

地理空間距離と言語距離 - ロマンシュ語圏を例に -

東京外国語大学 語学研究所 特別研究員: 清宮貴雅

ワークショップ ことばの「距離」を考える

2026年8月1日(土)

於西南学院大学大学院講義室202・203号室



01

背景（計量方言学の手法）

02

目的

03

方法

04

結果

05

考察と結論

06

参考文献

01 計量方言学の手法

類似度指標 / 距離指標 について

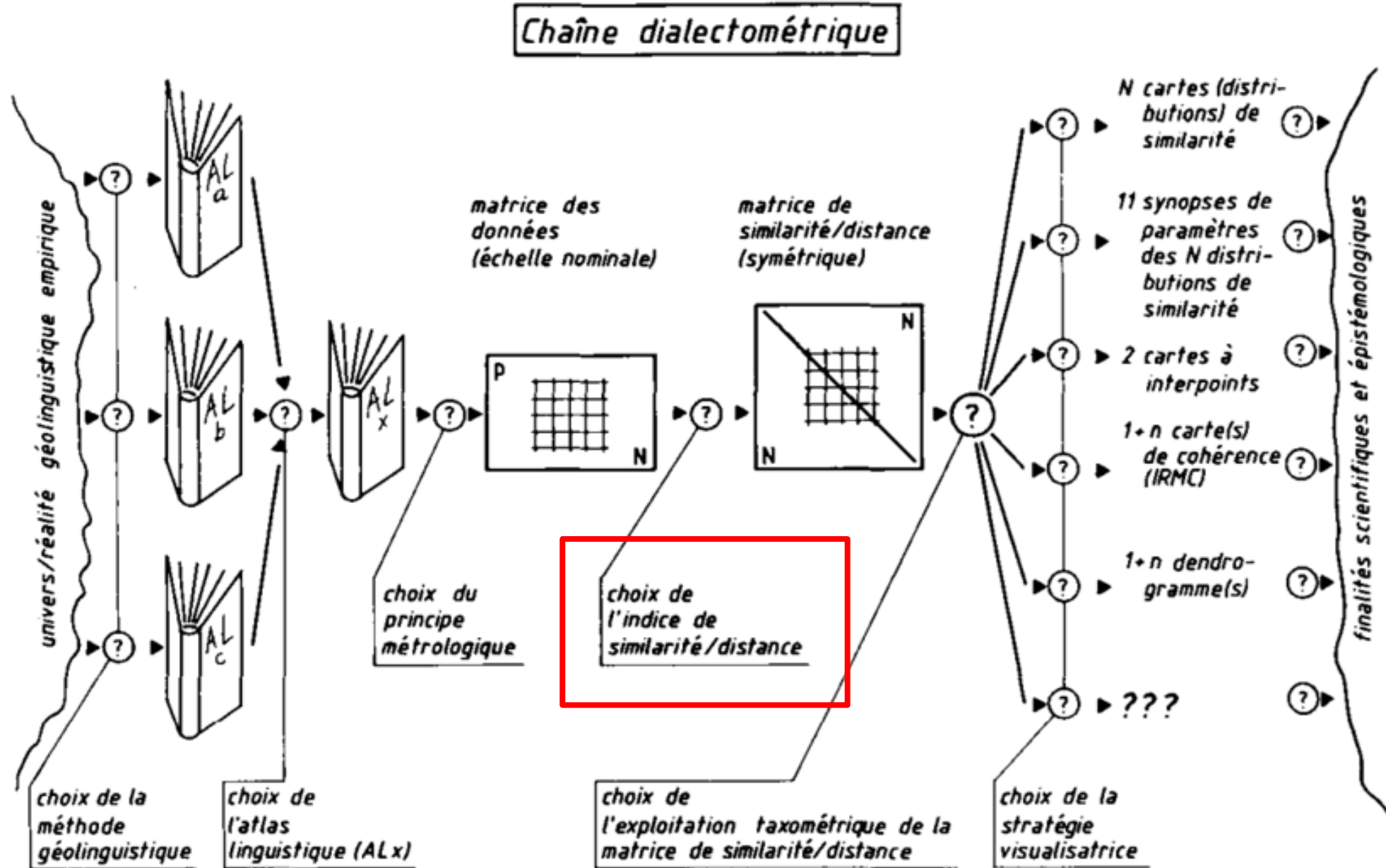


図1 計量方言学的分析の処理手順 (Goeb1 1992:431より)

Background

類似度指標 / 距離指標

■ レーベンシュタイン距離 による数値化

- ・ Nerbonne et al. (1999)、Heeringa & Nerbonne (2001)、Valls et al. (2012)、清宮 (2020) など
- ・ 2つの文字列を同じ形にするために必要な、最小の編集[=挿入、削除、置換]回数を言語間の距離とする

例)

1.
猫 [kat]

2.
[kæt]

3.
[kɛət]

[kat] → [kæt] : [a]と[æ]の置換 = 距離 1

[kat] → [kɛət] : [a]と[ɛ]の置換 + [ə]の挿入 = 距離 2

[kɛət] → [kæt] : [ɛ]と[æ]の置換 + [ə]の削除 = 距離 2

欠点

- ・ 同根語 (cognates) でない語形の比較に不向き (Dunn 2014 : 195, Greenhill 2011 : 694)
- ・ 音声学的に似た音であるかどうかを考慮しない

Background

類似度指標 / 距離指標

■ 重み付きレーベンシュタイン距離 による数値化

- ・ 鍵水 (2013) など
- ・ 挿入・削除・置換に一律のコストを与えるのではなく、操作の種類や文字の組み合わせごとに異なるコストを設定する
- ・ 弁別素性・音響音声学的類似性・調音音声学的類似性などから、各分節音に数値を与え、各分節音間のコストを計算する
- ・ 音素間の類似性に応じてコストを変えるため、音声的な差を反映しやすい

欠点

- ・ 同根語でない語形の比較に不向き

Background

類似度指標 / 距離指標

■ バイナリ距離 による数値化

- ・ Séguy (1971)、Goebel (1992)、Lafkioui (2008) など
- ・ 2つの文字列が一致していれば距離 0、不一致であれば距離 1 とする
- ・ 一致/不一致の基準は研究者によって異なる

例)	1.	2.	3.	4.	5.
魚	[pwæsõ]	[pwæsõ]	[pɛsõ]	[pɛʃ]	[tryte]

欠点

- ・ 語形の比較 / 語源の比較のどちらかのみ分析可能
- ・ 言語情報をフルで使うことができない

[pwæsõ] 対 [pwæsõ] : 一致 = 距離 0
[pwæsõ] 対 [tryte] : 不一致 = 距離 1
[pwæsõ] 対 [pɛsõ] : 一致 (同根語 / 同じ語)
不一致 (音声)
[pwæsõ] 対 [pɛʃ] : 一致 (同根語)
不一致 (別の語)

Background

類似度指標 / 距離指標

■ 段階的非類似度 による数値化

- ・ Yarimizu et al. (2004)、Kawaguchi (2007)、清宮 (2020) など
- ・ 2つの文字列の一致度を複数段階で数値化する
- ・ 一致/不一致の基準、段階の数は研究者によって異なる

例) 1. 2. 3. 4. 5.
 魚 [pwæsõ] [pwæsõ] [pesõ] [peɸ] [tryte]

欠点

- ・ 各段階へ与える数値が恣意的になりやすい
- ・ 適切な段階数を設定しにくい

[pwæsõ] 対 [pwæsõ] : 一致
[pwæsõ] 対 [pesõ] : 部分一致 (同じ語)
[pwæsõ] 対 [peɸ] : 部分一致 (同根語)
[pwæsõ] 対 [tryte] : 不一致

Background

類似度指標 / 距離指標

■ 各数値化の利点/欠点

	レーヴェンシュタイン距離	重み付け レーヴェンシュタイン距離	バイナリ距離	段階的非類似度
利点	客観的(機械的)に距離を計算	音声的類似性を加味して、客観的(機械的)に距離を計算	あらゆる語形を比較可能	あらゆる語形を比較可能 言語情報を最大限に活用可能
欠点	同根語間以外の比較が不可能 音声的類似性を無視	同根語間以外の比較が不可能 重み付けが複雑	言語情報をフル活用しにくい	各段階へ与える数値が恣意的 適切な段階数の設定が難しい

Background

物理距離と言語距離の相関

■ Séguy (1971)

- ・ 地点間の言語距離は、物理距離が大きくなるにつ入れて増加する
- ・ 増加率は次第に小さくなる

■ Heeringa & Nerbonne (2001)

- ・ 方言差は地理的に連続して増加する
- ・ 線形よりも対数的に増加する

■ Pickl et al. (2014)

- ・ 同じ地理距離でも、川を挟む地点間では言語距離が大きくなる

■ 清宮 ほか (2026)、持橋 ほか (2026) [24K23937 非公開研究会での報告]

- ・ 地点間の言語距離は、物理距離が大きくなるにつ入れて増加する傾向にある
- ・ 谷や尾根の開析度（標高など）や傾斜も言語距離に影響を与える

本研究の目的

段階的非類似度の一致から不一致までを最大4段階に設定した場合
各段階にどの程度の数値を与えるべきか検証する

- ・物理距離(km)を固定した状態で、重み付けを変更し、比較する
- ・山々で隔てられる、ロマンシュ語の方言で分析をする

03 方法

Methods

分析手順

01 データ収集

言語データ

- ・AIS(言語地図)
- ・語源辞典

地理データ

- ・Google maps
- ・OpenStreetMap

02 段階 / 数値 の設定

→ 各段階に割り当てる
数値を複数セット作成

例) 4, 3, 2, 0
4, 3, 1, 0 など

地点間の語形の関係
性を数値化 (距離)

03 言語距離 / 物理距離の計算

言語距離

- ・語ごとの距離の和
を0~1に平均化

物理距離

- ・foot-hikingから各地
点間の移動距離
(km)を計算

04 統計分析

→ 言語距離行列と
物理距離行列で
Mantel test を実行

Mantel r を算出

02を複数回変更し、
02~04を複数回実行

Methods

01 データ収集 (1/3)

■ Jaberg & Jud (1928-1940) *Sprach- und Sachatlas Italiens und der Südschweiz*
『イタリアおよび南スイス言語・民俗地図』 (AIS)

- ・ 対象地域：イタリア共和国、スイス南東部、クロアチア・スロベニアの一部地域
- ・ 調査地点数：450地点
- ・ 調査時期：1908～1927
- ・ 1705枚の言語地図
- ・ ロマンシュ語圏は全19地点

Methods

01 データ収集 (2/3)

- 言語データ
 - ・ 100語版スワデシュリスト (Swadesh 1955) に記載されている語から56コンセプト (=語) の地図

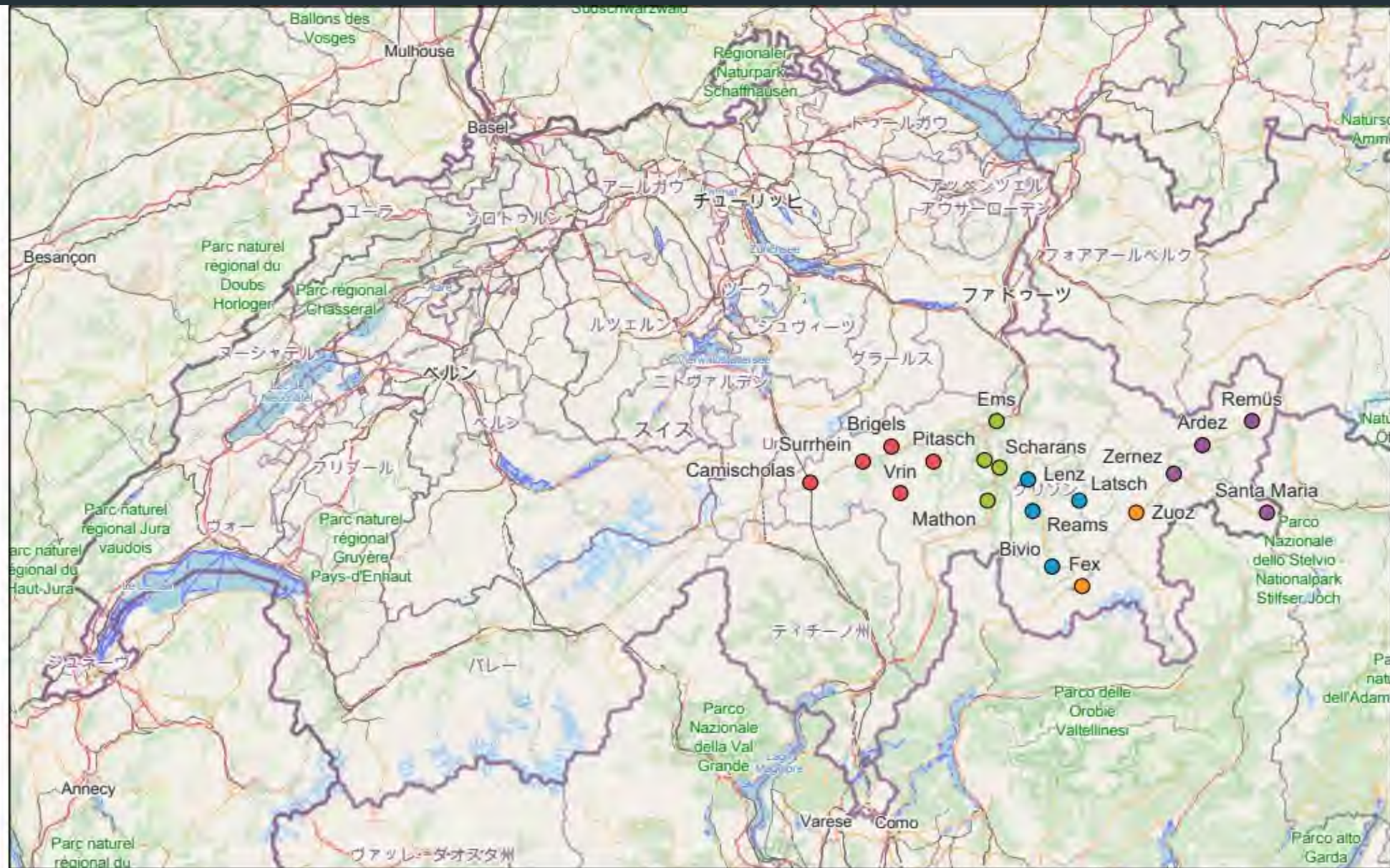
Category	<i>n</i>	
Quantity_Number	3	many, one, two
Human_Animal	5	man, fish, bird, dog, louse
Plant_Organic	10	leaf, root, bark, skin, meat, blood, bone, egg, tail, feather
Body	13	hair, head, ear, eye, nose, mouth, tongue, nail, foot, knee, neck, heart, liver
Verbs	5	to drink, eat (imperative), to sleep, to die, to sit down
Nature	11	sun, moon, star, water, stone, sand, land, smoke, fire, road, mountain
Colors	5	red, green, yellow, white, black
Qualities	4	cold, plenty, good, dry
ALL	56	

■言語データ

- ・各地図の語形をIPAへ変換＞語源辞書などで語源を推定＞コーパス化（19*56のデータセット）

地点1 ***sɛ́zar gǐw*** = [sɛːzər ɟiw] (= *senser giu*) < [surs. *ser* < lat. *sēdēre*] + [lat *dēorsum*]

- 16



29/07/2026

romansh_data

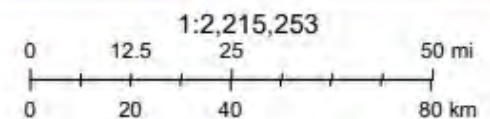
● sursilvan

● surmiran

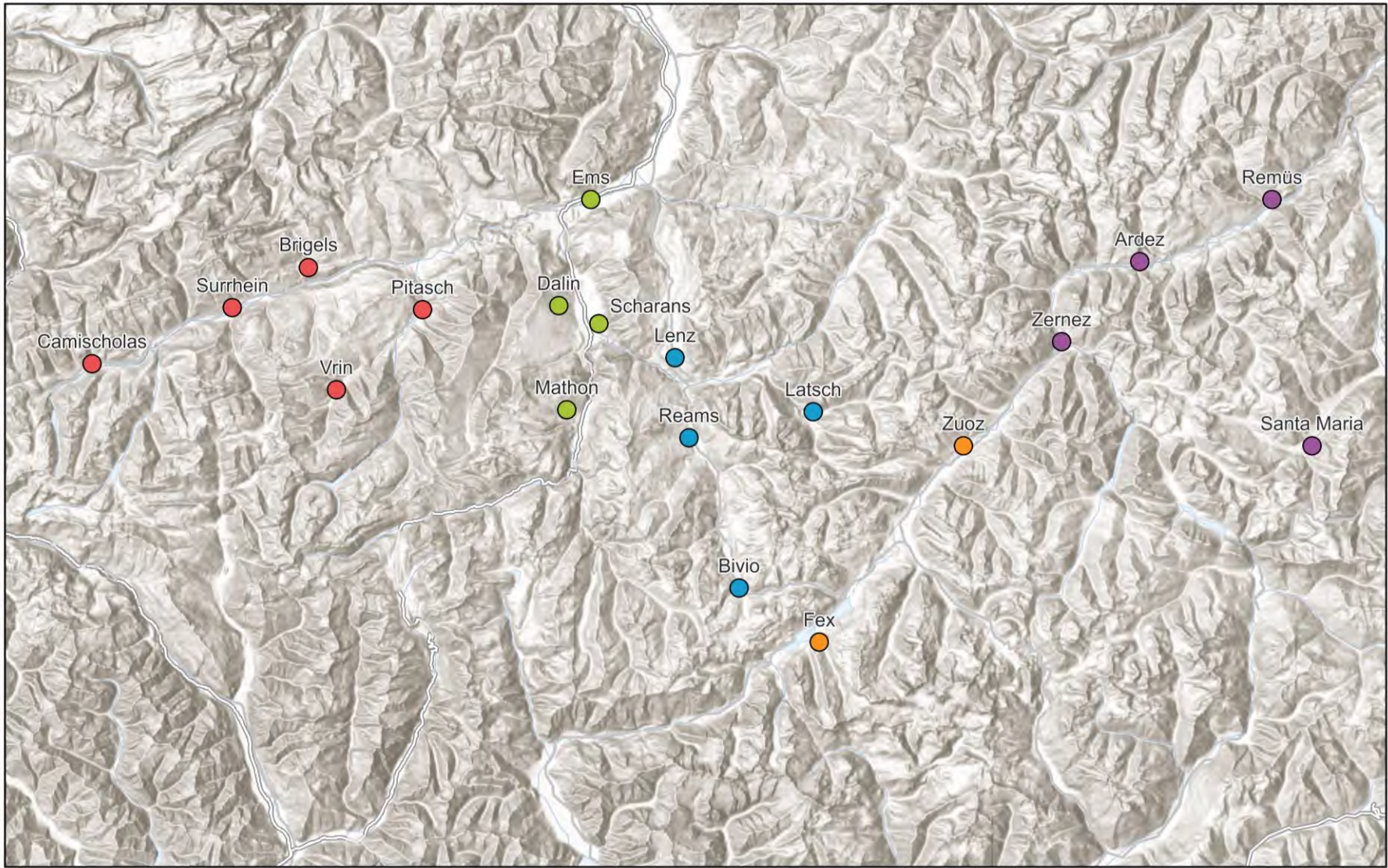
● sutsilvan

● vallader

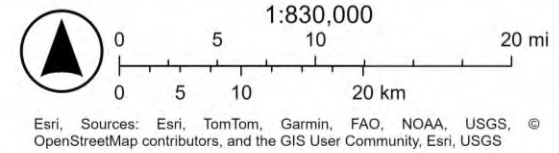
● puter



Map data © OpenStreetMap contributors, Microsoft, Facebook, Google, Esri
Community Maps contributors, Map layer by Esri



29/07/2026



Methods

02 段階 / 数値の設定 (1/4)

■ 段階の設定

例)	1.	2.	3.	4.	5.	6.*
魚	[pwæsõ]	[pwæsõ]	[pɛsõ]	[pɛʃ]	[tryte]	[powason]

lat. *piscis* > > 4. [pɛʃ]
 > afr. *peis* ON > 1., 2. [pwæsõ], 3. [pɛsõ] > 6. [powason]

lat. *tructa* > 5. [tryte]

同一語源 + 音声一致	: 1. – 2.
同一語源 + 音声不一致	: 1. – 3., 2. – 3.
同一語源 (別派生 / 借用)	: 1. – 4., 1. – 6., 2. – 4., 2. – 6., 3. – 4., 3. – 6., 4. – 6.
別語源	: 1. – 5., 2. – 5., 3. – 5., 4. – 5., 5. – 6.

Methods

02 段階 / 数値の設定 (2/4)

■ 段階の設定

- ・ 地点間の語形の関係性を名義尺度A～Dに変換

A (= 完全一致)	: IPA一致 + etymology一致
B (= 音声変異)	: IPA不一致 + etymology一致
C (= 同一語源)	: IPA不一致 + etymology部分一致
D (= 不一致)	: IPA不一致 + etymology不一致

例)	魚	IPA	語源	
	1.	[pwæsõ]	afr. peis ON lat. piscis	} A B C D
	2.	[pwæsõ]	afr. peis ON lat. piscis	
	3.	[pɛsõ]	afr. peis ON lat. piscis	
	4.	[pɛʃ]	lat. piscis	
	5.	[tryte]	lat. tructa	

Methods

02 段階 / 数値の設定 (3/4)

■数値の設定

- ・ 演算のためA、B、C、Dに値を付与
- ・ 一致度が高いほど高得点になるように設定（最終的に非類似度に変換）
 - > Aが最も値が高く、Dが最も値が低い

Methods

02 段階 / 数値の設定 (4/4)

	重み付けの説明	A	B	C	D
パターン1	完全一致かどうか	4	0	0	0
パターン2	同じ語であるかどうか	4	4	0	0
パターン3	同じ語であるか、同根語であるか ($A=B$ 、 $C>1/2*A$)	4	4	3	0
パターン4	同じ語であるか、同根語であるか ($A=B$ 、 $C=1/2*A$)	4	4	2	0
パターン5	同じ語であるか、同根語であるか ($A>B>C$ 、 $C=1/2*A$)	4	3	2	0
パターン6	同じ語であるか、同根語であるか ($A>B>C$ 、 $C<1/2*A$)	4	3	1	0
パターン7	同じ語であるか、同根語であるか ($A>B>C$ 、 $C<1/2*A$) 不一致との差を増加	5	4	2	0
パターン8	同じ語であるか、同根語であるか ($A>B>C$ 、 $C=1/2*A$) 不一致との差を増加	5	4	2.5	0
パターン9	同じ語であるか、同根語であるか ($A>B>C$ 、 $C=1/2*A$) 同じ語である場合の差を減少 不一致との差を増加	5	4.5	2.5	0

Methods

03 言語距離 / 物理距離の計算 (1/3)

■ 言語距離の計算

- ・ 合計得点ではパターン間の単純比較ができないため、非類似度を0～1の間に調整
- ・ 地点 x と地点 y を比較し、語彙項目 j の点数を S_j 、ABCDの最大値を w_{max} とすると、非類似度は：

$$\text{distance}(x, y) = 1 - \frac{\sum_{j=1}^n S_j}{w_{max} \times n} \quad \Leftrightarrow \quad \text{distance}(x, y) = 1 - \frac{\text{合計点}}{\text{最大可能点}}$$

例) $A=4$ であり、2地点間の56語の合計点（類似度）が195点だった場合

$$\begin{aligned} \text{distance}(x, y) &= 1 - \frac{195}{4 \times 56} \\ &= 0.1295 \end{aligned}$$

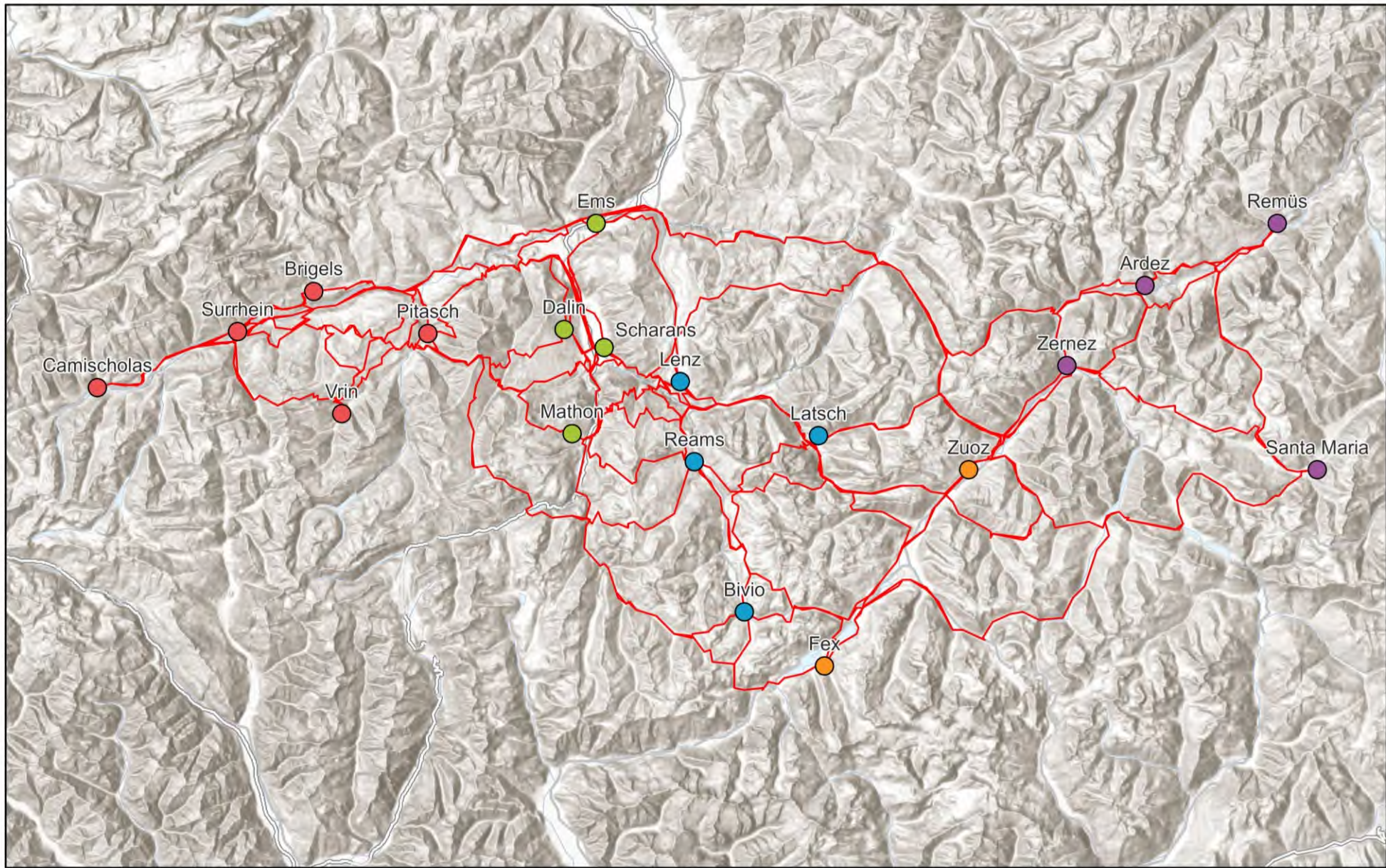
-
- ・ 地点間（19地点×19地点）の対称距離行列（言語距離）を作成

Methods

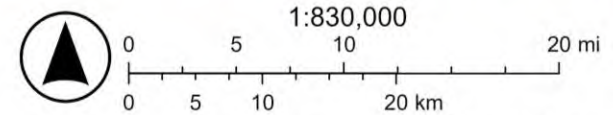
03 言語距離 / 物理距離の計算 (2/3)

■ 物理距離の計算

- ・ 19地点のすべての組合せ、合計171組について、2地点間の経路を計算
- ・ OpenStreetMapの道路・歩道・登山道データを利用するopenrouteserviceを使用
- ・ 条件 foot-hiking：徒歩向きの道
 recommended：最短距離ではなく、ハイキングに適した経路
- ・ 計算から、各径路の距離をkmで取得
- ・ 地点間（19地点×19地点）の対称距離行列（物理距離）を作成



30/07/2026



Esri, Sources: Esri, TomTom, Garmin, FAO, NOAA, USGS, © OpenStreetMap contributors, and the GIS User Community, Esri, USGS

Methods

04 統計分析

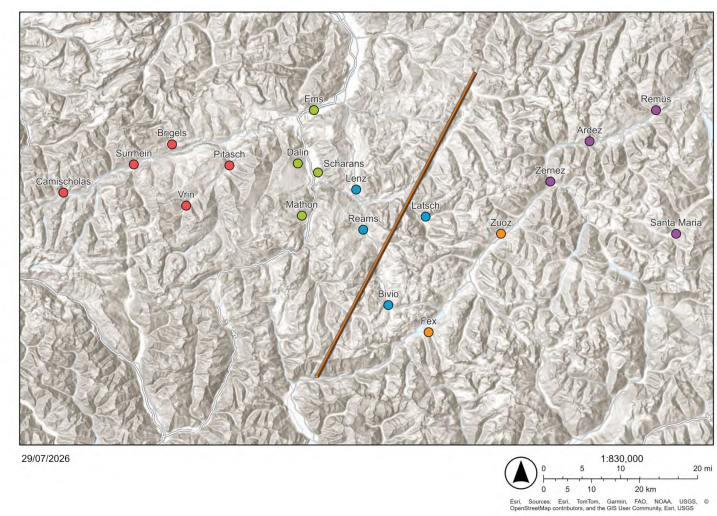
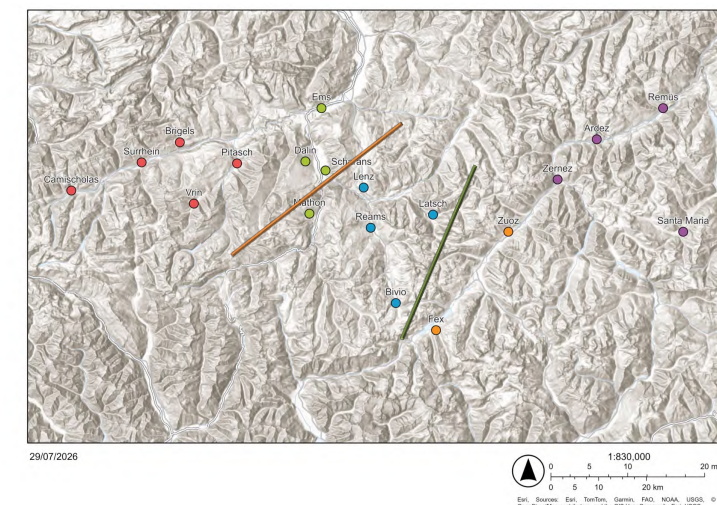
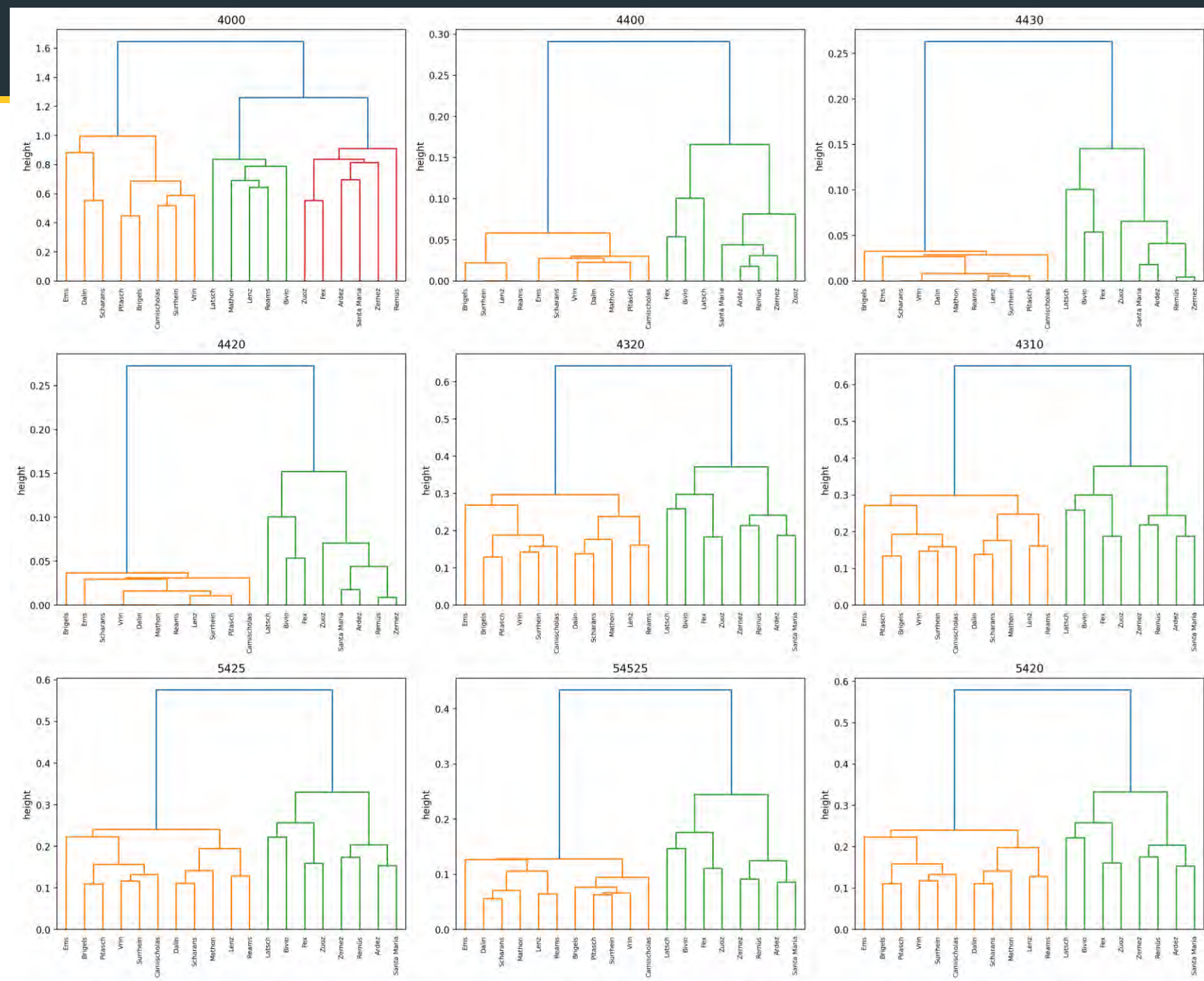
■ 階層的クラスタリング

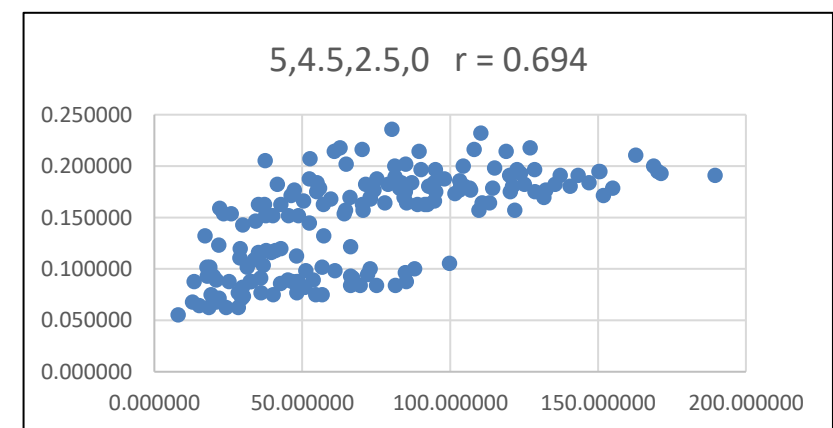
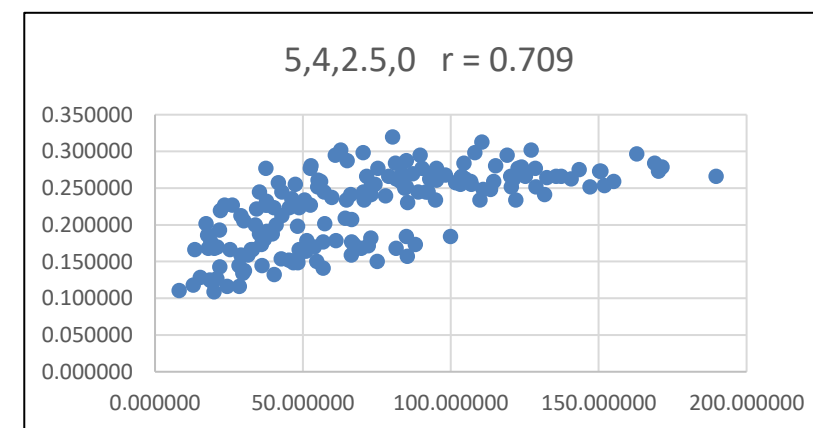
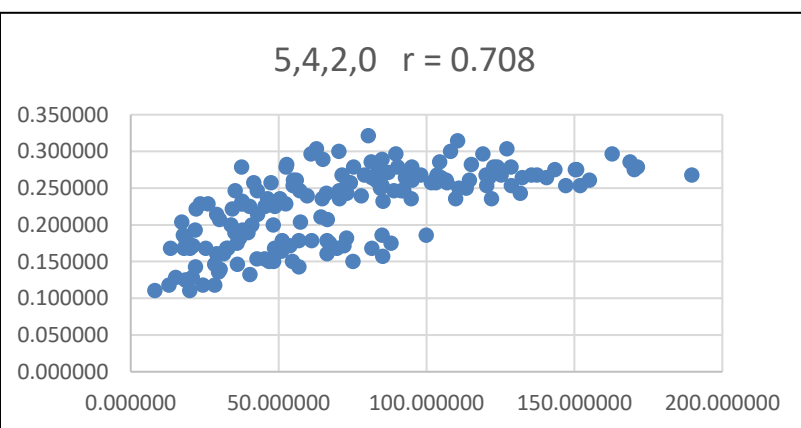
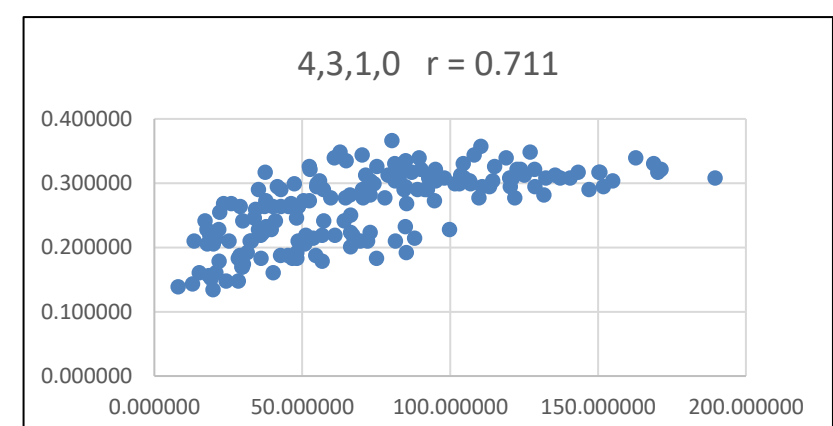
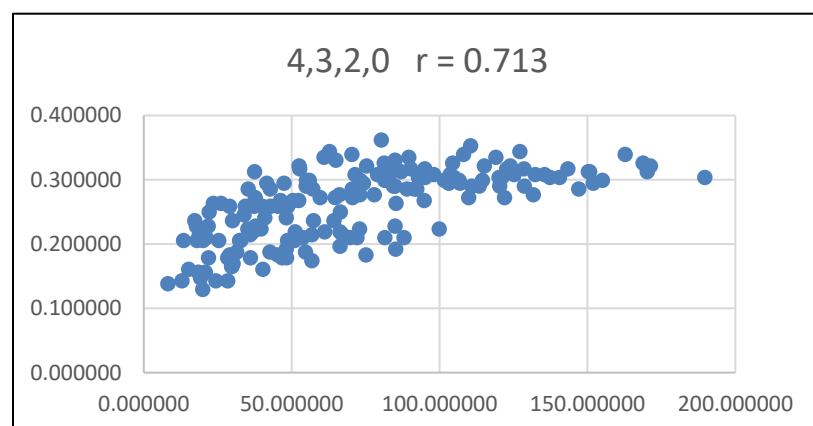
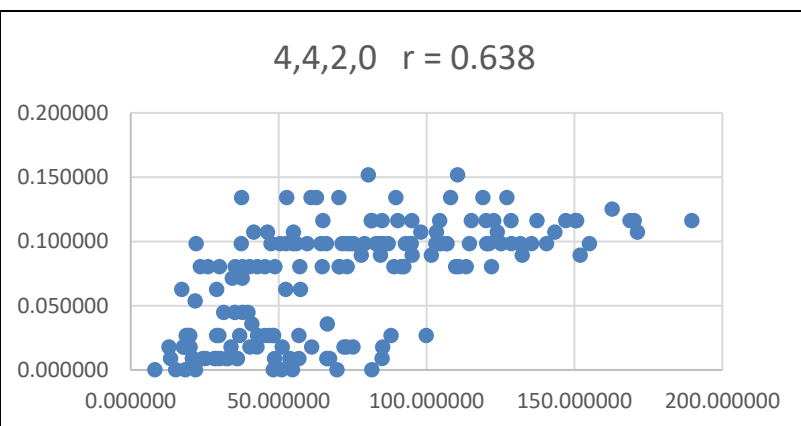
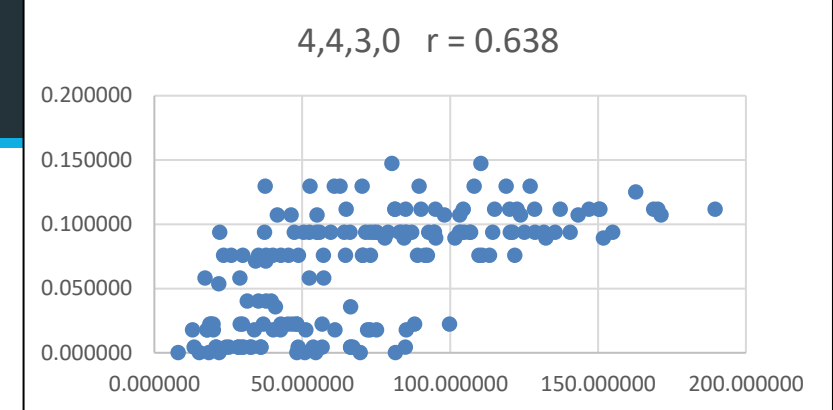
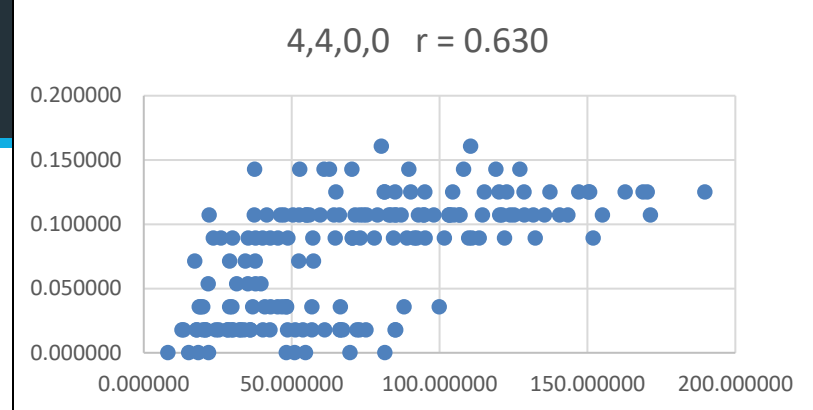
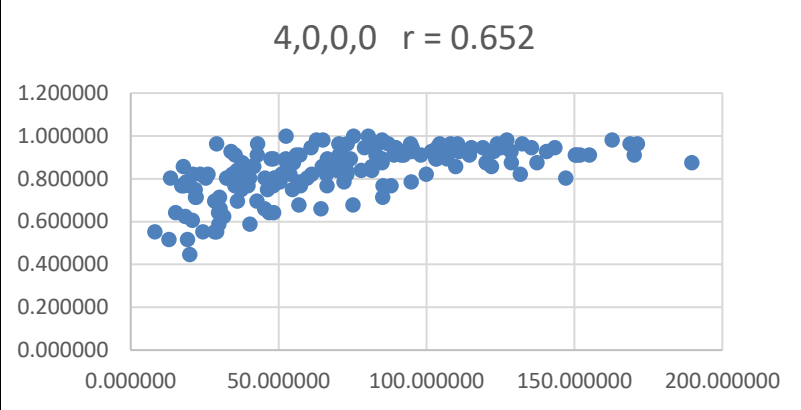
- ・ 言語距離（56語）から19地点を分類
- ・ 平方ユークリッド距離・ワード法
→ 言語距離の数値化（4, 0, 0, 0 など）が妥当かどうかを判断

■ Mantel test

- ・ 2つの行列間の相関を表す相関係数
 - > 言語距離行列 × 物理距離行列
 - > 物理距離が遠いペアほど、言語距離も大きいを見る
- ・ Spearman Mantel r
 - > $r = -1 \sim +1$
 - $r > 0$: 物理距離が遠いほど言語距離も大きい
 - $r \approx 0$: 対応関係が弱い
 - $r < 0$: 物理距離が遠いほど言語距離が小さい

04 結果





タイトル＝重み付けの数値 (A, B, C, D)、 r = Spearman Mantel r

縦軸：言語距離 (56語)、横軸：物理距離 (km)

Results

結果 : Spearman Mantel r

category	4,0,0,0	4,4,0,0	4,4,3,0	4,4,2,0	4,3,2,0	4,3,1,0	5,4,2,0	5,4,2.5,0	5,4.5,2.5,0
ALL (n = 56)	0.652	0.630	0.638	0.638	0.713	0.711	0.708	0.709	0.694
Quantity_Number (n = 3)	0.453	0.025	0.025	0.025	0.342	0.342	0.342	0.342	0.342
Human_Animal (n = 5)	0.461	N/A	N/A	N/A	0.461	0.461	0.461	0.461	0.461
Plant_Organic (n = 10)	0.285	0.507	0.507	0.507	0.490	0.490	0.496	0.496	0.497
Body (n = 13)	0.493	0.431	0.431	0.431	0.504	0.504	0.483	0.483	0.469
Verbs (n = 5)	0.600	0.500	0.564	0.564	0.646	0.627	0.627	0.630	0.618
Nature (n = 11)	0.517	0.403	0.403	0.403	0.503	0.503	0.492	0.492	0.487
Colors (n = 5)	0.455	0.560	0.560	0.560	0.566	0.566	0.566	0.566	0.566
Qualities (n = 4)	0.338	N/A	N/A	N/A	0.338	0.338	0.338	0.338	0.338

Results

結果: Spearman Mantel r

category	4,0,0,0	4,4,0,0	4,4,3,0	4,4,2,0	4,3,2,0	4,3,1,0	5,4,2,0	5,4,2.5,0	5,4.5,2.5,0
ALL (n = 56)	0.652	0.630	0.638	0.638	0.713	0.711	0.708	0.709	0.694
Quantity_Number (n = 3)	0.453	0.025	0.025	0.025	0.342	0.342	0.342	0.342	0.342
Human_Animal (n = 5)	0.461	N/A	N/A	N/A	0.461	0.461	0.461	0.461	0.461
Plant_Organic (n = 10)	0.285	0.507	0.507	0.507	0.490	0.490	0.496	0.496	0.497
Body (n = 13)	0.493	0.431	0.431	0.431	0.504	0.504	0.483	0.483	0.469
Verbs (n = 5)	0.600	0.500	0.564	0.564	0.646	0.627	0.627	0.630	0.618
Nature (n = 11)	0.517	0.403	0.403	0.403	0.503	0.503	0.492	0.492	0.487
Colors (n = 5)	0.455	0.560	0.560	0.560	0.566	0.566	0.566	0.566	0.566
Qualities (n = 4)	0.338	N/A	N/A	N/A	0.338	0.338	0.338	0.338	0.338

Results

結果: Spearman Mantel r

category	4,0,0,0	4,4,0,0	4,4,3,0	4,4,2,0	4,3,2,0	4,3,1,0	5,4,2,0	5,4,2.5,0	5,4.5,2.5,0
ALL (n = 56)	0.652	0.630	0.638	0.638	0.713	0.711	0.708	0.709	0.694
Quantity_Number (n = 3)	0.453	0.025	0.025	0.025	0.342	0.342	0.342	0.342	0.342
Human_Animal (n = 5)	0.461	N/A	N/A	N/A	0.461	0.461	0.461	0.461	0.461
Plant_Organic (n = 10)	0.285	0.507	0.507	0.507	0.490	0.490	0.496	0.496	0.497
Body (n = 13)	0.493	0.431	0.431	0.431	0.504	0.504	0.483	0.483	0.469
Verbs (n = 5)	0.600	0.500	0.564	0.564	0.646	0.627	0.627	0.630	0.618
Nature (n = 11)	0.517	0.403	0.403	0.403	0.503	0.503	0.492	0.492	0.487
Colors (n = 5)	0.455	0.560	0.560	0.560	0.566	0.566	0.566	0.566	0.566
Qualities (n = 4)	0.338	N/A	N/A	N/A	0.338	0.338	0.338	0.338	0.338

05 考察と結論

本研究の目的

段階的非類似度の一致から不一致までを最大4段階に設定した場合
各段階にどの程度の数値を与えるべきか検証する

- ・物理距離(km)を固定した状態で、重み付けを変更し、比較する
- ・山々で隔てられる、ロマンシュ語の方言で分析をする

Analysis and Conclusion

考察

■ Spearman Mantel test

- ・ etymologyの一致を中心に評価する配点（4,4,0,0 / 4,4,3,0 / 4,4,2,0）では、全ての地点で同一語源の語形となる場合には言語距離の計測が不可能
- ・ 完全一致のみを評価する配点（4,0,0,0 = バイナリ）よりも、IPA不一致とetymology部分一致も評価する配点（4,3,2,0 / 4,3,1,0 / 5,4,2,0 / 5,4,2.5,0 / 5,4.5,2.5,0）の方が、物理距離との相関が高い

■ 今後の課題

- ・ 地図データに欠損がある場合の点数化の方法

References (1/3)

- AIS = Jaberg, K & Jud, J. (1928-1940) *Sprach- und sachatlas italiens und der Südschweiz*. 8 vols. Zofingen: Ringier.
- Goebel, H. (1992) “Problèmes et méthodes de la dialectométrie actuelle (avec application à l’AIS).” In G. Aurrekoetxea & X. Videgain (eds.), *Nazioarteko dialektologia biltzarra. Agiriak / Actes du Congrès international de dialectologie* (Iker 7). Bilbo: Euskaltzaindia, 429-475.
- Greenhill, S. J. (2011) “Levenshtein Distances Fail to Identify Language Relationships Accurately.” *Computational Linguistics* 37(4): 689-698.
- Heeringa, W. & Nerbonne, J. (2001) “Dialect areas and dialect continua”. *Language Variation and Change* 13(3): 375-400.
- Kawaguchi, Y. (2007) “Is it possible to measure the distance between near languages ? A case study of French dialects”. In J.-M. Eloy & T. Ó hlfearnáin (eds.), *Langues proches – Langues collaterals*. Paris: L’Harmattan, 81-88.
- Lafkioui, M. (2008) “Dialectometry Analyses of Berber Lexis.” *Folia Orientalia* 44: 71–88.
- Nerbonne, J., Heeringa, W. & Kleiweg, P. (1999) “Edit distance and dialect proximity.” In D. Sankoff & J. B. Kruskal (eds.), *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, 2nd ed. Stanford, CA: CSLI Press, V-XVI.

References (2/3)

- Pickl, S., Spettl, A., Pröll, S., Elspaß, S., König, W. & Schmidt, V. (2014) “Linguistic Distances in Dialectometric Intensity Estimation.” *Journal of Linguistic Geography* 2(1): 25–40.
- Séguy, J. (1971) “La relation entre la distance spatiale et la distance lexicale.” *Revue de linguistique romane* 35(139-140): 335-357.
- Swadesh, M. (1955) “Towards greater accuracy in lexicostatistic dating.” *International Journal of American Linguistics* 21(2): 121-137.
- Valls, E., J., Prokić, J., Wieling, M., Clua, E. & Lloret, M.-R. (2012) “Applying the Levenshtein Distance to Catalan dialects: A brief comparison of two dialectometric approaches” *VERBA* 39: 35-61.
- Yarimizu, K., Kawaguchi, Y & Ichikawa, M. (2005) “Multi Analysis in Dialectology – A Case study of the Standardization in the Environs of Paris.” In Y. Kawaguchi, S. Zaima, T. Takagaki, K. Shibano & M. Usami (eds.), *Linguistic Informatics: State of the Art and the Future*. Amsterdam/Philadelphia: John Benjamins, 99-119.
- 清宮貴雅 (2020)「ロマンシュ語とフランコプロヴァンス語の類似性の計量方言学的分析」『ロマンス語研究』53: 123-132.
- 鎌水兼貴 (2013)「言語間距離における「重み付け」」『明海日本語』18 (2): 179-194.

References (3/3)

[非公開研究会での発表]

清宮貴雅、渡部展也、Anh Phan「地形距離と言語類似度の関係性 - ロマンシュ語を例に -」
「時空言語学」の創成：地理と歴史を融合した言語の変化と発展への新たなアプローチ
対面研究会，2026年3月19日（大阪 / ハイブリッド）。

持橋大地、渡部展也、清宮貴雅「GISによる地形コスト・交通ちゅだんの言語距離への影響（中間報告）」
「時空言語学」の創成：地理と歴史を融合した言語の変化と発展への新たなアプローチ
対面研究会，2026年6月20日（大阪 / ハイブリッド）。

ご清聴ありがとうございました。

本研究はJSPS科研費 JP24K23937, JP25K00472の助成を受けたものです。

連絡先：s.takamasa1993@gmail.com

所属：東京外国語大学 語学研究所 特別研究員 清宮貴雅