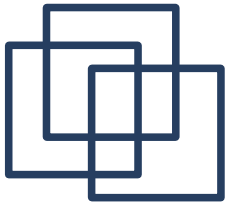


Temporal stability of features in Jazyki Mira

Oleg Belyaev
(obelyaev@gmail.com)

Defining stability

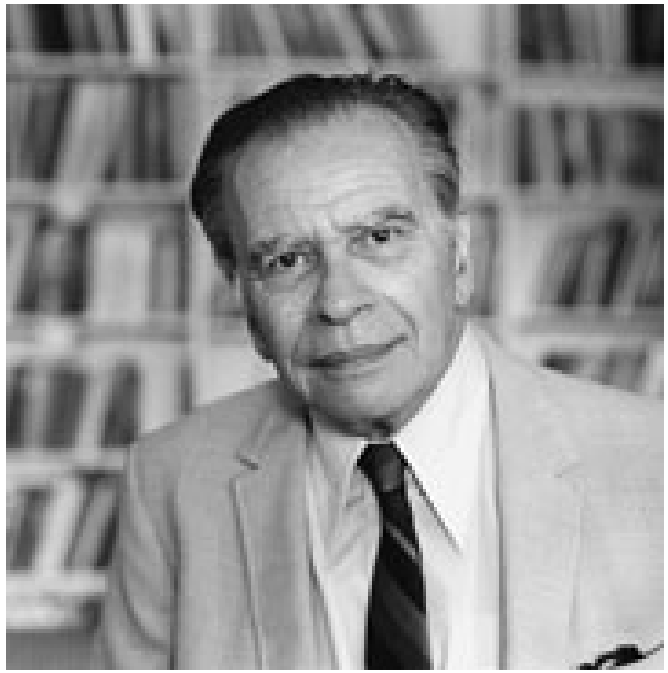
- A concept of an inherent 'stability' of a linguistic feature can be used to explain the historic development of languages.
- Some features are more prone to change, while others are less so.
- Claims are often made that some parts of grammar are more 'stable' than others. E. g. «Morphology is more stable than syntax».

Defining stability

- Edward Sapir, «Language»: slowly modifiable features which are more peculiar to the core or 'genius' of a language

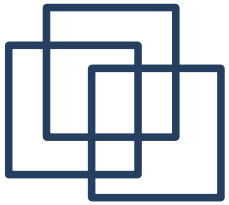
Defining stability

- Joseph H. Greenberg:



If a particular phenomenon can arise very frequently and is highly stable once it occurs, it should be universal or near universal (...). If it tends to come into existence often and in various ways, but its stability is low, it should be found fairly often but distributed relatively evenly among genetic linguistic stocks (...). If a particular property rarely arises but is highly stable when it occurs, it should be fairly frequent on a global basis but be largely confined to a few linguistic stocks (...). If it occurs only rarely and is unstable when it occurs, it should be highly infrequent or non-existent and sporadic in its geographical and genetic distribution (...)

(Greenberg 1978: 76)



Defining stability

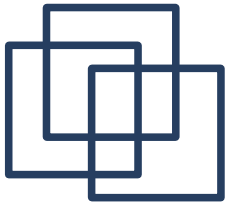
- Johanna Nichols (Nichols 1992, Nichols 1994)

  A set of precise metrics based on the frequencies of features among different families is proposed and applied.

- Also cf. Maslova 2004, Dahl 2004 for similar metrics.

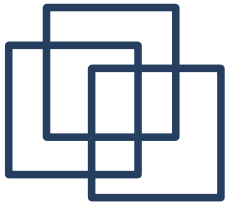
Defining stability

- Wichmann and Holman (n. d.)
- Uses a metric similar to Nichols (1992, 1994), but it is applied to WALS data, making it the first stability calculation to be based on a large typological database.
- Stability is defined as the probability that a given feature would remain unchanged during an arbitrary period of time.



Defining stability

- The values we can obtain from typology are not and cannot be true mathematic probabilities.
- What we can do is to use a metric which would serve as an approximation of which features are more stable than other features.



Stability in Jazyki Mira

- Jazyki Mira: a typological database similar to WALS, with 3821 binary features for 318 languages of Eurasia.
- Stability as a means of comparison with WALS: apply the same metric to JM's data.
- Wichmann & Holman include stability values for WALS features in binary form.

The metric used

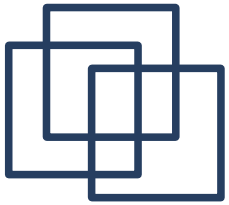
- The metric uses two frequency values, *R* and *U*.
- For *R*, one should count the proportion of language *pairs* for which the feature has the same value in each of the genera present (genera defined as per Dryer 2001).
- Then, one should calculate a weighted average of the values for all genera with the weight being the square root of the number of languages for a given genus:

$$R = \frac{\sum r_i * w_i}{\sum w_i}, w_i = \sqrt{n_i}.$$

The metric used

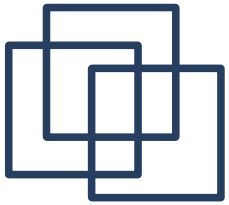
- The value for *U* is calculated in a similar way: one gets the proportion of all the pairs of *unrelated* languages where the value of the feature is different.
- The final value of stability is calculated via the formula (Albatineh et al. 2006 cited as explanation):

$$S = \frac{R - U}{1 - U}.$$



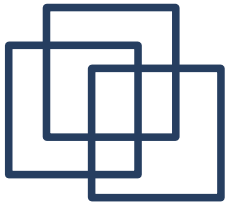
Application to JM

- All features are considered to be attested for all languages.
- The classification is the same as the one used by Wichmann & Holman: genera for calculating *R* and WALS families for calculating *U*.
- 57 reliable correspondences originally established.
- In this overview only features with $U < 0.9$ are considered.



The results

- See your handouts
- Four-way categorization by Wichmann & Holman:
- very stable: 51.8 – 100.0
- stable: 32.8 – 51.7
- unstable: 19.2 – 32.7
- very unstable: -62.8 – 18.9
- (for binary values of WALS features)



The results

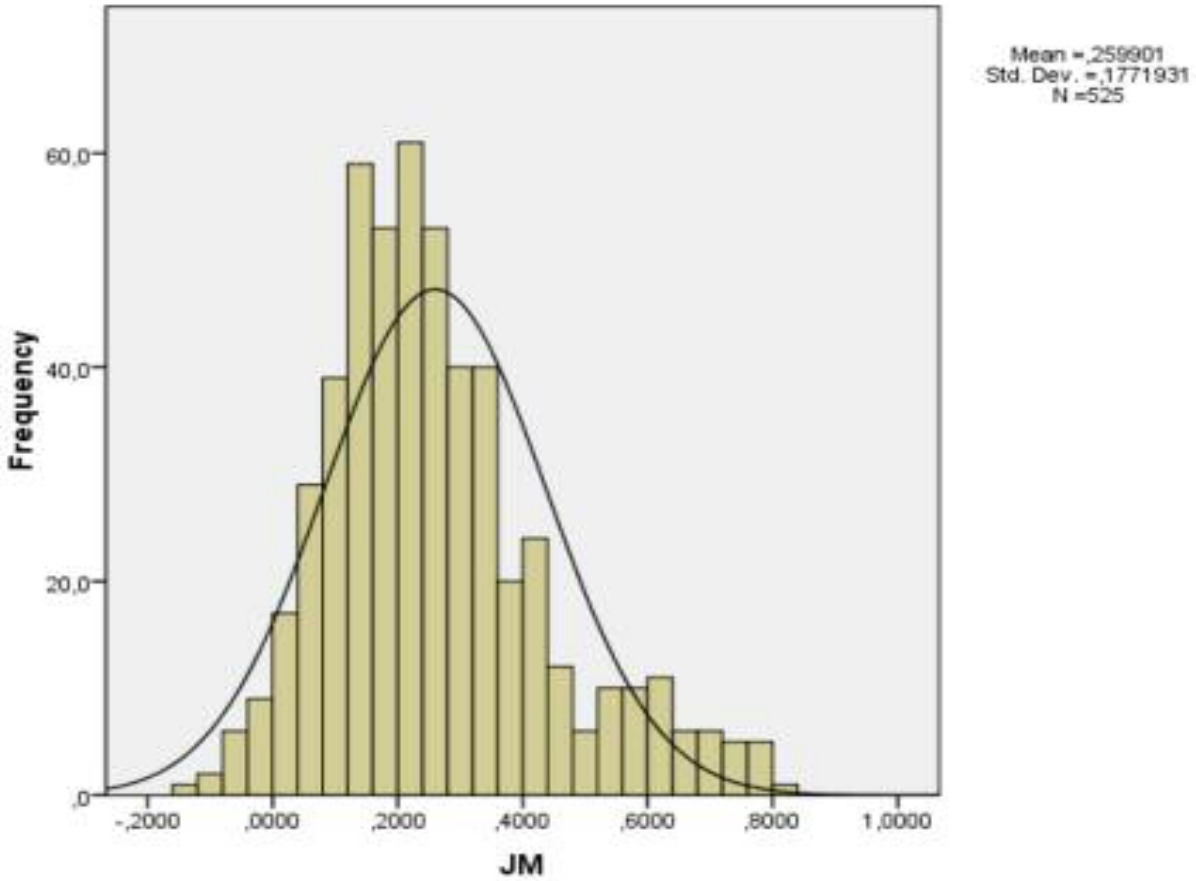
- See your handouts.
- Of the 42 feature pairs, 23 (~55%) fall into the same categories,
- 13 (~31%) are in adjacent categories (blue),
- 6 (~14%) have no correlation at all (red).
- This means that the correlation is acceptable for 86% of feature pairs.

The results

- For the 13 statements in the literature analyzed, there seems to be no correlation in only 2 cases (nasal vowels and numeral classifiers), and these have questionable interpretation for WALS.
- VSO word order: probably not enough data in JM.

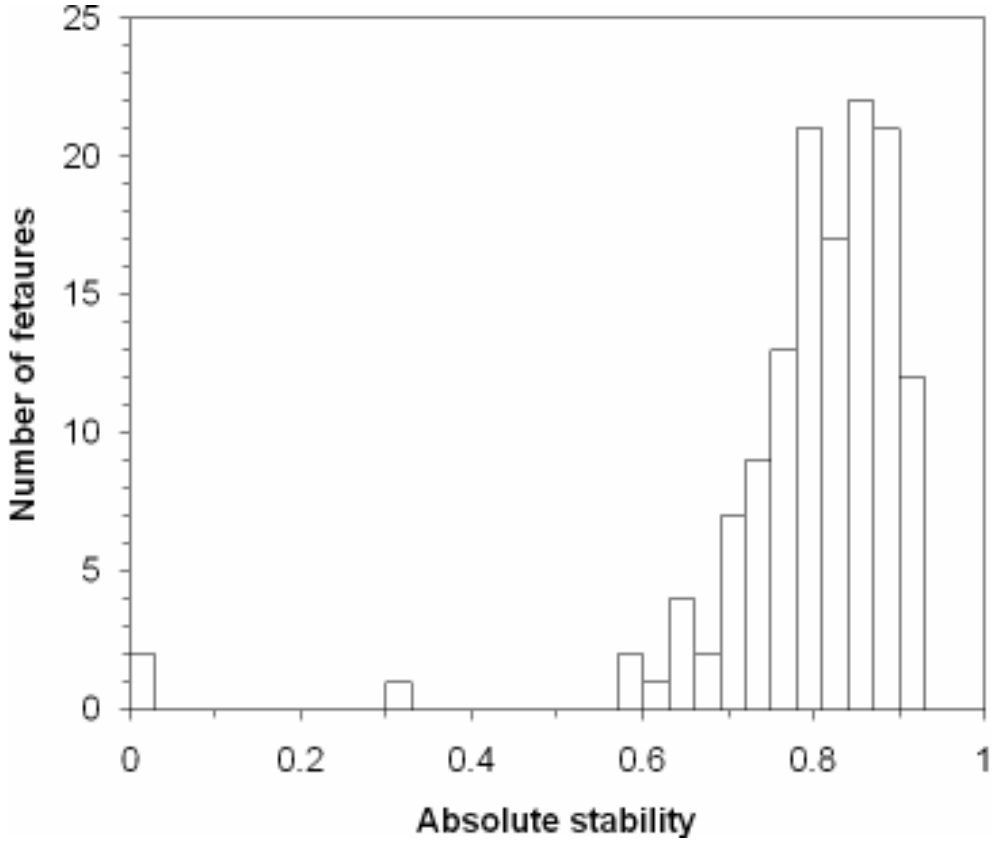
The results

- The distribution of JM stabilities for features where $U < 0.9$ (525 features)



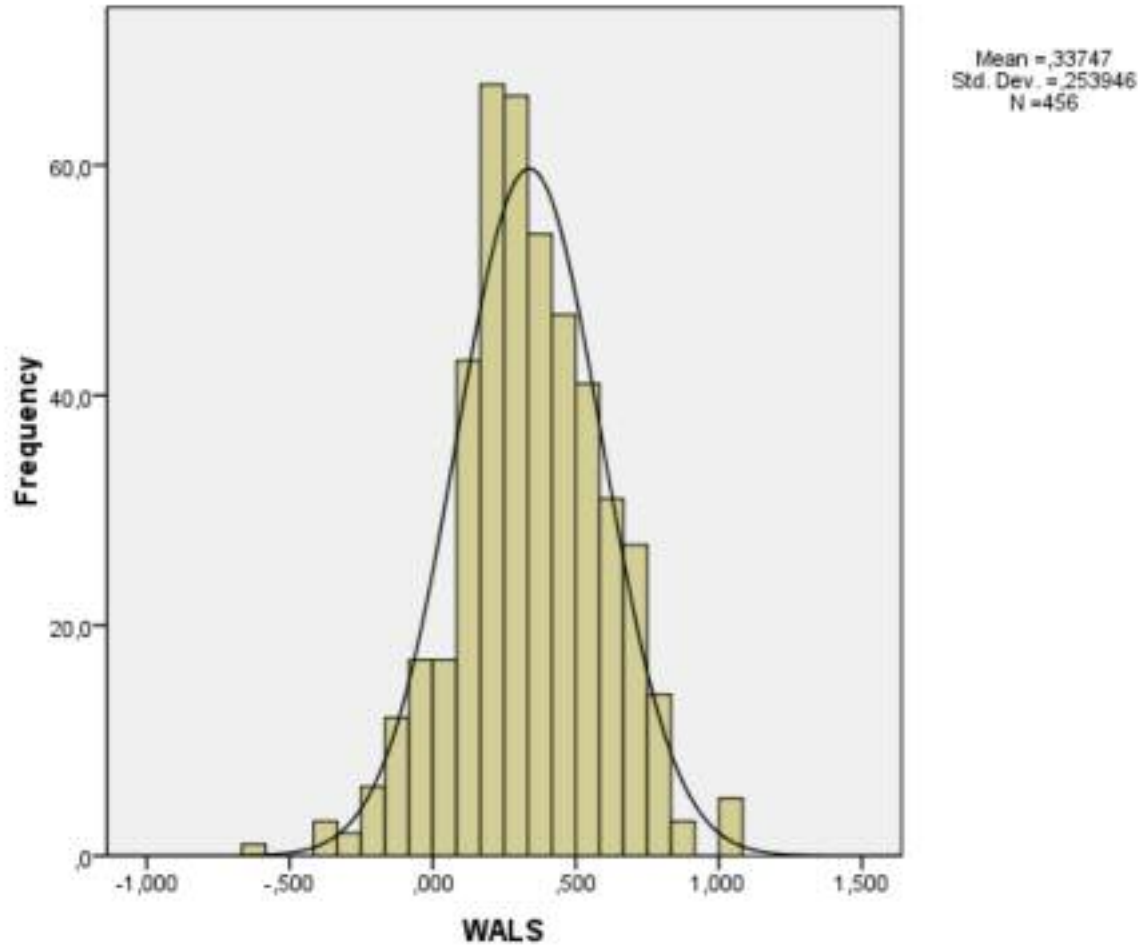
The results

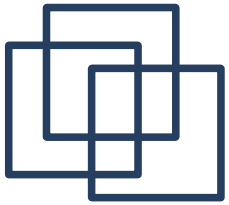
- The WALS distribution included in Wichmann & Holman's paper (for multi-value features!)



The results

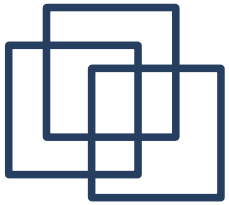
- The WALS histogram for decomposed binary features





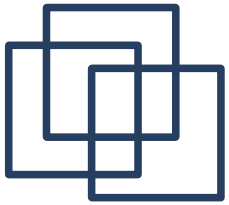
The results

- This seems to show that, while the two databases are different in structure and scope, the data contained in them is in fact comparable and can be used for conducting similar research.
- Which database is more suited for calculating stability remains, as of yet, unclear.
- Examples of highly stable features in both databases: gender, ergative alignment, SVO word order.



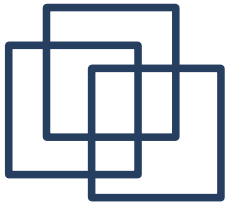
Testing the results

- Conducting a simulation similar to the one done by Wichmann and Holman.
- Using the obtained values as weights when calculating typological similarity (for later constructing phylogenetic trees; cf., e. g., Wichmann & Saunders 2007 and Polyakov & Solovyev 2006).
- Double-checking the metric on two databases.



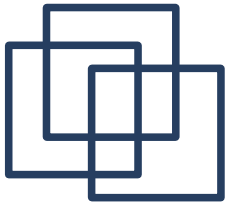
Problems

- JM only contains binary features, while counting stability could be useful for some multi-value features. E. g., the stability of «number of vowels» vs. the stability of «8 vowels».
- Binary features are sometimes too gradual to provide adequate statistics; for this reason, and for the sake of comparison with WALS, combining them into one would sometimes be useful.
- Stability can probably be decomposed into stability *per se* and borrowability



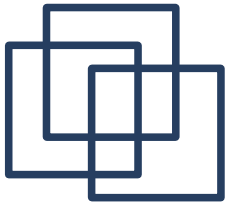
Possible solutions

- It is possible to extend JM with additional markup which would group binary features into multi-value ones.
- Unifying several features into one is a purely computational problem, but the sheer number of JM's features requires one to manually sort through them.



Conclusions

- The results seem to show that WALS and JM are comparable with regard to their usefulness in conducting quantitative research.
- Geographical limits of JM do not seem to impact stability in a significant way.
- To make comparison more fruitful it is crucial, however, to find better ways of establishing feature correspondences.
- Two databases allow double-checking data.



Acknowledgements

- Special thanks to Søren Wichmann, Eric W. Holman, Vladimir Polyakov, Valery Solovyev and Dmitry Egorov for their aid, suggestions and helpful discussion.

References

- Albatineh, Ahmed N., Magdalena Niewiadomska-Bugaj, and Daniel Mihalko. 2006. On similarity indices and correction for chance agreement. Journal of Classification 23. 301-313.
- Dahl, Östen. 2004. The Growth and Maintenance of Linguistic Complexity. Amsterdam: John Benjamins.
- Dryer, Matthew S. 2001. List of genera, available at http://www.wals.info .
- Greenberg, Joseph H. 1978. Diachrony, synchrony and language universals. In Universals of Human Language, Vol. III: Word Structure, ed. by Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik, pp. 47–82. Stanford: Stanford University Press.
- Maslova, Elena. 2004. Динамика типологических распределений и стабильность языковых типов [Dynamics of typological distributions and stability of language types]. Voprosy jazykoznanija 5.3-16.
- Nichols, Johanna. 1992. Linguistic Diversity in Space and Time. Chicago: The University of Chicago Press.
- Nichols, Johanna. 1995. Diachronically stable structural features. Historical Linguistics 1993. Selected Papers from the 11th International Conference on Hitorical Linguistics, Los Angeles 16-20 August 1993, Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Polyakov, Vladimir N. and Solovyev, Valery D. 2006. Компьютерные модели и методы в типологии и компаративистике [Computational models and methods in typology and comparative linguistics]. Kazan.
- Wichmann, Søren, and Eric W. Holman (pending publication). Assessing temporal stability for linguistic typological features. Available online: http://email.eva.mpg.de/~wichmann/WichmannHolmanIniSubmit.pdf
- Wichmann, Søren and Arpiar Saunders. 2007. How to use typological databases in historical linguistic research. Diachronica 24.2: 373-404. Available online: http://email.eva.mpg.de/~wichmann/ICHL05-FINALSUBMIT2.pdf .