

Employing cluster analysis to explore the diachronic dynamics of Slavic aspect

The talk demonstrates how cluster analysis can help discover changes among patterns of stem derivation which are essential for the Slavic aspect system.

The backbone of this system builds on two productive patterns of stem-derivation (Breu 2000; Wiemer/Seržant 2017): either an unprefixated verb stem (simplex, imperfective/IPFV1) is prefixed (see 1), or an already prefixed stem (perfective/PFV) is additionally suffixed (see 2) yielding another imperfective stem (IPFV2). If the two stems replace each other under clearly definable conditions without change of lexical meaning, the stems are not just synonyms, but so-called aspect (perfective—imperfective) pairs. Compare Czech

	IPFV1		PFV		IPFV2	
(1)	<i>psá-t</i>	⇒	<i>na-psa-t</i>			‘write’
(2)			<i>pře-psa-t</i>	⇒	<i>pře-pis-ova-t</i>	‘rewrite’

Pairs of type (1) arise if the prefix meaning overlaps with a semantic component of the simplex (Janda et al. 2013). If such pairs acquire another ipfv. stem (IPFV2) via suffixation, aspect triplets result (e.g., *dělit* ⇒ *roz-dělit* ⇒ *rozdě-l-ova-t* ‘separate’). Triplets can “survive”, e.g. if IPFV1 and IPFV2 remain synonyms, but develop preferences for different subsets of functions associated with imperfective aspect. Otherwise triplets decay: the lexical meanings of the imperfective stems dissimilate, or one of them becomes obsolete. Triplets vary considerably as for their (i) semantics and polysemy, (ii) frequency (both IPFV1 vs IPFV2 from the same triplet and entire triplets), (iii) “age”, and (iv) persistence. We search for patterns in the interplay of these parameters, in particular: do the persistence (or decay) of triplets and their meaning ranges correlate with their age, prefixes, semantic features, or with frequency?

For this purpose we use a comprehensive database of potential triplets in Czech, Polish and Russian for 1750-2017, subdivided into six subperiods. Frequencies drawn from corpora (see References) are used for cluster analysis applied to these subperiods, for both IPFV2 and IPFV1. By employing time series clustering with non-Euclidean distance measures we propose a certain methodological transfer from engineering (Iglesias/Kastner 2013) and economy (Roelofsen 2018) to linguistic data. Contrary to previous cluster analyses (e.g., Divjak/Fieller 2014), diachronic change figures as a variable, and clusters are based on frequency changes, not (as in Jansegers/Gries 2019) on extensive annotation. The obtained clusters provide a starting point for the qualitative analysis of selected typical and untypical representatives, with different prefixes and inherited vs borrowed (before or after 1750) roots from each cluster. Persistence rates and changes in meaning alternations and the lexical relation between IPFV1 and IPFV2 are established.

Preliminary observations, based on the Czech data (651 triplets), suggest that the lexical relation between IPFV1 and IPFV2 stems, and thus the subsystem of triplets, shows considerable diachronic stability, since

- triplets with IPFV2 attested before 1750 have decreasing frequency, but they have persisted.
- newer triplets have low, but increasing frequency.

Deadjectival triplets differ: simplex stems appeared after the prefixed stems and more frequently tend to disappear again (e.g., Czech †*slabit* ⇐ *oslabit* ⇒ *oslabovat* ‘weaken’).

In the talk we will compare the results of this explorative method for Czech, Polish and Russian and specify conclusions concerning the stability of the aspect system.

References

- Breu, Walter. 2000: Zur Position des Slavischen in einer Typologie des Verbalaspekts (Form, Funktion, Ebenenhierarchie und lexikalische Interaktion). In: Breu, Walter (ed.): *Probleme der Interaktion von Lexik und Aspekt (ILA)*. Tübingen: Niemeyer, 21-54.
- Divjak, Dagmar & Nick Fieller. 2014: Cluster analysis. Finding structure in linguistic data. In: Dylan, Glenn & Justyna A. Robinson (eds.): *Corpus Methods for Semantics (Quantitative studies in polysemy and synonymy)*. Amsterdam, Philadelphia: Benjamins, 405-441.
- Iglesias, Felix & Wolfgang Kastner. 2013: Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns. *Energies* 6, 579-597.
doi:10.3390/en6020579
- Janda, Laura A., Anna Endresen, Julia Kuznetsova, Olga Lyashevskaya, Anastasia Makarova, Tore Nessel & Svetlana Sokolova. 2013: *Why Russian Aspectual Prefixes Aren't Empty (Prefixes As Verbal Classifiers)*. Bloomington (Indiana): Slavica.
- Janseger, Marlies & Stefan Th. Gries. 2019: Towards a dynamic behavioral profile: A diachronic study of polysemous *sentir* in Spanish. *Corpus Linguistics and Linguistic Theory*, 1-43.
- Roelofsen, Pjotr. 2018: *Time series clustering*. Amsterdam: Vrije Universiteit Amsterdam (Master Thesis). Available at <https://www.math.vu.nl/~sbhulai/papers/thesis-roelofsen.pdf> [03.01.2020.]
- Wiemer, Björn & Ilja A. Seržant. 2017: Diachrony and typology of Slavic aspect: What does morphology tell us? In: Bisang, Walter & Andrej Malchukov (eds.): *Unity and diversity in grammaticalization scenarios*. Berlin: Language Science Press, 230-307.

Corpora used

1) Czech

- Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Jelínek, T., Kovářiková, D., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondříčka, P., Zasina, A.: *Korpus SYN, verze 7 z 29. 11. 2018*. Ústav Českého národního korpusu FF UK, Prague 2017. (Available at <http://www.korpus.cz>)
- Kučera, K., Řehořková, A., Stluka, M.: *DIAKORP: Diachronní korpus, verze 6 z 18. 12. 2015*. Ústav Českého národního korpusu FF UK, Prague 2015. (Available at <http://www.korpus.cz>)

2) Polish

- Electronic corpus of 17th and 18th century Polish texts (*KorBa*): <http://clip.ipipan.waw.pl/KORBA>
- NKJP – Polish National Corpus: <http://nkjp.pl/>
- corpus of 12 mln tokens (texts from 1750-1917) compiled for *DiAsPol*
- *DiAsPol* parallel corpus (see <http://www.diaspol.uw.edu.pl/eng/>)

3) Russian

- Russian National Corpus: <http://www.ruscorpora.ru/new/>