Daria Alfimova

Kiel University; Cluster of Excellence ROOTS, Kiel

CIRCUM-BALTIC OBJECT MARKING AGAINST A BROADER AREAL PERSPECTIVE

The problem of cross-linguistic comparison of argument coding has been attracting the attention of linguists at least since the seminal work by [Jakobson 1936]. Today, there are plenty of studies dedicated to it as well as such indispensable typological tools as ValPal [Hartmann et al. (eds.) 2013], BivalTyp [Say (ed.) 2020], the Slavic flagging database [Seržant et al. 2021]. The current study deals with the clustering of argument coding patterns in the Eurasian macroarea with a focus on the languages of the Circum-Baltic (CB) area. The study aims to find out whether the languages within the CB area have more coding patterns in common compared to the languages that are genealogically or areally related to them but do not belong to the area. In terms of the number of examined languages, my approach to coding patterns is broader than the one presented by Koptjevskaja-Tamm and Wälchli [Koptjevskaja-Tamm, Wälchli 2001] but less comprehensive than in recent works by Say [2014, 2018].

Method. My data is subsampled from the overall sample published in BivalTyp (accessed on July 14, 2021). I consider 105 predicates in 31 languages of Eurasia with ≥ 500000 speakers according to Ethnologue. Only 2% of the patterns missing in the overall sample are filled by me (considering all available sources including consultations with native speakers). To make coding patterns comparable, I suggest abstract concepts for the second arguments (P-arguments) of all predicates in the subsample, following the main ideas from [Levinson, Meira 2003; Haspelmath 2010]. For example, the Russian pattern *iz*/*s* + genitive and German *aus* + dative are getting labeled as FROM since in both languages these patterns are mostly used with motion verbs and denote the starting point of movement. The total amount of obtained labels for P-arguments is 21; the first arguments of the predi-

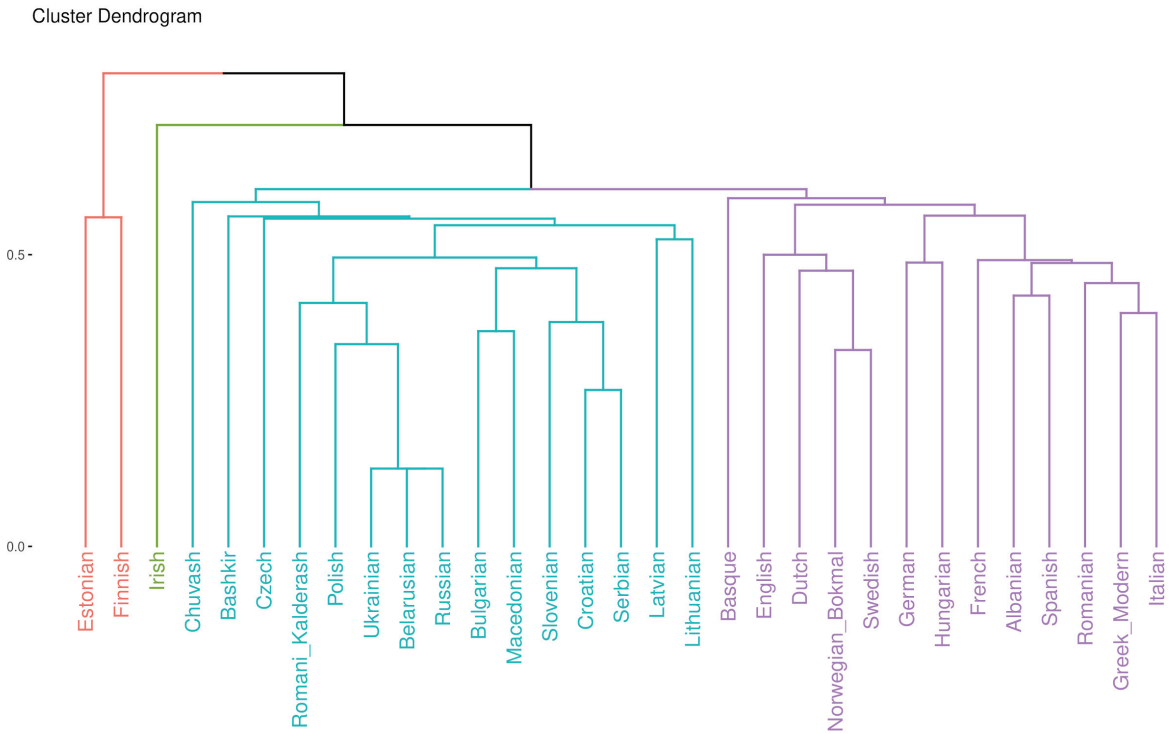

Figure 1. Cluster dendrogram for all languages of the subsample

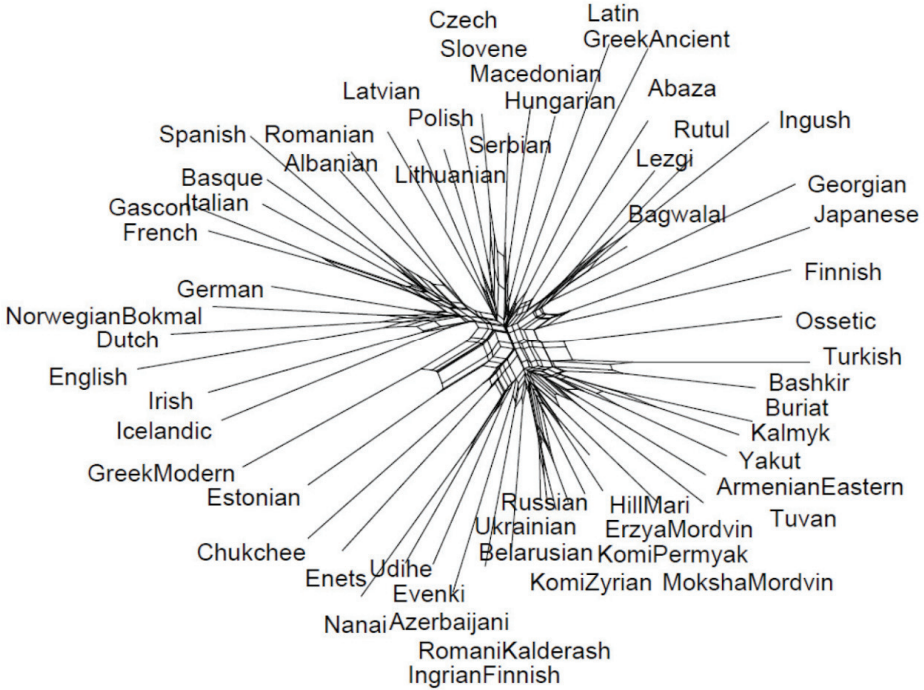

Figure 2. Visualization from Figure 8 from [Say 2018]

cates are not taken into account. The hierarchical clustering analysis is applied to the data drawing on the *hclust* function [R Core Team 2020]. The result is plotted in Figure 1 using the *factoextra* package [Kassambara, Mundt 2020].

Results and discussion. As Figure 1 shows, there is no solid opposition of CB languages vs. non-CB languages. Four clusters can be distinguished: 1) Northern-European Cluster (uniting Finnish and Estonian); 2) Insular Cluster (Irish only); 3) Central / Eastern-European Cluster (Balto-Slavic languages and some of the indigenous languages of Russia as well as Kalderash Romani; 15 in total); 4) Southern / Western-European Cluster (13 in total). This split has to be explained from both genealogical and areal perspectives. Genealogical clusters are clearly visible: Slavic and Germanic subclusters are essential parts of two separate larger clusters. It is noteworthy that the obtained distribution of Slavic languages corresponds to one from [Seržant et al. (to appear)]. Areal explanations apply to the node connecting the Balto-Slavic and two Turkic (Bashkir, Chuvash) languages, the German-Hungarian subcluster, and the "Balkanoic" subcluster consisting of Romance languages, Modern Greek and Albanian. As expected, this clustering partially overlaps with the result published by S. Say [Say 2018], cf. Figure 2 and Figure 1. In further research, more predicates should be examined (preferably with cross-linguistically diverse P-arguments coding).

References

Hartmann et al. (eds.) 2013 — I. Hartmann, M. Haspelmath, B. Taylor (eds.). *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. Available at http://valpal.info (accessed on 03.09.2021).

Haspelmath 2010 — M. Haspelmath. Comparative concepts and descriptive categories in crosslinguistic studies. *Language.* 2010. Vol. 86. No. 3. P. 663–687.

Jakobson 1936 — R. Jakobson. Beitrag zur allgemeinen Kasuslehre. Gesamtbedeutungen der russischen Kasus. R. Jakobson. *Selected Writings*. Vol. 2: *Word and Language*. 1936. S. 23–71.

Kassambara, Mundt 2020 — A. Kassambara, F. Mundt. *factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version*

1.0.7. 2020. Available at: https://CRAN.R-project.org/package=factoextra (accessed on 03.09.2021).

Koptjevskaja-Tamm, Wälchli 2001 — M. Koptjevskaja-Tamm, B. Wälchli. The Circum-Baltic languages. An areal-typological approach. Ö. Dahl, M. Koptjevskaja-Tamm (eds.). *Circum-Baltic languages*. Vol. 2: *Grammar and typology*. (Studies in Language Companion Series 55). Amsterdam; Philadelphia: John Benjamins Pub., 2001. P. 615–750.

Levinson, Meira 2003 — S. Levinson, S. Meira, The Language and Cognition Group. 'Natural concepts' in the spatial topological domain — Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language*. 2003. Vol. 79. No. 3. P. 485–516.

R Core Team 2020 — R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. Available at: https://www.R-project.org/ (accessed on 03.09.2021).

Say 2014 — S. Say. Bivalent verb classes in the languages of Europe: A quantitative typological study. *Language dynamics and change*. 2014. Vol. 4. No. 1. P. 116–166.

Say 2018 — S. Say. Markirovaniye aktantov dvukhmestnykh predikatov: Predvaritelnyye itogi tipologicheskogo issledovaniya [Bivalent predicates' argument encoding: Preliminary results of a typological study]. S. Say (ed.). *Valentnostnyye klassy dvukhmestnykh predikatov v raznostrukturnykh yazykakh* [Valency classes of two-place predicates]. St. Petersburg: Institute for Linguistic Studies, RAS, 2018. P. 116–166.

Say (ed.) 2020 — S. Say (ed.). *BivalTyp: Typological database of bivalent verbs and their encoding frames.* Saint Petersburg: Institute for Linguistic Studies RAS, 2020. Available at: https://www.bivaltyp.info (accessed on 03.09.2021).

Seržant et al. 2021 — I. A. Seržant, B. Wiemer, E. Bužarovska, M. Ivanová, M. Makartsev, S. Savić, D. Sitchinava, K. Skwarska, M. Uhlik. Dataset for the paper "Areal and diachronic trends in argument flagging across Slavic". Available at https://zenodo.org/record/4749564#.YTLZ5I4zaUm (accessed on 03.09.2021).

Seržant et al. (to appear) — I. A. Seržant, B. Wiemer, E. Bužarovska, M. Ivanová, M. Makartsev, S. Savić, D. Sitchinava, K. Skwarska, M. Uhlik. Areal and diachronic trends in argument flagging across Slavic. Alignment and Alignment Change in the Indo-European Family. To appear.