Pronoun consonant patterns: Deep inheritance or erosion effects?

Harald Hammarström
Uppsala Universitet
harald.hammarstrom@lingfil.uu.se

18 Apr 2025 Princeton

Pronouns and Deep Language Families

	1PSG	2PSG	
Indo-European	**mē*	**te*	Ruhlen (1994:16)
Nivkh	*me-*	*ti*	Ruhlen (1994:16)

Eurasiatic?

Wintu	*ni*	*mi*	Nichols and Peterson (1996:344)
Kiowa	*nɔ̀*	*àm*	Nichols and Peterson (1996:344)

Amerind?

Western Dani	*an*	*kat*	Ross (1995:156)
Awin	*no*	*gu*	Ross (1995:156)

Trans New Guinea?

Pronouns Canons

- The pronouns resemblances often boil down to single consonant (+ vowel?) correspondendences for 1PSG and 2PSG
- Ambitious proponents argue for such macro-families as
  - **Eurasiatic** (m-/t-) (Greenberg 1997, 2002; Ruhlen 1994)
  - A (wide-range) **Trans New Guinea** (na-/ka-) (Ross 1995)
  - **Amerind** (nV-/mV-) (Nichols and Peterson 1996)
  - ... and others in Africa (Babaev 2009), Australia (Blake 1991; Harvey 2003) and New Guinea (Ross 1995, 2001, 2005; Voorhoeve 1987; Wurm 1971, 1975) etc.

Argument

"Getting down to brass tacks, how in the Hell are you going to explain general American n- 'I' except genetically? It's disturbing, I know, but (more) non-committal conservatism is only dodging, after all, isn't it?" Sapir (1918 in a letter; quoted in Darnell and Sherzer 1971: 27)

- Not chance
- Not borrowing
- Not sound symbolism
- Not some 'mysterious' contact influence (Campbell 1994; Güldemann 2017; Nichols 2012; Osada 2001)
- ... then it must be genealogical inheritance

Another Possibility: Erosion

- Suppose a meaning is very stable = seldom undergoes form replacement
- The the form is subject to erosion, i.e., various phonetic lenition processes
- Ultimately such erosion processes may lead to a monosyllable with “unmarked” sounds
- These monosyllables might resemble each other more than expected from just any two (uneroded) forms
- ... if so, short pronoun resemblances across families may reflect convergent erosive evolution, **not** deep families

What is Erosion?

- The term ‘erosion’ (or, rather, its German counterpart) in linguistics goes back at least to 1822

  Dies wird wohl erleichtert durch verloren gehende Bedeutung der Elemente, und **Abschleifung** *der Laute in langem Gebrauch. (von Humboldt 1822: 306, quoted in Lindström 2004:205)*

- Used — without a stringent definition — to refer to (sound) changes that reduce the phonetic form of a word
- Equated by many modern authors with successive *lenition*

What is Lenition?

- First used by Thurneysen 1909 as

  a mutation of consonants which normally originated in a reduction of the energy employed in their articulation (Thurneysen 1909:74)

- Used — without a stringent definition — to refer to sound changes that result in "a 'relaxation' or 'weakening' of articulatory effort" (Bauer 1988, 2008; Gurevich 2011, inter alia)
- I have not found an exhaustive list of which sound changes count as lenition according to any author
- There is a list of "processes that most commonly fall under the label of *lenition*" (Gurevich 2011)

Erosion, Lenition or What?

- For the purpose of the present talk
  - It does not matter exactly what erosion / lenition is
  - Enough to assume a process which *over time reduces the entropy of the distribution of forms not being replaced*, i.e., that the space of forms is compressed
- Nevertheless, I prefer to investigate a process which approaches reality i.e., lenition
- I also will not address the question of *why* erosion / lenition happens

Lenition according to Gurevich (2011)

Lenition Type	Example
Degemination	kk → k / V _ V
Deaspiration	p^h → p
Voicing	p → b / V _ V
Spirantization	b → ß / V _ V
Flapping	r → ɾ / V _ V
Debuccalization	p → ʔ / _ C
Gliding	b → w / V _
Loss	h → Ø
Devoicing	b → p / _$

Lenition Augmented and Formalized by HH

Lenition Type	Definition
1. Degemination	X: → X
2. Deaspiration [and deejectivization]	C^h, C’ → C
3. [Intervocalic] Voicing	[P̥F̥] → [P̬F̬] / V _ V
4. [Intervocalic] Spirantization	[P̬F̬] → ɣ, β, ð
5. [Intervocalic] Flapping	ʈ, ɖ, r → ɾ / V _ V
6. [Post-Vocalic] Debuccalization	P̥ → ʔ, S̥ → h / V _
7. [Post-Vocalic] Gliding	t, c → j, p, k → w / V _
8. [Postvocalic] Loss [of Glottal and Glide]	h, ʔ, w, y → Ø / V _
9. [Final] Devoicing	C̬ → C̥ / _$
10. Click Loss	$!^C$ → C
11. Vowel Denasalisation	Ṽ → V
12. Nasalisation	VN → Ṽ
13. Accompaniment Loss	$[PN]^{wj}$ → [PN]
14. Liquid/Nasal + C Cluster Simplification	[LN]P → PP, [LN]N → NN
15. [Intervocalic Spirant to Glide]	ɣ, β, ð → w, w, j / V _ V
16. [Vowel merger]	VV → V̄
17. [Intervocalic / final liquid lenition]	V^BL → w, V^FL → j, r → ɣ / V _ [V$]
18. [Initial glottal loss]	h, ʔ → Ø / $ _
19. [Affricatization]	P̪S → ʧ, ʤ

S = Sibilant, P = Plosive, F = Fricative, ! = Click, N = Nasal, L = Liquid, R = Rhotic, V̄ = L-H B-F V, V^B = Back V, V^F = Front V

Data: ASJP

ASJP Database v 20 (2022) `https://asjp.clld.org/`

# Wordlists	10 168
# ISO 639-3 languages	5 676
# Meanings per language	40
Total # words	482 117

Transcription impoverished (compared to IPA) with only 7 vowels and 34 consonants, no tone

Example: ASJP Swedish

https://asjp.clld.org/languages/SWEDISH

ASJP | Home | Wordlists | Meanings | Sources

Wordlist Swedish

Glottocode: swed1254 | ISO 639-3: swe

Compiled by Viveka Velupillai

Showing 1 to 40 of 40 entries

← Previous | 1 | Next →

No.	Meaning	Concepticon	Word	Loan
1	I	I	yog	False
2	you	THOU	du	False
3	we	WE	vi	False
11	one	ONE	et	False
12	two	TWO	tv~o	False
18	person	PERSON	mEniSxE	False
19	fish	FISH	fisk	False
21	dog	DOG	h3nd~	False
22	louse	LOUSE	lus	False
23	tree	TREE	trEd	False
25	leaf	LEAF	lev	False
28	skin	SKIN	Sx~in	False
30	blood	BLOOD	bl~ud	False
31	bone	BONE	ben	False
34	horn	HORN (ANATOMY)	hun	False
39	ear	EAR	3rE	False

Coordinates WGS84	60°N, 15°E 60.00, 15.00
number of speakers	9,606,320
status	alive

Classification

WALS
IE > Germanic

Glottolog
Indo European > Germanic > Northwestgermanic > Northgermanic > Northscandinavian > East Centralswedic > Eastswedic

Ethnologue
Indo European > Germanic > North > Eastscandinavian > Danish Swedish > Swedish

Example: Erosion

mEniSxE:	mEniSxE → mEnihxE → mE*ihxE → mEi-hxE → mEhxE → mExE → mEVE → mEwE → mEE → mE
fisk:	fisk → fihk → fik → fi7 → fi
du:	du

- A natural measure of (non-)erosion $E(X)$ is # steps towards saturation
  - E(mEniSxE) = 9
  - E(fisk) = 4
  - E(du) = 0

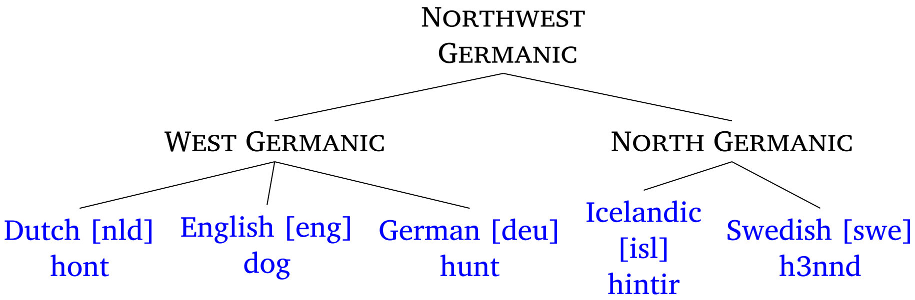
Erosion: ASJP Global Statistics

Total # words	482 117
Total # different forms	187 005
Total # forms after erosion	8 855
Avg erosion $E(x)$	6.1775

Stability

- Ever so often the form for a given meaning is replaced by a (new, non-cognate) form
- We would like to know the (approximate) age of each word form
- We have
  - Tree topologies (from glottolog.org)
  - Word forms at the leaves
- Would like to have reliable cognate judgments over all forms, but this is not globally available

Gauging Stability: Approach

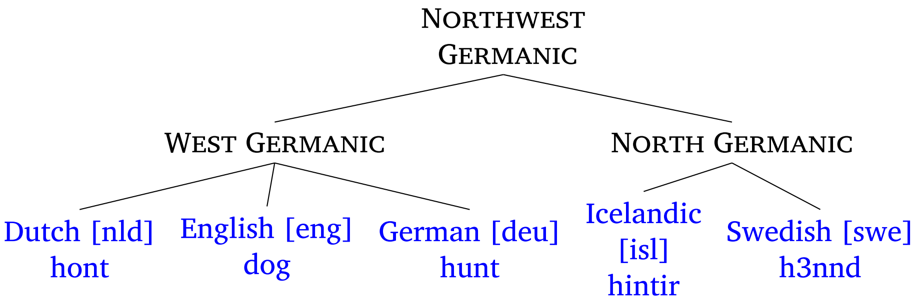


- If we had reliable cognate judgments, it would be easy to argue the **hund* form is older, and the **dog*-form an innovation
- Can we estimate the same thing from comparing form similarity and be correct often enough?
- We will need some simplifying assumptions ...

Form Similarity

- Long tradition of previous work on cognate detection (e.g., List 2014, Kondrak 2009, Steiner et al. 2011) use some variant of Edit Distance
- Essentially, the smallest number of substitutions, deletions and insertions to get from one string to the other
$$\begin{aligned} D(hund, hond) &= 1 \\ D(hund, dog) &= 4 \end{aligned}$$
- Can be informed by common sound correspondences
- The distance can be normalized to the length of the longer string (yields a score 0..1)
- Can be turned into a similarity score $S(x, y) = 1 - D(x, y)$
- Can be turned into a probability of cognacy
$PC(x, y) = \frac{|z \in Y | D(x,z) > D(x,y)|}{|\{z | z \in Y\}|}$ if Y is the set of forms of y:s language

Gauging Stability



- $P(form, node)$ = Probability that a cognate of *form* was present at node
- Then compare, e.g., $P(x, root)$ for each x found at the leaves
- E.g., we expect to obtain

  $P(hund, \texttt{Northwest Germanic}) > P(dog, \texttt{Northwest Germanic})$

Three Assumptions

- Form similarity to cognacy: Cognates are more similar in form than non-cognates

$$PC(x,y) > PC(x,z) \Longleftrightarrow x,y \text{ are cognate but y, z are not}$$

- Unique cognate appearance: Cognates can only appear once, no loans, chance identity unlikely
- Multiple occurrence is the dominant evidence for cognate retention

Quite far-reaching assumptions, hopefully counterbalanced by the size of the data

Cognate Evolution Formally (cf. Rönchen et al. 2024)

- $P(x, p)$ = Probability that a cognate of form x was present at p
- If we know at least one branch has the form x, let
  - c_x denote the child of p under which x is ultimately found and
  - $c_1, \ldots, c_n$ the other children of p
  - r_i denote the probability of a retention along the branch to c_i (constant $\Leftarrow$ equal branch lengths)
- If p is a leaf node with form y then $P(x, p) = PC(x, y)$ ($\Leftarrow$ form similarity to cognacy)
- Otherwise $P(x, p) =$
  $P(x$ was retained at c_x but lost in all of $c_i)$ +
  $P(x$ was retained at c_x and at least one more child $c_i)$ ($\Leftarrow$ unique)

$$P(x, p) = P(x, c_x) \cdot r_x \prod_i (1 - r_i) \cdot (1 - P(x, c_i)) +$$

$$P(x, c_x) \cdot (1 - \prod_i (1 - P(x, c_i)))$$

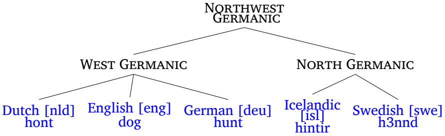
Gauging Cognate Age

- The expression

$$P(x,p) = P(x,c_x) \cdot r_x \prod_i (1-r_i) \cdot (1-P(x,c_i)) +$$
$$P(x,c_x) \cdot (1 - \prod_i (1-P(x,c_i)))$$

- crucially depends on the probability $\prod_i (1-P(x,c_i))$ ("x is absent all the other branches") independent of branch lengths/r_i
- If x is absent all the other branches, c_x may nevertheless have been at the parent, the likelihood of this ($r_x \prod_i (1-r_i)$) depends on the number of siblings and branch lengths/r_i
- Let us assume the probability of retention is dominated by the cases where the cognate is retained in more than one branch ($\Leftarrow$ multiple occurrence is the dominant evidence for cognate retention)

Example: Simplified Germanic



PC(x, y)

	dog	h3nnd	hintir	hont
dog	1.00	0.23	0.22	0.42
h3nnd	0.23	1.00	0.59	0.66
hintir	0.22	0.59	1.00	0.58
hont	0.42	0.66	0.58	1.00
hunt	0.35	0.66	0.58	0.94

- $P(dog, \texttt{West Germanic}) = 0.623 + 0.377 \cdot r_2 \cdot (1 - r_1) \cdot (1 - r_3)$
- $P(hont, \texttt{West Germanic}) = 0.963 + 0.037 \cdot r_1 \cdot (1 - r_2) \cdot (1 - r_3)$
- $P(hont, \texttt{Northwest Germanic}) \gtrsim 0.58$
- $P(dog, \texttt{Northwest Germanic}) \gtrsim 0.15$
- $P(hunt, \texttt{Northwest Germanic}) \gtrsim 0.58$
- $P(hintir, \texttt{Northwest Germanic}) \gtrsim 0.49$
- $P(h3nnd, \texttt{Northwest Germanic}) \gtrsim 0.52$

Example: Dog in Indo-European

- Computed on 301 Indo-European lgs in ASJP

Rank	x	$P(x, \texttt{Indo-European})$	Language(s)
1	Sun	0.49	Eastern Armenian [hye], Western Armenian [hyw]
2	kanis	0.28	Latin [lat]
3	qen	0.27	Gheg Albanian [aln]
4	ki	0.26	Breton [bre], Cornish [cor], Welsh [cym]
5	sunis	0.25	Old Prussian [prg]
6	TEn	0.22	Northern Tosk Albanian [als]
7	hunds	0.21	Gothic [got]
8	Su3	0.21	Lithuanian [lit]
9	gue	0.18	Tsakonian [tsd]
10	ku	0.15	Irish [gle], Early Irish [sga], Tokharian A [xto], Tokharian B [txb]
11	sax	0.11	Bashkardi [bsg]
12	koyni	0.10	Megleno Romanian [ruq]
13	span	0.10	Avestan [ave]
14	SoC	0.10	Wakhi [wbl]
15	kod	0.10	Shughni [sgh], Yazgulyam [yah]
...	...	...	...
159	dog	0.00	English [eng], Scots [sco], Bislama [bis], Sea Island Creole English [gul], Ghanaian Pidgin English [gpe], Hawai'i Creole English [hwc], Cameroon Pidgin [wes], Krio [kri], Kriol [rop], Nigerian Pidgin [pcm], Pichi [fpe], Torres Strait Creole [tcs]
...	...	...	...
168	pero	0.00	Spanish [spa]
...	...	...	...
184	SyE	0.00	French [fra]
...	...	...	...
187	iru	0.00	Chavacano [cbk]

Expected Age

- For a form x we obtain a series of probabilities

$$P(x, leaf) \rightarrow P(x, parent_1) \rightarrow P(x, parent_2) \rightarrow \ldots \rightarrow P(x, root)$$

English [eng]	Macro-English	...	N Sea Germanic	W Germanic	NW Germanic	Germanic	Indo-European
P(dog, n) 1.00	0.78	...	0.04	0.01	0.00	0.00	0.00

Dutch [nld]	Global Dutch	Modern Dutch	W Germanic	NW Germanic	Germanic	Indo-European
P(hont, n) 1.00	1.00	0.69	0.57	0.35	0.25	0.08

- Assuming equal length steps between successive parents we can convert this to an "expected age" $A(x)$

$$A(x) = \sum_{p_i} i \cdot (P(x, p_i) - P(x, p_{i-1}))$$

- $A(dog) \gtrsim 3.02$, i.e., just before the Anglic node
- $A(hont) \gtrsim 3.95$, i.e., just after Northwest Germanic

Meaning Stability in Indo-European

- Stability $S(m)$ of meaning m = Average expected age of all forms for meaning m

#	m	$S(m)$	#	m	$S(m)$	#	m	$S(m)$	#	m	$S(m)$
1	name	4.10	11	night	3.77	21	star	3.53	31	louse	3.16
2	ear	4.09	12	blood	3.75	22	die	3.50	32	fire	3.15
3	nose	4.07	13	horn	3.73	23	skin	3.43	33	eye	3.11
4	you	4.06	14	dog	3.68	24	drink	3.39	34	breast	3.06
5	two	4.05	15	leaf	3.65	25	bone	3.36	35	tree	3.06
6	tooth	4.02	16	sun	3.64	26	stone	3.29	36	hear	3.04
7	hand	3.99	17	full	3.63	27	knee	3.28	37	person	2.95
8	tongue	3.91	18	new	3.58	28	come	3.26	38	mountain	2.92
9	water	3.88	19	one	3.55	29	see	3.20	39	liver	2.91
10	fish	3.85	20	we	3.54	30	I	3.19	40	path	2.81

- Pros & cons with respect to other extant measures of stability (e.g., Starostin 2007:825, Kruskal et al. 1971, 1973; Pagel et al. 2013) which, however, all require explicit cognate coding

Pronoun Stability Across Families

Family	# lgs	I		you		we	
		Rank	$S(m)$	Rank	$S(m)$	Rank	$S(m)$
Austronesian	1164.67	20	3.52	35	3.09	30	3.26
Atlantic-Congo	902.00	6	3.77	11	3.71	32	3.29
Sino-Tibetan	534.33	2	3.85	4	3.33	33	2.50
Indo-European	510.33	30	3.19	4	4.06	20	3.54
Afro-Asiatic	306.67	8	3.15	27	2.78	29	2.74
Nuclear Trans New Guinea	305.67	1	3.61	2	3.60	7	3.29
Austroasiatic	156.00	33	2.50	21	2.80	40	2.08
Tai-Kadai	132.67	22	2.82	19	2.96	31	2.53
Pama-Nyungan	117.33	5	2.93	3	2.97	14	2.76
Otomanguean	102.33	5	3.07	24	2.81	36	2.53
Mande	73.33	6	3.10	9	3.02	36	2.39
Uto-Aztecan	72.00	4	3.18	33	2.70	29	2.75
Nuclear Torricelli	62.00	7	2.54	9	2.46	14	2.39
Dravidian	59.67	18	2.58	5	2.86	24	2.39
Uralic	59.33	13	2.77	21	2.64	16	2.69
Arawakan	55.67	4	2.80	6	2.71	17	2.38
Tupian	53.33	16	2.78	20	2.71	19	2.74
Central Sudanic	51.33	3	3.87	30	2.63	21	3.00
Nilotic	50.00	6	3.11	10	2.99	21	2.80
Athabaskan-Eyak-Tlingit	37.00	2	3.32	5	3.03	23	2.60

Are Pronouns More Stable?

- Across the 20 families on the previous slide

	I	you	we
Avg Age	3.12	2.99	2.73
Size-Weighted Avg Age	3.41	3.29	3.02
Avg Rank	10.6	14.9	24.6
Size-Weighted Avg Rank	12.8	16.9	27.3
Median Rank	6	11	24

- On the 40-item list, personal pronouns `I` and `you` are on the top half and `we` on the bottom half
- NB: The 40 meanings of ASJP have been selected precisely for stability, so if we consider 100-, 200- or 1000- meaning lists, personal pronouns should be in the very top percentiles

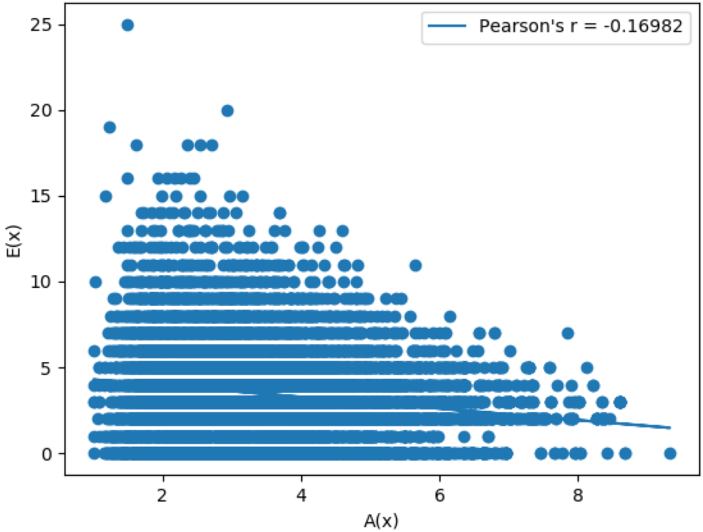
Erosion vs Stability

- Each form x has an expected age $A(x)$ and erosion potential $E(x)$

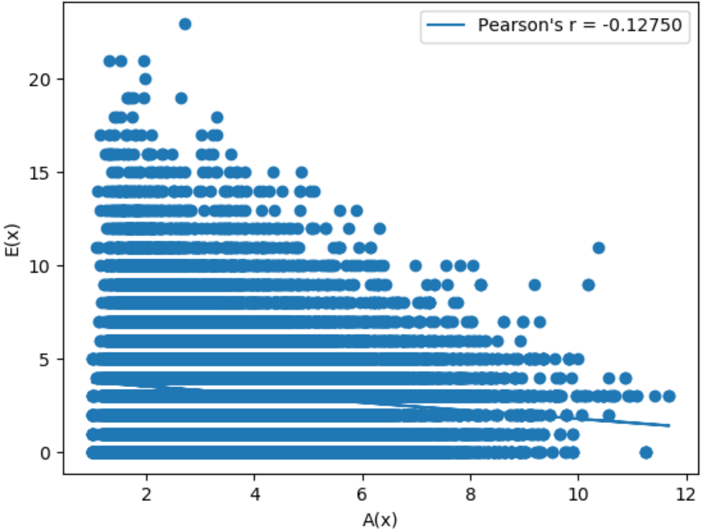
	$A(x)$	$E(x)$	
$A(dog)$	3.02	3	dog $\rightarrow$ dok $\rightarrow$ do7 $\rightarrow$ do
$A(hont)$	3.95	5	hont $\rightarrow$ hott $\rightarrow$ hot $\rightarrow$ ho7 $\rightarrow$ ho $\rightarrow$ o

- A meaning also has age and erosion as the average of those of its forms
- Ages and erosions of forms and meanings can (at least) be compared within the same family/tree
- The hypothesis the older the form, the more the erosion, i.e., increased $A(x)$ yields lower $E(x)$

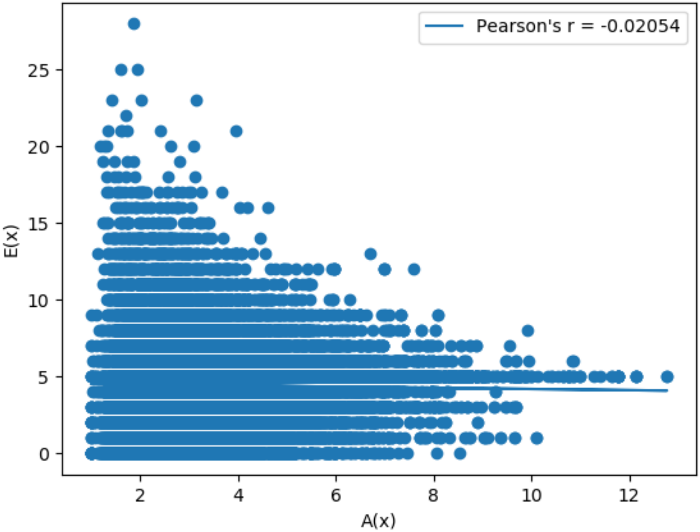
Example: All forms in Indo-European



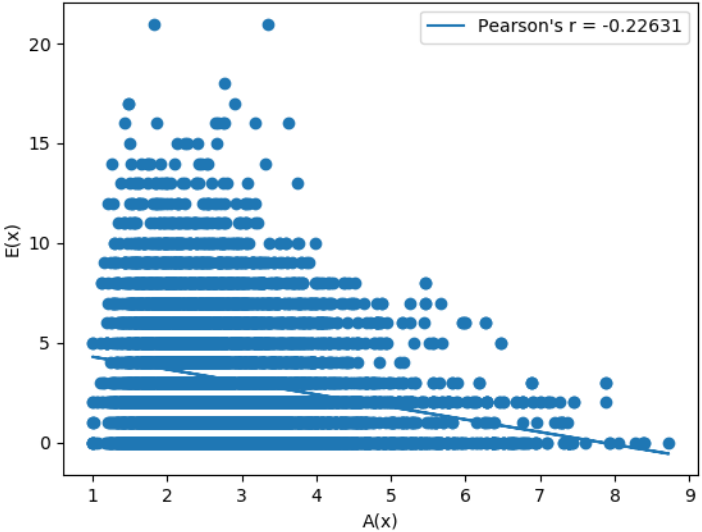
Example: All forms in Atlantic-Congo



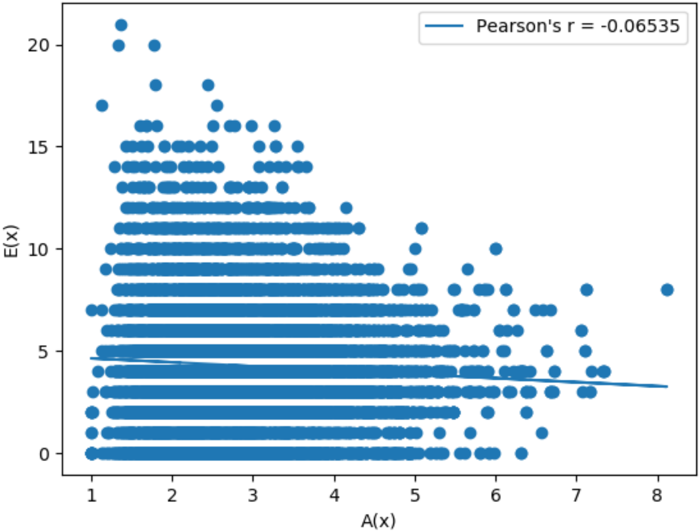
Example: All forms in Austronesian



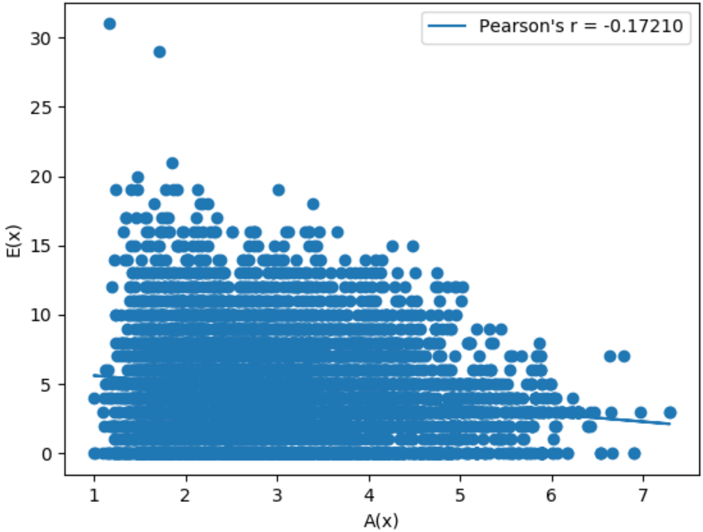
Example: All forms in Sino-Tibetan



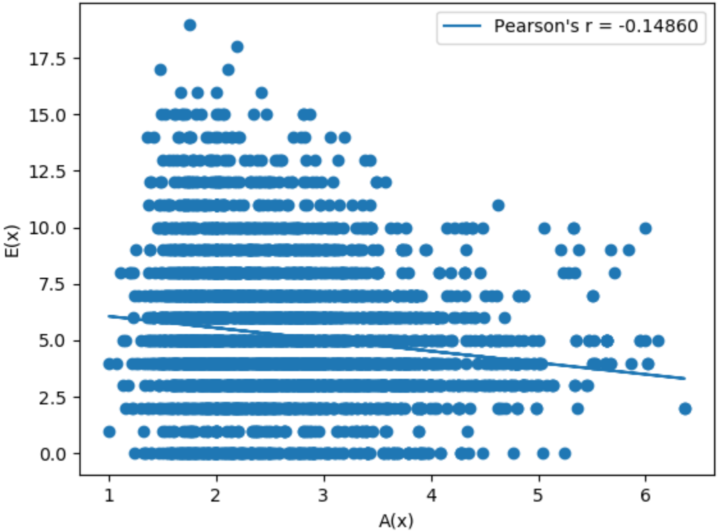
Example: All forms in Afro-Asiatic



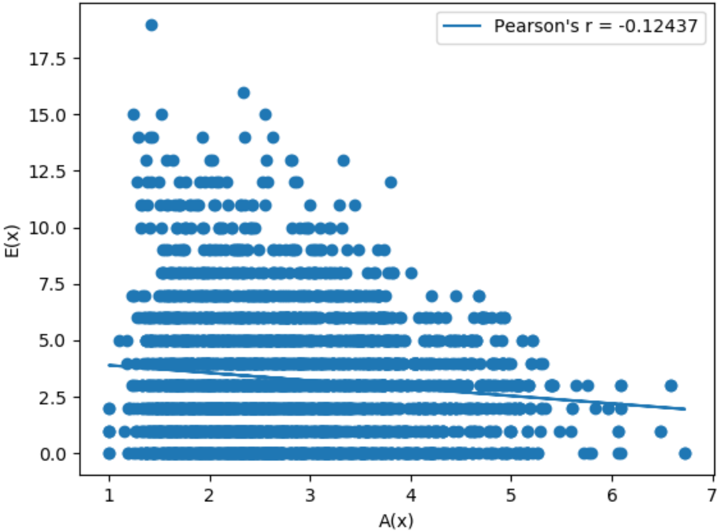
Example: All forms in Nuclear Trans New Guinea



Example: All forms in Pama-Nyungan

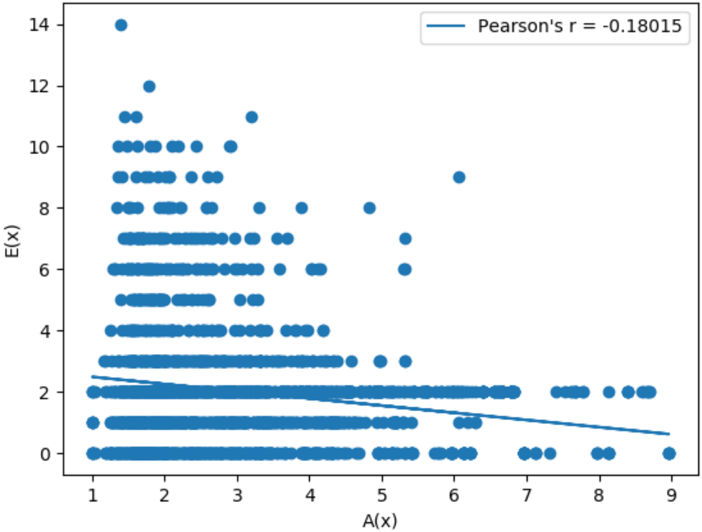


Example: All forms in Otomanguean

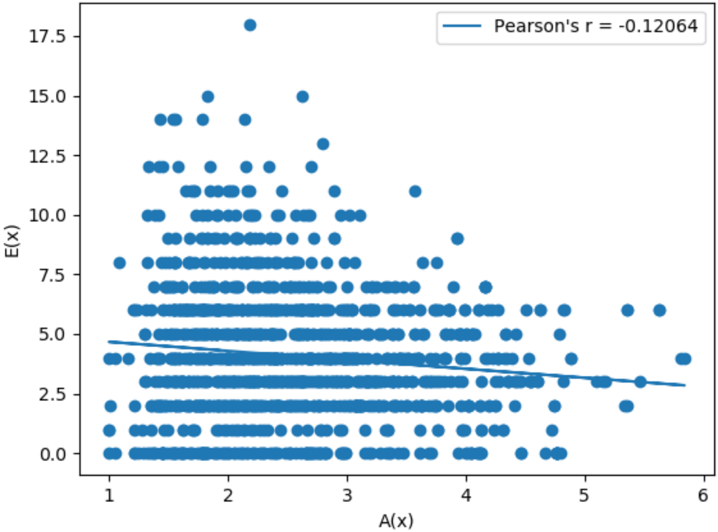


Example: All forms in Austroasiatic

Example: All forms in Tai-Kadai



Example: All forms in Dravidian



Correlations Age vs Erosion

	Pearson's r	Significance
Atlantic-Congo	-0.13	$p < 0.01$
Austronesian	-0.02	$p < 0.01$
Indo-European	-0.17	$p < 0.01$
Sino-Tibetan	-0.23	$p < 0.01$
Afro-Asiatic	-0.07	$p < 0.01$
Nuclear Trans New Guinea	-0.17	$p < 0.01$
Pama-Nyungan	-0.15	$p < 0.01$
Otomanguean	-0.12	$p < 0.01$
Austroasiatic	-0.08	$p < 0.01$
Tai-Kadai	-0.18	$p < 0.01$
Dravidian	-0.12	$p < 0.01$
Arawakan	-0.03	$p < 0.01$
Mande	-0.21	$p < 0.01$
Tupian	-0.09	$p < 0.01$
Uto-Aztecan	-0.08	$p < 0.01$
Central Sudanic	-0.12	$p < 0.01$
Nuclear Torricelli	-0.10	$p < 0.01$
Nilotic	-0.19	$p < 0.01$
Uralic	-0.10	$p < 0.01$
Athabaskan-Eyak-Tlingit	-0.17	$p < 0.01$

- Significance testing by doing 100 runs of random permutations of leaf forms to check how often r (randomized) $\leq r$ (real world)

Are Eroded Forms More Similar?

$E(w1)$	$E(w2)$	Avg $P(w1,w2)$
0	0	0.39
1	1	0.29
2	2	0.34
3	3	0.32
...	...	...
0	2	0.24
...	...	...
0	18	0.16

- Yes, but only slightly, ...

Conclusions

- Some evidence that indeed
  - `I`, `you` are in the top stability range
  - Older forms show (a little) more erosion
- This should indeed lead to pronouns being more similar than random across families (without implying they derive from a common protoform)
- But this should also hold, and slightly more so, for other very stable meanings `name`, `ear`, `nose`, `two`, ...
- Future work
  - Evaluation of age estimates on known datasets
  - Are some meanings more erosive than others?
  - Better understanding of many far reaching assumptions and averages
  - Quantify how much more similar on average two random forms become as time passes

Babaev, K. (2009). Once again on the comparison of personal pronouns in proto-languages. *Journal of Language Relationship*, 1:37–48.

Bauer, L. (1988). What is lenition? *Journal of Linguistics*, 24(2):381–392.

Bauer, L. (2008). Lenition revisited. *Journal of Linguistics*, 44:605–624.

Blake, B. J. (1991). The significance of pronouns in the history of australian languages. In Baldi, P., editor, *Patterns of Change, Change of Patterns: Linguistic change and reconstruction methodology*, volume 45 of *Trends in Linguistics: Studies and Monographs*, pages 435–450. Berlin: Mouton de Gruyter.

Campbell, L. (1994). Putting pronouns in proper perspective in proposals of remote relationships among native american languages. In *Proceedings of the meeting of the Society for the Study of the Indigenous Languages of the Americas and the Hokan-Penutian workshop*, volume 8 of *Survey of California and other Indian Languages*, pages 1–20. Berkeley: University of California.

Greenberg, J. H. (1997). The indo-european first and second person pronouns in the perspective of eurasiatic, especially chukotkan. *Anthropological Linguistics*, 39(2):187–195.

Greenberg, J. H. (2000, 2002). *Indo-European and its closest relatives: the Eurasiatic language family*. Stanford: Stanford University Press. 2 vols.

Gurevich, N. (2011). Lenition. In van Oostendorp, M., Ewen, C. J., Hume, E., and Rice, K., editors, *Phonological Processes*, volume III of *The Blackwell Companion to Phonology*, pages 1–17. Oxford: Blackwell.

Güldemann, T. (2017). A shared pronominal canon in the macro-sudan belt: typological, areal and genealogical perspectives. In Kramer, R. and Kießling, R., editors, *Mechthildian approaches to Afrikanistik: advances in language based research in Africa, Festschrift für Mechthild Reh*, pages 101–146. Köln: Rüdiger Köppe Verlag.

Harvey, M. (2003). Reconstruction of pronominals among the non-pama-nyungan languages. In Evans, N., editor, *The non-Pama-Nyungan languages of northern Australia: Comparative Studies of the continent's most linguistically complex region*, volume 552 of *Pacific Linguistics*, pages 475–513. Canberra: Research School of Pacific and Asian Studies, Australian National University, Canberra.

Kondrak, G. (2009). Identification of cognates and recurrent sound correspondences in word lists. *Traitement Automatique des Langues*, 50(2):201–235.

Kruskal, J. B., Dyen, I., and Black, P. (1971). The vocabulary method of reconstructing language trees: innovations and large-scale applications. In Hodson, F. R., Kendall, D. G., and Tautu, P., editors, *Mathematics in the Archaeological and Historical Sciences*, pages 361–380. Edinburgh: Edinburgh University Press.

Kruskal, J. B., Dyen, I., and Black, P. (1973). Some results from the vocabulary method of reconstructing language trees. In Dyen, I., editor, *Lexicostatistics in Genetic Linguistics*, pages 30–55. Berlin: Mouton.

Lindström, T. (2004). *The History of the Concept of Grammaticalisation*. PhD thesis, University of Sheffield.

List, J.-M. (2014). *Sequence comparison in historical linguistics*. PhD thesis, Düsseldorf: Heinrich Heine University.

Nichols, J. (2012). Selection for m:t pronominals in eurasia. In Lars, J. and Robbeets, M., editors, *Copies versus Cognates in Bound Morphology*, volume 215 of *Current issues in linguistic theory*, pages 47–69. Amsterdam: John Benjamins.

Nichols, J. and Peterson, D. A. (1996). The amerind personal pronouns. *Language*, 72(2):336–371.

Osada, T. (2001). Personal pronouns and related phenomena in the south asian linguistic area: Convergent features or convergence-resisting features? In Singh, R., Bhaskararao, P., and Subbarao, K., editors, *The yearbook of South Asian languages and linguistics 2001. Tokyo symposium on South Asian languages: Contact, convergence and typology*, pages 269–287. New Delhi: Sage.

Pagel, M., Atkinson, Q. D., Calude, A., and Meade, A. (2013). Ultraconserved words point to deep language ancestry across eurasia. *Proceedings of the National Academy of Sciences of the United States of America*, 110(21):8471–8476.

Ross, M. (1995). The great papuan pronoun hunt: Recalibrating our sights. In Baak, C., Bakker, M., and van der Meij, D., editors, *Tales from a concave world: Liber amicorum Bert Voorhoeve*, pages

139–168. Leiden: Department of Languages and Cultures of Southeast Asia and Oceania, Leiden University.

Ross, M. (2001). Is there an east papuan phylum? evidence from pronouns. In Pawley, A., Ross, M., and Tryon, D., editors, *The Boy from Bundaberg: Studies in Melanesian Linguistics in Honour of Tom Dutton*, volume 514 of *Pacific Linguistics*, pages 301–321. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Ross, M. D. (2005). Pronouns as a preliminary diagnostic for grouping papuan languages. In Pawley, A., Attenborough, R., Golson, J., and Hide, R., editors, *Papuan Pasts: Studies in the Cultural, Linguistic and Biological History of the Papuan-speaking Peoples*, volume 572 of *Pacific Linguistics*, pages 15–66. Canberra: Research School of Pacific and Asian Studies, Australian National University, Canberra.

Ruhlen, M. (1994). *On the Origin of Languages: Studies in Linguistic Taxonomy*. Stanford: Stanford University Press.

Rönchen, P., Wiklund, T., and Hammarström, H. (2024). Likelihood calculation in a multistate model of vocabulary evolution for linguistic dating. *Language Dynamics and Change*, 14:1–41.

Starostin, S. (2007). Opredelenije ustojčivosti bazisnoj leksiki. In *Trudy po jazykoznaniju*, pages 825–839. Moscow: Jazyki slavjanskix kul'tur.

Steiner, L., Stadler, P. F., and Cysouw, M. (2011). A pipeline for computational historical linguistics. *Language Dynamics & Change*, 1:89–127.

Thurneysen, R. (1998 [1909]). *A Grammar of Old Irish*. Dublin Institute for Advances Studies: School of Celtic Studies, Dublin Institute for Advances Studies.

Voorhoeve, C. L. (1987). Worming one's way through new guinea: the chase for the peripatetic pronouns. In Laycock, D. C. and Winter, W., editors, *A World of Language: Papers presented to Professor Stephen A. Wurm on his 65th Birthday*, volume 100 of *Pacific Linguistics: Series C*, pages 709–727. Canberra: Research School of Pacific and Asian Studies, Australian National University, Canberra.

Wurm, S. A. (1971). The papuan linguistic situation. In Sebeok, T. A., editor, *Linguistics in Oceania*, volume 8 of *Current Trends in Linguistics*, pages 541–657. Berlin: Mouton de Gruyter, Berlin.

Wurm, S. A. (1975). Personal pronouns. In Wurm, S. A., editor, *New Guinea Area Languages and Language Study Vol 1: Papuan Languages and the New Guinea linguistic scene*, volume 38 of *Pacific Linguistics: Series C*, pages 191–218. Canberra: Research School of Pacific and Asian Studies, Australian National University.