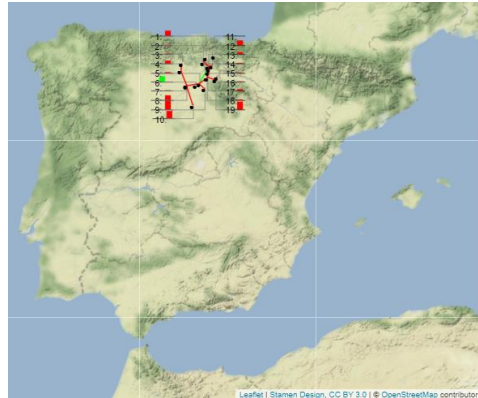


最近隣法による文字変異の地図化と語彙拡散の計量化

古スペイン語重子音文字 *ss* の歴史・地理的解釈

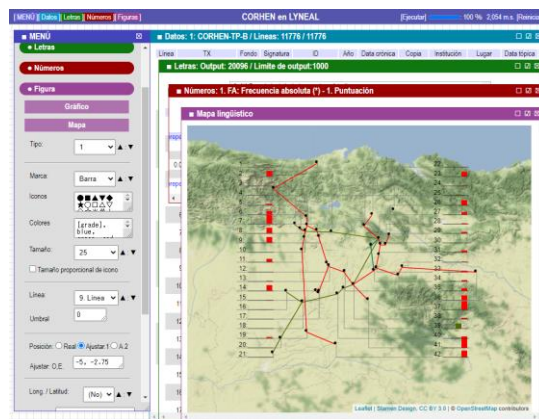
上田博人 (東京大学)・Leyre Martín Aizpuru (Universidad de Sevilla)

§ 1. スペイン語の歴史の初期(1000-1250)の様相はイベリア半島北部 (図 1)の各地で発行された 9-13 世紀の公証文書の中で観察される。



【図 1】

これらの文書を収録した歴史言語コーパス CORHEN («Corpus Histórico del Español Norteño», María Jesús Torrens Álvarez, Spain, CSIC)では転写された文書の発行地と年代が示されている。私たちはその検索・統計・グラフ・地図化システムを担当してきた(LYNEAL «Letras y Números en Análisis Lingüísticos»: 図 2)。



【図 2】

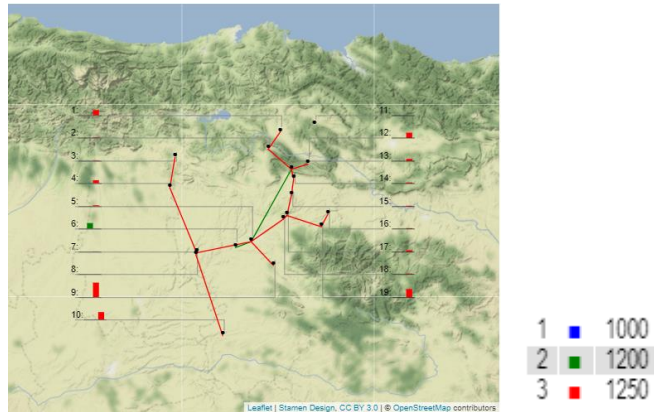
§ 2. 本発表ではこの資料を使って古スペイン語に見られる重子音文字 *ss* を扱う(図 3)。

【図 3】  *ssellada con su ssello* (CODEA-8, Valladolid, 1283)

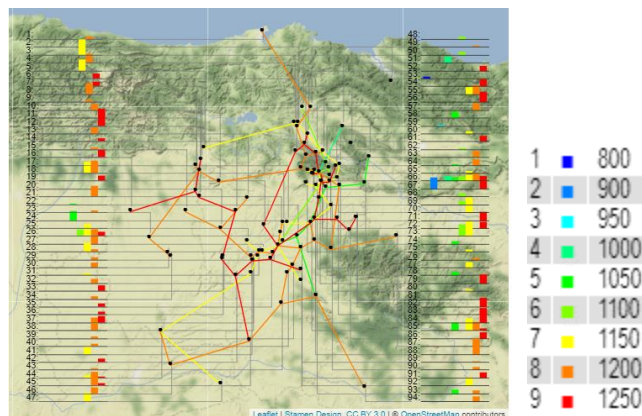
先行文献では『死の舞踏』(*Dança general de la muerte*, Sevilla, 1520)における *sson* (動詞 *ser*): *son* 「音」の意味弁別 (Morreale, 1996: 158, 173), 語中子音+*ss* (*Perssia*, *penssar* 「考える」)との関連(Sánchez-Prieto Borja, 1998: 138; Martín Aizpuru, 2019: 353-354), 筆記体文字の特徴(Sánchez-Prieto Borja, 2004: 442), 筆記者の特徴(Martín Aizpuru, 2019: 350-351)などが指摘された。加えて私たちは語中の-*ss*-との関連を見な

がら、その歴史・地理的解釈を試みる。

【図 4】の地図は語頭に#ss がある単語の相対頻度分布を示し(#ss / (#ss + #s)), 【図 5】の地図は語中に-ss-がある単語の相対頻度分布(-ss- / (-ss- + -s-))を示す。



【図 4】 #ss / (#ss + #s)



【図 5】 -ss- / (-ss- + -s-)

2 つの地図を比べると語頭の#ss の分布が広範囲に散発的であったことと(図 4), 語中の-ss-が広範囲に多発的であったことがわかる(図 5)。結線については後述する(§3)。

Trask (2000, *s.v. four-part analogy*)は「4 部類推」の例として東部アメリカ英語方言の *dive* の過去形 *dove* が(1)の式で説明しているが、これを(2)のように書き換える。

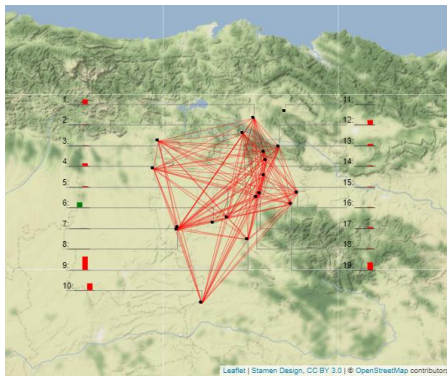
(1) A : B :: C : ? *drive : drove :: dive : ?* ? = *dove*

(2) Xi : Xj :: Yi : ? *drive : drove :: dive : ?* ? = Yj = *dove*

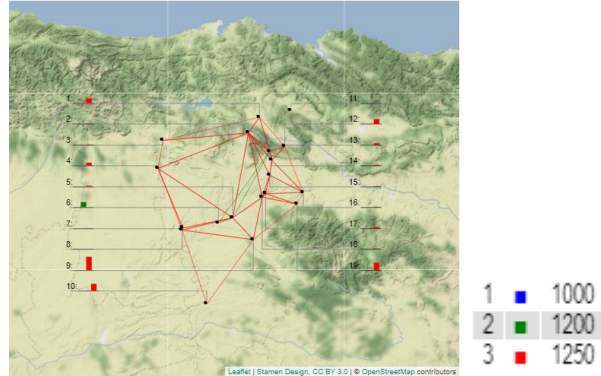
(2) の X と Y は動詞 *drive* と *dive* を示し、i と j は現在形と過去形を示す。この類推関係と同様にして、X と Y を語中と語頭とし、i と j を音韻と文字とすれば、次の並行関係が成り立つ: $-/s/ : -ss- :: \#s : ?$, $? = \#ss$ 。このとき、多くの場合、左辺の要素(要因 = $-/s/ : -ss-$)は右辺(結果 = $\#s / \#ss$)よりも、十分に優勢(広範囲多発)である(図 4, 5)。

§3. 上の地図(図 4, 5)では該当項目が出現した地点の間を最近隣線で結合した。LYNEAL システム内の地図視覚化プログラムはそれぞれの地点で回りの近隣地点の中から最短距離にある地点と結合させる(クラスター分析の「最近隣法」(nearest neighborhood)を用いた)。言語現象は一般に近隣に広がるので最近隣線は現実的である。地点を結ぶネットワーク(図 6)は混雑しやすいが、最近隣線はシンプルである(図 4, 5)。デロネイ図(図 6)はネットワークに比べればかなり見やすいが、それでもやや煩雑で

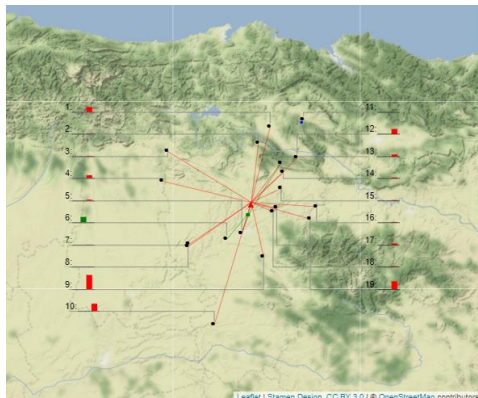
ある。放射線（図 8）は、その中心が示す言語的な意味がわかりにくい。外側の反応点を凸包（convex hull）の面で囲むことは面内に例外を多く含むとき、過剰に一般化するため不都合である（図 9, 1250 年代の地点 6）。ボロノイ図は広大な面を 1 点で代表させているので、やはり過剰に一般化している（図 10）。



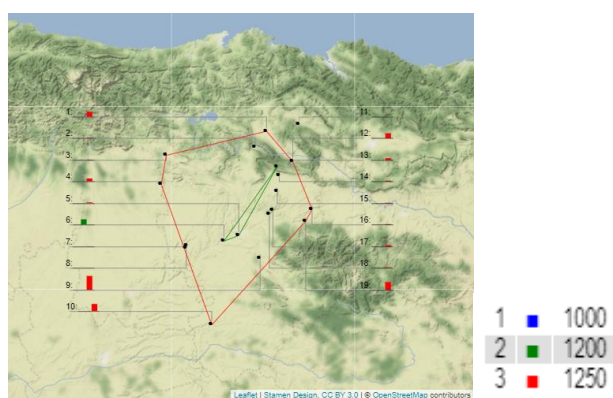
【図 6】 ネットワーク



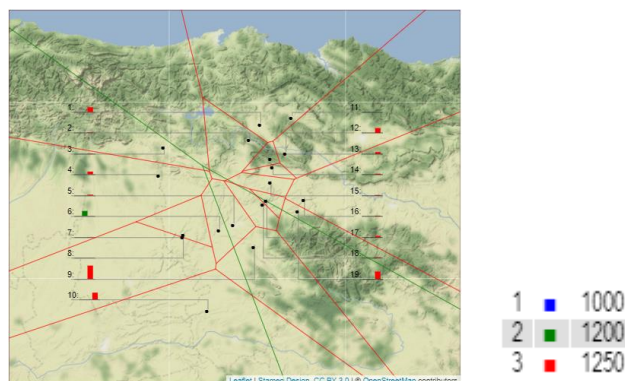
【図 7】 デロネイ図



【図 8】 放射線



【図 9】 凸包面



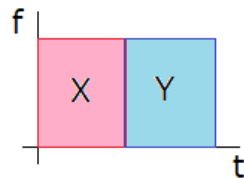
【図 10】 ボロノイ図

§4. 12-13 世紀に出現した母音間の-ss-は無声の/s/を示し(*passar* /pa'sar/ 「通る」), 有声の s (/z/)と弁別されていたが(*casa* /'kaza/ 「家」), 14 世紀以降両者は s (/s/)に統一された(*pasar* /pa'sar/, *casa* /'kasa/). 以下では消失した ss の全体(語中と語頭)を扱う。

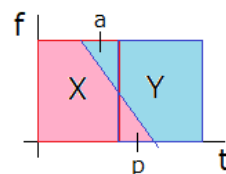
19 世紀の「青年文法学派」(Junggrammatiker, Neogrammarian)の言語学では音韻

変化は原則としてすべての単語に一律に例外なく生じる，とされた(Osthoff and Brugmann, 1878, cf. Lehmann 1967: 204)。一方，20世紀に，音韻変化の進行は単語ごとに異なる，という「語彙拡散」(lexical diffusion)が主張された(Wang, 1969)。

私たちは両者を二律背反とせず，段階的な解釈を採る。はじめに，音韻変化 $X \rightarrow Y$ がすべての単語に一律に例外なく生じる状態を【図 11】に示す(f: 頻度, t: 時間軸)。実際は X に音韻変化後に残存する語(p)があり，また Y に音韻変化の前に先行する語(a)がありうる(図 12)。



【図 11】 $X \rightarrow Y$

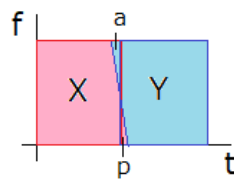


【図 12】 a: 先行例外 / p: 後続例外

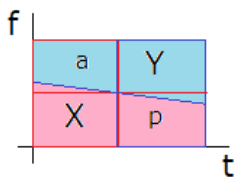
そこで，語彙拡散度(D)を次のように定義する：

$$D = (a + p) / (X + Y)$$

D が最小($\rightarrow 0$)になるのは， $a = p \rightarrow 0$ のときであり(図 13)，最大($\rightarrow 1$)になるのは $a + p = X + Y$ のときである(このとき X と Y が並立し， $X \rightarrow Y$ という変化がない：図 14)。

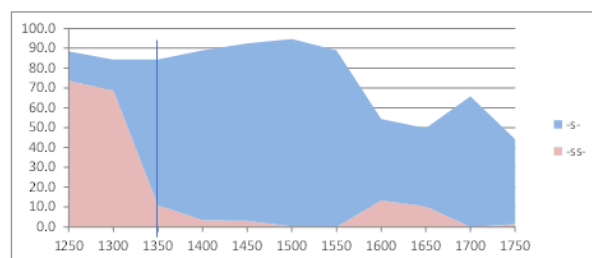


【図 13】 最小の $D = 0$ に近似



【図 14】 最大の $D = 1$ に近似

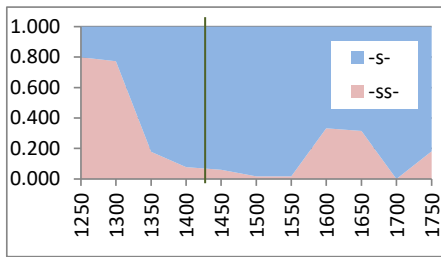
この語彙拡散度(D)を音韻変化ではなく文字変化(ss \rightarrow s)に適用すると，【図 15】のようになる(資料：CODEA «Corpus de Documentos Españoles Anteriores a 1800», Pedro Sánchez-Prieto Borja y GITHE)。



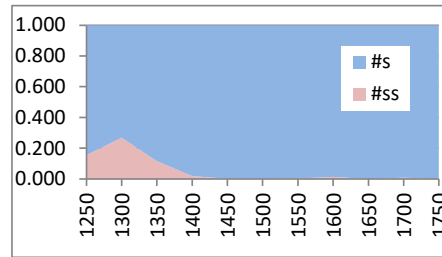
【図 15】 語中の -ss- \rightarrow -s-, $D = .087$

しかし，出現頻度(絶対頻度)を使って求めた語彙拡散度(D)は，出現頻度が異なるケースの語彙拡散度と比較できない。そこで，出現頻度(絶対頻度)の代わりに比較する語形の相対頻度を用いると，相対頻度の和はどの年代でも 1 になるので，全年代の相対頻度の和は年代数(N)と等しくなる。よって，相対頻度の語彙拡散度(D')の式は次のように簡単になる。

$$D' = (a + p) / N$$



【図 16】 語中の-ss- → -s-, $D' = 0.146$



【図 17】 語頭の#ss → #s, $D' = 0.053$

【図 17】では世紀を通じて#sが優勢であったので、 D' を $\#ss / (\#ss + \#s)$ で計算した。すなわち、最初に(1250) #ss → #sの変化があった、と解釈した。

§5. 上の【図 16, 17】によれば 1550 まで-ss-と#ssは並行していた。しかし、語中の-ss-の相対頻度を見ると 1250, 1300 の年代で優勢であったが(図 16), 語頭では#ssが優勢になることはなかった(図 17)。よって、-ss-が#ssの要因であり、その逆ではないと考えられる。

最後に CODEA の 1050-1250 の年代の 624 文書について-ss-と#ssの有無(0 : 1~)を調べ次の表を作成した。これによれば、-ss-:0 と#s:0, および、-ss-:1~と#ss:1~の一致の頻度(123, 177)が有意に高いことがわかる($\chi^2 = 40.6, p < 0.01$)。よって、同一文書内に-ss-が使われていなければ#ssも使われていない、そして-ss-が使われていれば#ssも使われている、という傾向があった、といえる。

文書数	#ss: 0	#ss: 1~	和
-ss- : 0	123	10	133
-ss-: 1~	314	177	491
和	437	187	624

期待値	#ss: 0	#ss: 1~	和
-ss- : 0	93.1	39.9	133.0
-ss-: 1~	343.9	147.1	491.0
和	437.0	187.0	624.0

<引用文献>

- Lehmann, W. P. (1967), *A Reader in Nineteenth Century Historical Indo-European Linguistics*. Bloomington: Indiana University Press.
- Martín Aizpuru, L. (2019), *La «norma lingüística» de la cancellería real castellana (1230-1312): Fernando III, Alfonso X, Sancho IV y Fernando IV*, Tesis doctoral, Universidad de Salamanca.
- Morreale, M. (1996), «La Dança general de la muerte (II)», *Revista de Literatura Medieval*, VIII, pp. 111-177.
- Sánchez-Prieto Borja, P. (1998), *Cómo editar los textos medievales. Criterios para su presentación gráfica*. Madrid. Arco / Libros.
- _____. (2004), «La normalización del castellano escrito en el siglo XIII. Los caracteres de la lengua: Grafías y fonemas» en Rafael Cano (coord.) *Historia de la Lengua Española*, pp. 423-448. Barcelona. Ariel.
- Trask, R. L. (2000), *The Dictionary of Historical and Comparative Linguistics*, Edinburgh University Press.
- Wang, W. S-Y. (1969), «Competing Changes as a Cause of Residue», *Language*, vol. 45, pp.9-25.