

Some Aspects of Language are More Stable than Others

– A Comparison of Eight Methods –



dan.dediu@mpi.nl

Dan Dediu

* Language and Genetics Department
MPI for Psycholinguistics, Nijmegen
* Donders Institute for Brain, Cognition and Behaviour,
Radboud University Nijmegen

cysouw@uni-marburg.de

Michael Cysouw

* Research Unit Quantitative Language Comparison
Ludwig Maximilians Universität München
* Forschungszentrum Deutscher Sprachatlas
Philipps-Universität Marburg



Typological features could carry information about historical relationships between languages that are too deep (or too blurred by contact) for the historical-comparative method to identify. For this reason, it would be highly important to know about typological features that do not easily change, i.e. that are stable. However, it is not even clear whether the notion of stability applies to typological features. Maybe any attempt to estimate stability will be fragile, and estimates will differ between methods.

Goal

Investigate whether structural/typological aspects of language have an inherent stability, either high or low.

Approach

We compare various published methods proposing different operationalizations and estimates of the concept of structural stability, using the same large dataset of language families and features (the *World Atlas of Language Structures*, Haspelmath et al. 2005). To be able to compare all approaches, we had to work with 60 WALS features.

Methods Compared

- CM** Cysouw et al. (2008) using Mantel-statistic
- CC** Coherence method from Cysouw et al. (2008)
- CR** Rank method from Cysouw et al. (2008)
- D** Approach from Dediu (2011)
- M** Approach from Maslova (2004)
- P1** Parkvall (2008) using all families from WALS
- P2** Same as P1, using only 'widely accepted' families
- W** Approach from Wichmann & Holman (2009)

Results

Overall, the different methods show surprising agreement, with CC, CR, D and M being particularly close. After removing four bivariate outliers, a Principle Component Analysis shows more than 50% agreement between all methods, and more than 70% agreement between CC, CR, D, and M. Based on this, we propose a list of stable features from WALS.

Summary of Methods

Cysouw et al. (2008) compare each feature to the global 'average' of all features in WALS, and assign a value to how well each feature matches the global. They investigate three different operationalizations of this principle.

Dediu (2011) defines stability in terms of the rate of change experienced by a feature during the evolution of a language family. The method uses two Bayesian phylogenetic software packages, two data codings and two sources of historical linguistic classifications in order to control for spurious results.

Maslova (2004) uses pairs of closely related languages to derive an estimate of rate of change. The method works by comparing the number of consistent pairs (i.e. pairs of closely related languages that have the same type) with the number of differing pairs.

Parkvall (2008) uses the Herfindahl-Hirschman index (or Gini coefficient) over genealogical and areal units to estimate variability within units. Low variability is claimed to relate to high stability.

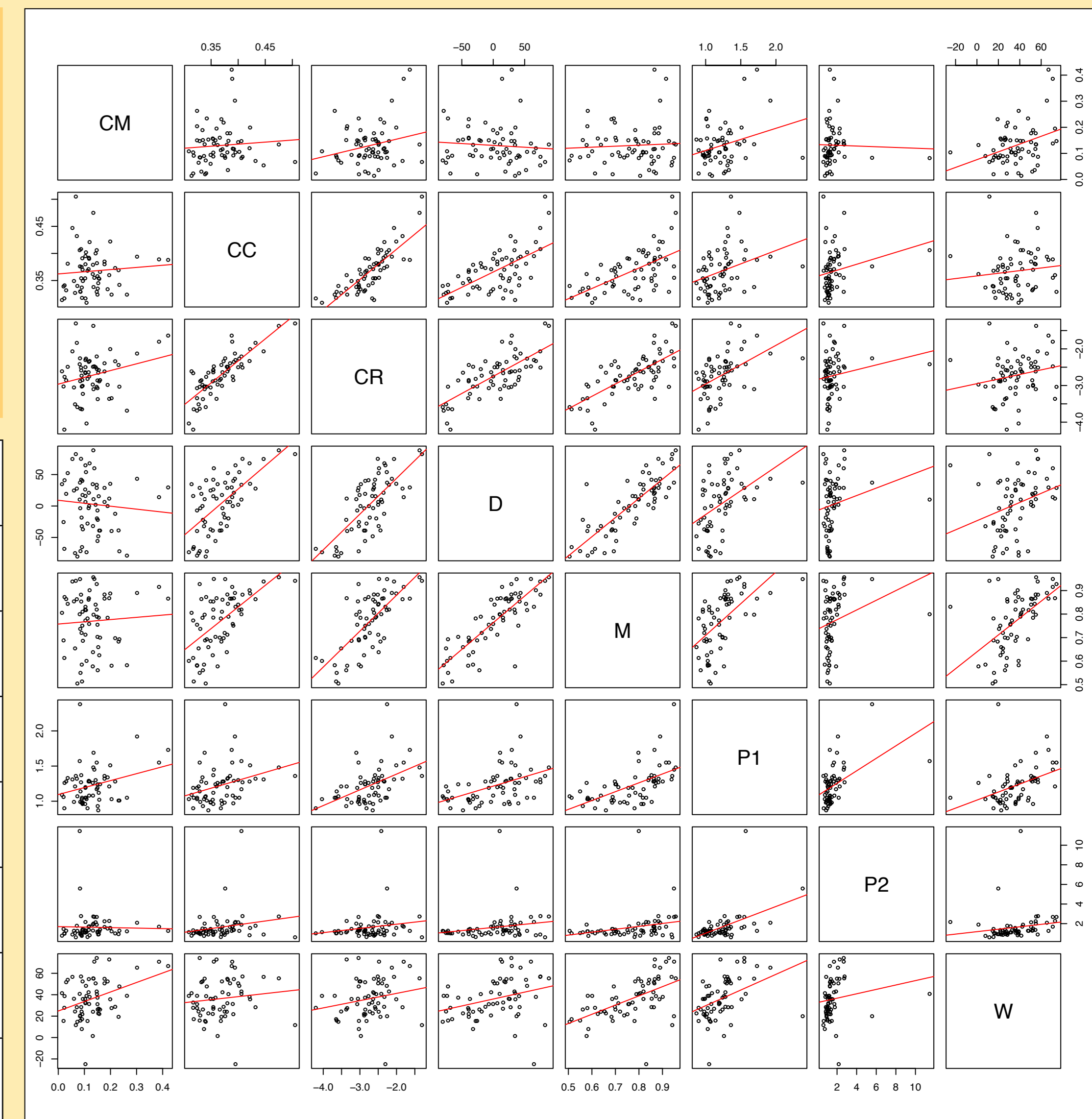
Wichmann & Holman (2009) estimate a feature's resistance to change irrespective of the causes of change. They propose three methods to estimate the relative stabilities, but their favourite is "metric C", which is based on the idea that related languages preferably share the value of a stable features, after correcting for overall tendencies.

Correlations between Methods

Table below shows correlations between methods: upper triangle Pearson's r , lower triangle Spearman's ρ
Within cells: top is the correlation estimate and bottom the p -value. Significances at 0.05 and 0.01 are marked with stars. Strongest correlations are marked bold.

Figure to the right plots of all correlations (with outliers)

CM	0.11 0.413	0.14 0.307	-0.14 0.309	-0.03 0.834	0.28* 0.035	0.25 0.061	0.21 0.116
0.15 0.264	CC	0.81** 1.11e-14	0.62** 2.40e-7	0.57** 2.76e-6	0.36** 0.006	0.36** 0.006	0.28* 0.034
0.13 0.345	0.81** 1.39e-14	CR	0.72** 1.57e-10	0.69** 1.96e-9	0.46** 3.51e-4	0.28* 0.038	0.29* 0.026
-0.16 0.241	0.57** 3.74e-6	0.66** 4.86e-8	D	0.83** 4.44e-16	0.51** 4.80e-5	0.44** 6.45e-4	0.52** 3.45e-5
-0.01 0.947	0.56** 4.65e-6	0.64** 1.71e-7	0.83** <2.2e-16	M	0.64** 7.13e-8	0.40** 0.002	0.76** 5.27e-12
0.16 0.220	0.36** 0.006	0.47** 2.01e-4	0.54** 1.80e-5	0.58** 1.91e-6	P1	0.59** 1.38e-6	0.59** 1.74e-6
0.29* 0.031	0.34* 0.011	0.14 0.294	0.32* 0.015	0.48** 1.84e-4	0.51** 7.06e-5	P2	0.64** 1.10e-7
0.18 0.178	0.30* 0.022	0.32* 0.015	0.55** 7.67e-6	0.74** 4.03e-11	0.58** 2.32e-6	0.54** 1.51e-5	W



Principal Component Analysis

Loading in the same direction is marked bold

The first component (PC 1) explains more than half of the variance: it represents the agreement between the different methods.

The second component (PC 2) shows agreement between CC, CR, D and M against the other methods. These methods measure similar stability.

The third component (PC 3) shows agreement between CM, CC, CR, which are three methods from the same authors.

% variance explained	PC 1	PC 2	PC 3
	53.8%	17.3%	12.4%
CM	0.10	0.57	-0.67
CC	0.35	-0.32	-0.41
CR	0.38	-0.34	-0.39
D	0.41	-0.29	0.16
M	0.43	-0.13	0.24
P1	0.36	0.30	0.02
P2	0.33	0.38	0.17
W	0.36	0.35	0.35

References

- Cysouw, M., M. Albu & A. Dress (2008) Analyzing feature consistency using dissimilarity matrices. *STUF*, 61: 263–279.
- Dediu, D. (2011) A bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Proc R Soc B*, 278: 474–479.
- Haspelmath, M., M. S. Dryer, D. Gil & B. Comrie, eds. (2005) *The World Atlas of Language Structures*. Oxford University Press, UK.
- Maslova, Elena. 2004. Dinamika tipologic eskix raspredelenij i stabil'nost' jazykovyx tipov. *Voprosy Jazykoznanija* 2004(5). 3-16.
- Parkvall, M. (2008) Which parts of language are most stable? *STUF*, 61: 234–250.
- Wichmann, S. & E. W. Holman (2009) *Assessing Temporal Stability for Linguistic Typological Features*. LINCOM Europa, München.

Stable Features

The loading of each feature on the first principle component can be interpreted as the strength of agreement between the methods about the stability. We list here the top 20 'most stable features' from WALS. Note the frequent occurrence of word order features and phonemic features.

No.	WALS Feature	First Principle Component
1	18: Absence of Common Consonants	4.41
2	11: Front Rounded Vowels	3.48
3	136: M-T Pronouns	3.28
4	86: Order of Genitive and Noun	3.28
5	83: Order of Object and Verb	3.21
6	85: Order of Adposition and Noun Phrase	2.94
7	73: The Optative	2.81
8	80: Verbal Number and Suppletion	2.61
9	82: Order of Subject and Verb	2.35
10	119: Nominal and Locational Predication	2.25
11	10: Vowel Nasalization	2.14
12	6: Uvular Consonants	1.94
13	107: Passive Constructions	1.87
14	89: Order of Numeral and Noun	1.45
15	118: Predicative Adjectives	1.38
16	9: The Velar Nasal	1.37
17	7: Glottalized Consonants	1.36
18	87: Order of Adjective and Noun	1.31
19	13: Tone	1.22
20	44: Gender Distinctions in Pronouns	1.16