性・年齢・学歴)の特徴とはならないと考えられるからである。しかし、たとえば性差については、男性間の会話の話題の多くはスポーツ、仕事、自動車、女性であり、女性間の会話の話題の多くは家族、仕事、家、衣服、友人、男性であるという(García Mouton 1991: 71)。両者に共通する話題として仕事の他に共有する知人の噂も考えられる。これらの共通の話題を除外した特有の話題選択の傾向がそれぞれの話し言葉の特徴づ

けているならば、話題語も含めて分析対象とすべきである。話題語は名詞、部分的には形容詞に多いので英語とスペイン語の最大の高頻度の名詞を比較すると、英語: *time, year, people, way, man* であり (Lindquist 2009 [2016]: 39, British National Corpus, 1 億語), 私たちの三種資料(110 万語)では *niño, casa, día, mesa, silla* であって、まったく異なる。この中の *silla* の頻度は Juilland and Chang-Rodríguez (1964: 50 万語), 27, Justicia (1995: 53 語), 2470 (A1:455, A2:1070, A3:945), Santander (Martínez y Ueda 2021, 2023: 11 万 5 千語), 3 である。このように資料によって頻度が大きく異なるのは話題語の特性である。よって、*silla* のような話題性の高い名詞は「言語の文体特徴」とは区別して「話題の文体特徴」とする。

以上見てきたように、文体の問題は確立した言語の「体系」(sistema)の問題ではなく、むしろ無標の「慣用」(norma)から逸脱・変異した有標的「言」(habla)の問題である(Coseriu 1973 [1981])。「言：観察の対象の、無限に変化し、変化しうる具体的諸特徴」(同:47)は選択された語形の様々な使用頻度によって示される(無限に変化する文体的変種)。一方「慣用：対象の特定の機能から独立した、慣習的、共有的で多かれ少なかれ恒常的な諸特徴」は多数の使用頻度によって示される安定した無標の慣用的変種である。最後に「体系：不可欠な、すなわち機能的な諸特徴」は全変種を統合した全体によって示される。従って、言語の変異を対象とする文体論研究は言→慣用→体系という言語研究の中心的過程の最初に位置する基礎であって、体系から逸脱した特殊な現象を扱う付随的・補足的な研究ではないと考える。

プログラム-R

5.1. 相対標準偏差

```
sdp=function(A) {n=length(A); sqrt((n-1)/n)*sd(A)}
#sdp; population standard dev. / sd: unbiased standard dev.
sdpr=function(A) {n=length(A); sd(A)/mean(A)/sqrt(n)}
#relative unbiased standard dev.
> A=c(3, 5, 8, 9, 12); sd(A); sdp(A); sdpr(A) # (1)
# 3.507136; 3.136877; 0.2119512
> A=c(10,0,0,0,0); sd(A); sdp(A); sdpr(A) # (2)
# 4.472136; 4; 1
> A=c(10,10,10,10,10); sd(A); sdp(A); sdpr(A) # (3)
# 0; 0; 0
```

* sd は R 標準関数不偏標準偏差, sdp は母標準偏差, sdpr は相対標準偏差を返す (→上田 2016, 4.6.1)。

5.2. 規定距離

```
DM=function(X=D, d='e', s=F){
  #d='e':ユークリッド距離/'m':平均距離/'r':規定距離, s:標準化
  n=nrow(X); p=ncol(X); R=matrix(rep(0,p*p),p,p)
  colnames(R)=colnames(X); rownames(R)=colnames(X)
  if(s) X=dd(X,stats(X,1,'sdpr')) # 標準化
  for(i in 1:p){for (j in 1:i-1){
    if(d=='e') R[i,j]= sum((X[,i]-X[,j])^2)^(1/2) # d='e':ユークリッド距離
    if(d=='m') R[i,j]=(sum((X[,i]-X[,j])^2)/n)^(1/2)# d='m':平均距離
    if(d=='r') R[i,j]=sum((X[,i]-X[,j])^2)/(sum(X[,i]^2)+sum(X[,j]^2))
    # d='r'; 規定距離
    R[j,i]=R[i,j]
  }}; R
} #距離行列(Distance matrix)
```

* 規定距離は DM(D, d='r')とする(D:行列)。(規定距離→上田 2016: 6.4.6)

5.3. 単調性

```
Monotony=function(A){
  a=d=0; for(i in 2:length(A)){
    if(A[i]>A[i-1]) a=a+A[i]-A[i-1] else d=d+A[i-1]-A[i]
  }; (a-d)/(a+d)
} #Monotony [-1, 1]
```

* A:数値ベクトル (→上田 2016: 4.7.1)

5.4. L字型分布

```
ldi = function(A){n=length(A); mx=max(A); (mx*n - sum(A)) / (mx*(n - 1))}
```

* A:数値ベクトル (→上田 2016: 4.6.12)

5.5. 集中分析

```
Concent=function(D,sel){
  p=ncol(D); n=nrow(D)
  Qp=(1:p)-(p+1)/2; Qp #重み横ベクトル
  Qpp=matrix(Qp, nrow=n, ncol=p, byrow=T)
  Qn=(1:n)-(n+1)/2; Qn #連番縦ベクトル
  Qnn=matrix(Qn, nrow=n, ncol=p, byrow=F)
  for(i in 1:1000) {
    W=D
    if(sel==1|sel==3){
      A=rowSums((D*Qpp)^3); A=CubeRoot(A/rowSums(D)); D=D[order(A),]}
    if(sel==2|sel==3){
      A=colSums((D*Qnn)^3); A=CubeRoot(A/colSums(D)); D=D[,order(A)]}
    if(sel<3 | all(D==W)) break
  }; D
} # Concentration D:df, sel=0: none, 1:row, 2: col, 3:all
```

* D: 行列; 行の並べ替えは sel=1 (→上田 2016: 7.2).

参考文献

- Alliende, Felipe. 1987. "Perfil 4, cuatro procedimientos rápidos para determinar la legibilidad de un texto", *Lectura y Vida. Revista Latinoamericana de Lectura*, año 8, n° 4, sin número de página.
http://www.lecturayvida.fahce.unlp.edu.ar/numeros/a8n4/08_04_Alliende.pdf
- Alvar, Manuel y Bernard Pottier. 1983. *Morfología histórica del español*. Madrid. Gredos.
- Batchelor, R. E. 2006. *Using Spanish Synonyms*. Cambridge University Press.
- Bustos Tovar. 1974. *Contribución al estudio del cultismo léxico medieval*. Anejo del Boletín de la Real Academia Española. Anejo XXVIII, Madrid.
- Bybee, Joan. 2007. *Frequency of Use and the Organization of Language*. Oxford University Press.
- Camus, Bruno. 2022. *La formación de palabras*. Madrid. Atco / Libros.
- Corominas, Joan. 1976. *Breve diccionario etimológico de la lengua castellana*. Madrid. Gredos.
- Coseriu, Eugenio. 1973. *Teoría del lenguaje y lingüística general. Cinco estudios*.

- Madrid. Gredos. 原誠・上田博人(訳)『人間の学としての言語学. コセリウ言語学選集 2 言語体系』東京：三修社.
- Crawford, Alan N. 1985. "Fórmula y gráfico para determinar la comprensibilidad de textos del nivel primario en castellano", *Lectura y Vida. Revista Latinoamericana de Lectura*, año 6, nº 4, sin número de página.
http://www.lecturayvida.fahce.unlp.edu.ar/numeros/a6n4/06_04_Crawford.pdf
- Divjak, Dagmar. 2019. *Frequency in Language. Memory, Attention and Learning*. Cambridge University Press.
- Domínguez, Giuseppe. 2021. *Diccionario de la RAE en modo texto plano*.
<https://www.giuseppe.net/blog/archivo/2015/10/29/diccionario-de-la-rae-en-modo-texto-plano/>
- Ferrer García, C. E Pascual Gaspar, J. A. Láñez Gadea. 2009. "Legibilidad y comprensibilidad de la información individual y consolidada en las empresas cotizadas españolas", *XV Congreso AECA*, 2009.
http://www.aeca1.org/pub/on_line/comunicaciones_xvcongresoaecca/cd/27a.pdf
- Frago, Juan Antonio. 2015. *Don Quijote. Lengua y sociedad*. Madrid. Arco/libros.
- Frías Delgado, Antonio. 2009. "Distribución de frecuencias de la longitud de las palabras en español: aspectos diacrónicos y de estilometría", *A Survey on Corpus-based Research. Panorama de investigaciones basadas en corpus*, Asociación Española de Lingüística del Corpus.
<https://www.um.es/lacell/aelinco/contenido/pdf/51.pdf> [26/9/2023]
- 福井芳男. 1978. 「文体論とリファテール」 Michael Riffaterre. 1971. *Essais de Stylistique Structurale*. Flammarion. 福井芳男・宮原信・川本皓嗣・今井成美(訳)『文体論序説』東京：朝日出版社. 371-384.
- García de Diego, Vicente. 1970. *Gramática histórica española*. Madrid. Gredos.
- García Mouton, Pilar. 1999. *Cómo hablan las mujeres*. Madrid. Arco/Libros.
- García Mouton, Pilar. 2003. *Así hablan las mujeres. Curiosidades y tópicos del uso femenino del lenguaje*. Madrid. Las Esferas de los Libros.
- García-Page Sánchez, Mario. 2011. *Cuestiones de Morfología Española*. 2a. ed. Madrid. Editorial Centro de Estudios Ramón Areces.
- Gómez de Silva, Guido. 1988. *Breve diccionario etimológico de la lengua española*. México. El Colegio de México.
- Gómez Guinovart, Javier. 1999a. *La escritura asistida por ordenador. Problemas de sintaxis y de estilo*. Universidad de Vigo. Servicio de Publicaciones.
- Gómez Guinovart, Xavier. 1999b. "Bases lingüísticas y computacionales del procesamiento de la impropiiedad estilística y la legibilidad", *Revista española de lingüística aplicada*, Vol. Extra 1, 153-174.

<https://dialnet.unirioja.es/servlet/articulo?codigo=227029>

- Greenberg, Joseph J. 1963. "Phoneme Distribution and Language Universals", in Joseph J. Greenberg (ed.). *Universals of Language*, The Massachusetts Institute of Technology, 61-72.
- Greenberg, Joseph J. 1966. *Language Universals*. The Hague. Mouton de Gruyter.
- Guiraud, Pierre. 1957. *La Stylistique*. Presses Universitaires de France. 佐藤信夫 (訳) 1959. 『文体論—ことばのスタイル』東京：白水社.
- Gutiérrez, María Luz. 1978. *Estructuras sintácticas del español actual*. Madrid. Sociedad General Española de Librería.
- Hartfeld, Helmut. 1972. *El «Quijote» como obra de arte del lenguaje*. Madrid. Consejo Superior de Investigaciones Científicas.
- 秦隆昌 1999. 『ロマンス諸語対照スペイン語語源小辞典素案』東京：信山社.
- 波多野完治. 1958. 『ことばと文章の心理学』東京：新潮社
- 波多野完治. 1965. 『文章心理学<新稿>』東京：大日本図書.
- 波多野完治. 1988. 『文章心理学入門』東京：小学館.
- 林四郎. 1966. 「言語行動のタイプ」日本文体論協会(編)『文体論入門』東京：三省堂, 252-276.
- 林四郎. 1979. 「『坊ちゃん』の会話構成」山口仲美(編)『論集日本語研究 8. 文章・文体』東京：有精堂, 209-219.
- 林巨樹. 1979. 「井上靖の文体」山口仲美(編)『論集日本語研究 8. 文章・文体』東京：有精堂, 199-208.
- Herdan, G. 1956. *Language as Choice and Change*. Groningen: Noordhoff N. V.
- Hough, Graham. 1969. *Style and Stylistics*. Routledge. 四宮満(訳) 1972. 『文体と文体論』東京：松柏社.
- Hualde, José Ignacio, Antxon Olarrea y Anna María Escobar. 2001. *Introducción a la lingüística hispánica*. Cambridge University Press.
- Irizarry, Estelle. 1997. *Informática y literatura. Análisis de textos hispánicos*. Barcelona. Proyecto A Ediciones.
- 石橋幸太郎(編).1977. 『現代言語学辞典』東京：成美堂.
- Juilland, Alphonse and Eugene Chang-Rodríguez. 1964. *Frequency Dictionary of Spanish Words*. The Hague: Mouton.
- Justicia, Fernando. 1995. *El desarrollo del vocabulario. Diccionario de frecuencias*. Universidad de Granada.
- 樺島忠夫. 1968 『表現の解剖. 続文章工学』東京：三省堂.
- 樺島忠夫・寿岳章子. 1979. 「文体の統計的観察」山口仲美(編)『論集日本語研究 8. 文章・文体』東京：有精堂, 178-198.
- 計量国語学会. 2017. 『データで学ぶ日本語学入門』東京：朝倉書店.

- 金明哲. 2016. 「計量文献学の基礎研究とその応用」村上征勝・金明哲・土山玄・上阪彩香『計量文献学の射程』東京：勉誠出版, 57-114.
- Kock, Josse de. 1983. *Elementos para una estilística computacional*. Madrid. Editorial Cooquio.
- Kock, Josse de. (ed.) 2001. *Gramática española. Enseñanza e investigación. Apuntes metodológicos*. Ediciones Universidad Salamanca.
- Lang, Mervyn F. 1992. *Formación de palabras en español. Morfología derivativa productiva en el léxico moderno*. Madr. Cátedra.
- Lindquist, Hans. 2009. *Corpus Linguistics and the Description of English*. Edinburgh University Press. 渡辺秀樹・大森文子・加野まきみ・小塚良孝(訳)『英語コーパスを活用した言語研究』東京：大修館書店.
- Martinet, André. 1970. *Elementos de lingüística general*. Julio Calonge Ruiz (trad.) 2a ed. Madrid: Gredos.
- Martínez, Inmaculada y Hiroto Ueda. 2021a. *Inventario léxico de PRESEEA-Santander*.
<https://h-ueda.sakura.ne.jp/kenkyu/chiri/inventario-santander.pdf>
- Martínez, Inmaculada y Hiroto Ueda. 2021b. "Aspectos estadísticos del español oral. Análisis del corpus sociolingüístico de Santander (España)", *X Congreso Asiático de Hispanistas*, 25-16 de junio de 2021, Universidad Hankuk de Estudios Extranjeros, Seúl..
<https://h-ueda.sakura.ne.jp/kenkyu/goi/aspectos-estadisticos/aspectos-estadisticos.pdf>
- Martínez, Inmaculada y Hiroto Ueda. 2023. *Inventario morfológico de PRESEEA-Santander*.
<https://h-ueda.sakura.ne.jp/kenkyu/chiri/estilo/inventario-morf.pdf>
- Miles, Josephine. 1964. *Eras & Modes in English Poetry*. Berkeley. University of California Press.
- Miller, George. 1951. *Language and Communication*. Eduardo Goigorsky y Silvia Delpy (trad.) 1979. *Lenguaje y comunicación*. Buenos Aires. Amorortu editores.
- Molinié, Georges. 1992. *La Stylistique*. Presses Universitaires de France. 大浜博(訳)『文体の科学』東京：白水社.
- Moreno Fernández, Francisco. 1992. *Sociolinguistics and Stylistic Variation*. LynX, A Monographic Series in Linguistics and World Perception, 3.
- Moreno, Antonio, Guillermo de la Madrid, Manuel Alcántara, Ana González, José M. Guirao and Raúl de la Torre. 2005. "The Spanish Corpus", in Emanuela Cresti and Massimo Moneglia (eds.) *C-ORAL-Rom. Integrated*

- Reference Corpora for Spoken Romance Languages*. Amsterdam. John Benjamins Publishing Company.
- 中村明(編). 1977. 『作家の文体』 東京：筑摩書房.
- 中村明. 2016. 『日本語文体論』 東京：岩波書店.
- Navarro Tomás, Tomás. 1966. *Estudios de fonología española*. New York: Las Americas Publishing Company.
- Nieto Caballero, Guadalupe y Pablo Ruano San Segundo. 2020. *Estilística de corpus. Nuevos enfoques en el análisis de textos literarios*. Bern. Peter Lang.
- 太田強正. 2012. 『スペイン語語源辞典』 横浜：春風社.
- 岡本信照. 2021. 『スペイン語の語源』 東京：白水社.
- Parodi, Giovanni. 2007. *Working with Spanish Corpora*. New York. Continuum.
- R Core Team (2021) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
<https://www.R-project.org/>.
- Real Academia Española y Asociación de Academias de la Lengua Española. 2009. *Nueva gramática de la lengua española*. 2 vols. Madrid. Espasa Libros.
- Real Academia Española. 2023a. *Banco de datos (CREA) [en línea]. Corpus de referencia del español actual*. <<https://www.rae.es>> [2023/11/14]
- Real Academia Española. 2023b. *Banco de datos (CORPES XXI) [en línea]. Corpus del Español del Siglo XXI (CORPES)*. <<http://www.rae.es>> [2023/11/14]
- Ríos Hernández, Iván Neftalí. 2017. "Un acercamiento a la legibilidad de textos relacionados con el campo de la salud", Quito, Ecuador : CIESPA.
<https://repositorio.flacsoandes.edu.ec/handle/10469/13252>
- Rojo, Guillermo. 2021. *Introducción a la lingüística de corpus en español*. London. Routledge.
- 齊藤俊雄(編). 1992. 『英語英文学とコンピュータ』 東京：英潮社.
- Saporta, Sol. 1959. "Morpheme Alternants in Spanish", in Henry R. Kahane and Angelica Pietrangeli (eds.), *Structural Studies on Spanish Themes*, University of Illinois Press, 15-162.
- Saporta, Sol. 1963. "Phonemes Distribution and Language Universals", Joseph H. Greenberg (ed.) *Universals of Language*. Cambridge. The M.I.T. Press, 61-72.
- Serrano-Dolader, David. 1999. "La derivación verbal y la parasíntesis", en Ignacio Bosque y Violeta Demonte (dir.). 1999. *Gramática descriptiva de la lengua española*, vol. 3, Madrid. Espasa, 4683-4755.
- Stubbs, Michael. 2002. *Words and Phrases: Corpus Studies of Lexical Semantics*.

- Blackwell Publishing*. 南出康世・石川慎一郎(訳). 2006 『コーパス語彙意味論：語から句へ』 東京：研究社.
- 田中久美子. 2021. 『言語とフラクタル』 東京大学出版会.
- 田中春美(主幹). 1988. 『現代言語学辞典』 東京：成美堂.
- 寺崎英樹. 2019. 『スペイン語文法シリーズ.2.語形変化・語形成』 東京：大学書林.
- Tonko Milic, Louis. 1967. *A Quantitative Approach to the Style of Jonathan Swift*. The Hague. Mouton.
- 上田博人. 1987. 『スペイン語の語彙の頻度と拡がり』 東京外国語大学語学研究所.
<https://h-ueda.sakura.ne.jp/kenkyu/goi/frec-disp/frec-disp-0.pdf>
- 上田博人. 2016. 『言語研究のための数値データ分析法』
<https://h-ueda.sakura.ne.jp/gengo/>
- Ueda, Hiroto y Antonio Moreno Sandoval. 2017. *Análisis de datos cuantitativos para estudios lingüísticos*.
<https://h-ueda.sakura.ne.jp/gengo/4-numeros/doc/numeros-es.pdf>
- Whaley, Lindsay J. 1997. *Introduction to Typology. The Unity and Diversity of Language*. Thousand Oaks, California. Sage Publications. 大堀壽夫・古賀裕章・山泉実(訳). 2006. 『言語類型論入門』 東京：岩波書店.
- WhatMough, J. 1956. *Language*. 蛭沼寿雄・久野暉訳 1960 『言語 現代における総合的考察』 東京：岩波書店.
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York. Springer-Verlag.
- Wierzbicka, Anna. 1967. *I jezyku dla wszystkich*, Warszawa. 小原雅俊・石井哲士朗・阿部優子訳. 2011. 『アンナ先生の言語学入門』 東京外国語大学出版会.
- Williams, C. B. 1970. *Style and Vocabulary. Numerical Studies*. New York. Hafner Publishing.
- 安本美典. 1960. 『文章心理学の新領域』 東京：創元社.
- 安本美典. 1965. 『文章心理学入門』 東京：誠信書房.
- 安本美典. 1977. 「現代の文体研究」 『岩波講座日本語 10 文体』 東京：岩波書店, 395-423.
- 安本美典. 2009. 「計量文体論・文章心理学」 計量国語学会 『計量国語学事典』 東京：朝倉書店, 253-273.
- 山本貴光. 2014. 『文体の科学』 東京：新潮社.
- 吉岡健一. 1996. 「計量文体学研究の展望」 アンソニー・ケイ 『文章の計量』 (吉岡健一訳) 東京：南雲堂.

Zipf, George Kingsley. 1936. *The Psycho-biology of Language. An Introduction to Dynamic Philology*. London: Routledge.

Zipf, George Kingsley. 1949. *Human behavior and the Principle of Least Effort. An Introduction to Human Ecology*. MansfielsAddison-Weskey Press.