

A database of pragmaticalization: discourse formulae of Russian, Serbian and Slovene

Anton Buzanov, anton.buzanov.00@gmail.com
Polina Byčkova, polyatomson@gmail.com
Arina Molčanova, arina_molchanova@bk.ru
Anna Postnikova, anna.postn90@mail.ru
Daria Ryžova, daria.ryzhova@mail.ru
Miljka Stanković, miljkastankovic@gmail.com
HSE University, Moscow

The paper presents a database of discourse formulae designed as a tool for typological analysis of pragmaticalization processes. Grammaticalization, i.e. the process in which a language unit becomes (more) grammatical (Lehmann 1982), has been under active consideration for more than half a century. Significant results have been achieved in this field, including the lists of potential lexical source meanings and grammatical target meanings (Kuteva et al. 2019), as well as the detailed descriptions of the mechanism itself (Hopper & Traugott 2003, Boye & Harder 2012). On the contrary, there are very few studies in pragmaticalization (Diewald 2011), i.e. a process in which a linguistic element becomes a pragmatic item, cf. a refusal marker). It is yet to be discovered, which meanings should be considered as pragmatic ones and what are the sources for pragmatic units.

To address this problem, we focus on studying the typology of discourse formulae. We define discourse formulae as multiword idiomatic reactions to another utterance. These units are evidently pragmatic, as they only occur in a dialogue and express the speaker's attitude to the interlocutor's words. Discourse formulae form a well-defined, though not a closed class of linguistic items. Instead, this class appears to be subject to constant renovation, and thus provides a very rich basis for the analysis of the pragmaticalization processes.

In order to study discourse formulae typologically, we are building a database which would aggregate data from structurally different languages. As a starting point, we use Russian, Serbian and Slovene formulae which already show substantial grammatical and lexical variability.

In the sample, there are formulae with very similar structure and meaning, cf. rus. *Bog znaet*, srp. *Bog zna*, slo. *Bog ve*. All of them consist of the noun 'God' and the verb 'to know' in present tense and express the speaker's inability to answer the question, cf. eng. *who knows*. There are also cases where formulae have the same components, but different semantics. For example, Serbian formula *Kako da ne* (lit. 'how to no') and the Slovene cognate *Kako da ne* are mostly used as a marker of confirmation, while the Russian structural parallel *Kak bi ne tak* conveys an opposite meaning of negation.

Meanwhile, some structures are only present in one of the three languages, cf. srp. *taman posla* 'just job.GEN.SG' and *kamo sreće* 'where happiness.GEN.SG' both denoting refusal, or a conjunction of Russian emphatic particles *nu i nu*, which denotes surprise. Interestingly, Russian has many formulae that consist exclusively of function words, while in Serbian and Slovene such structures are almost totally absent.

The database offers detailed information on each formula, including but not limited to glosses, general syntactic structure and variation (discourse formulae can have slight structural variability, e.g., switch emphatic particles), main pragmatic function, additional semantics, and the type of the context (see a fragment of annotation in Table 1). The database uses PostgreSQL syntax and allows for various queries. The interlinear glosses provided for every formula enable the search by a particular word or grammatical category (for instance, one can find all formulae containing a verb of speech, or an imperative). The features from different columns can also be combined: the search by a combination of a pragmatic function and an additional semantic tag will result in a list of approximate translation equivalents. We refrain from

forming the synonym clusters manually, as the database interface can provide clusters of formulae based on different parameters.

In our talk, we will present the database and provide preliminary theoretical generalizations based on the Slavic data. Other languages (such as German, English, and Chinese) will shortly be added to the database. With the sample increasing, more observations are to be made.

Language	Formula	Variation	Realisation	Glosses	Example	Pragmatic function	Additional semantics	Context type	Syntax
Serbian	Bog zna	<i>a Bog će ga znati</i> <i>Bog bi ga znao</i> <i>Bog bi znao</i>	<i>bog zna</i>	God know.PR S.3SG	– <i>Šta misliš, kako će se ovo svršiti?</i> – <i>upitao sam ga.</i> – <i>Bog bi ga znao odgovori on – videćemo.</i> ‘– What do you think, how will it end? – I asked him. – Who knows, he answered – we’ll see’.	refusal	refusal to answer cause: no info	polar question specific question	N V
Russian	kak by ne tak	<i>da kak by ne tak</i> <i>a vot kak by ne tak</i>	<i>kak by ne tak</i>	how COND not so	– <i>Možet, on uedet, a?</i> – <i>probormotal Goš s nadeždoj.</i> – <i>Kak by ne tak...</i> ‘Maybe he’ll leave? – Gosh muttered with hope. – You wish...’	negation	expert negative assessment	hypothesis	PronInt Cl
Slovene	kako da ne	-	<i>kako da ne</i>	how that no	– <i><...> pomeni, da si videl Stevensa v igri?</i> – <i>Kako da ne. V Nebraski.</i> ‘– It means that you have seen Stevens in the game? – Of course. In Nebraska’.	confirmati on	expert obvious	polar question hypothesis	PronInt Sconj Cl

Table 1. A fragment of the annotation in the database.

Abbreviations

Cl – clause, COND – conditional mood, GEN – genitive case, N – noun, PronInt – interrogative pronoun, PRS – present tense, Sconj – subordinate conjunction, SG – singular number, V – verb.

References

Boye, K., & Harder, P. (2012). A usage-based theory of grammatical status and grammaticalization. *Language*, 1-44.

- Diewald, G. (2011). Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics* 49 (2011), Nr. 2, 49(2), 365-390.
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge University Press.
- Kuteva, T., Heine, B., Hong, B., Long, H., Narrog, H., & Rhee, S. (2019). *World lexicon of grammaticalization*. Cambridge University Press.
- Lehmann, C. (1982). Thoughts on grammaticalization: A programmatic sketch, vol. 1. (Arbeiten des Kölner Universalien-Projekts, AKUP 48). Cologne: Institut für Sprachwissenschaft, University of Cologne.