

How to annotate a construction?

Any language allows the omission of phonologically realized arguments in certain context and to varying degrees. In my contribution, I would like to propose a possibility to annotate these syntactic zeros within a certain construction type of Present Day and Old Russian, the so-called Dative Subject Constructions (DSC).

Russian DSC have long been an object of interest for the linguists. While the status of their first argument is the focus of much attention, the frequency of their use has not been studied in detail, in particular in relation to their diachronic development. The present investigation intends to fill this gap by addressing, first, the following questions: 1) Are we dealing here with a productive construction type? and 2) Can we ascertain a significant increase or decrease of syntactical productivity of this construction type from the diachronic point of view?

The first argument in DSC can also be omitted. Traditionally, we distinguish two types of the zero subjects¹. In order to maintain this differentiation, I decided to talk about **PRO** as the subject of non-tensed clauses and **pro** if the speaker can decide in certain contexts between the overt realization of the subject or its omission (realization as **pro**). As for the dative subjects, notwithstanding their controversial subject status, we can maintain the same differentiation. For the null subject of non-finite clauses (**PRO**) the dative case have been assumed (see for example Comrie 1974). It is more interesting to observe constructions in which the overt first dative argument co-occur with the covert forms. First dative argument can be omitted (realized as **pro**) and the reference of the omitted pronoun can be elliptic and thus, reconstructed from the context (1). In the majority of cases, the omission of a pronoun results in the establishment of the indefinite reference (2).

What is it good for to annotate such syntactic relations in a modern corpus, which contains multiple annotation layers? Previous studies on DSC have revealed that the annotation of first null arguments would definitely be advantageous for an adequate research on productivity of this construction type. Thus, the analysis of the linguistic data, based on the SYNTAGRUS², a dependency-based treebank of Russian and RRuDi³ has shown that the annotation of the both corpora cannot meet all requirements of the present study. The SYNTAGRUS contents in fact a dependency-based annotation of the syntactical relations between the verbs (or predicatives) and their arguments (including the so-called dative subject relation). We have, however, to take into account the fact that only the overt occurrences of dative arguments are syntactically annotated. We need only look at the frequency of the covert occurrences of dative subjects (in some cases up to 90%⁴) and its variation across the different DSC, in order to comprehend the necessity of the annotation of the syntactical null arguments for the adequate research.

In my work, I aim at contributing to the solution of some challenging tasks. The first one is the question of how to mark these syntactic zeros. Coding such arguments is possible and have been realized within different project⁵ in different way. Here I would like to propose a to a large extent not theory-driven method of annotation of the dative null subjects, in consideration of the fact that they cannot be placed in a syntactic tree without being defined as the syntactical subject or not. Furthermore, it not always makes sense to trace their dependency to every potential lexeme, which it could co-occur with, by taking into account the assumption

¹ Both names can be traced back to the terminology of the Grammar of the Principles and Parameters and are used primarily in the context of the generative linguistics. Their use within the scope of this paper is based exclusively on convention and independent of any theory.

² <http://www.ruscorpora.ru/>

³ RRuDi - Regensburg Russian Diachronic Corpus: <https://www.slawistik.hu-berlin.de/de/member/meyerrol/subjekte/rrudi>

⁴ As for example in case of modal predicative *nel'zja* ('cannot') (Grillborzer 2014).

⁵ For example in AnCora (<http://clic.ub.edu/corpus/en>), null subject arguments are treated as 'empty' tokens.

of dealing with a syntactically productive construction type. On the other hand, we still need to find the methods to automatize the annotation process, i.e to identify the potential gaps automatically. The last part of my contribution is devoted to this issue.

Examples

(1) *I vy_i vstaěte, iděte umyvaetes', ili*
 AND YOU GET UP GO WASH YOUR FACE OR
ne umyvaetes'... esli ∅_i ne xočetsja.
 NOT WASH YOUR FACE IF **PRODAT**NOT WANT_{3PSg}
 ('And you get up, go and wash your face, or do not wash your face, if you don't want.')

[Evgenij Griškovec. OdnovrEmEnno (2004)]⁶

(2) *On dejstvitel'no okazalsja interesnym i prostym*
 HE INDEED TURNED OUT INTERESTING AND SIMPLE
v obščeenii sobesednikom, iz tex,
 IN CONVERSATION INTERLOCUTOR OF THOSE
s kotorymi ∅ prijatno obsudit' ljubuju temu ...
 WITH WHO **PRODAT**PLEASANT DISCUSS ANY SUBJECT

('He actually turned out as an interesting and conversable interlocutor, one of those with whom it is pleasant to discuss any subject.')

[Ol'ga Zueva. Skaži čto ja tebe nužna... // «Daša», 2004]⁵

References:

- Baayen, H. 1993. On frequency, transparency and productivity. In G. Booij & J. van Marle (Eds). *Yearbook of morphology 1992*. (pp. 181-208). Dordrecht/Boston/London: Kluwer.
- Comrie, B. 1974. The Second Dative: A Transformational Approach. *Slavic Transformational Syntax*, eds. R. Brecht and C. Chvany, Michigan Slavic Materials, No. 10, Ann Arbor: University of Michigan, 123–150.
- Franks, S. 1995. *Parameters of Slavic Morphosyntax*. Oxford: Oxford University Press.
- Goldberg, A. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago. University of Chicago Press.
- Grillborzer, Ch. 2014. *Sintaksis Konstrukcij s Pervym Datel'nyim Aktantom - Sinxronnyj i Diaxronnyj Analiz*. Dissertation thesis. University of Regensburg.
- Moore, J. and D. Perlmutter. (2000). What Does It Take to Be a Dative Subject? *Natural Language and Linguistic Theory* 18, 373–416.
- Testelec, J. G. (2001). *Vvedenie v obščij sintaksis*. Izdatel'stvo RGGU, Moskva.

Corpora:

- RNC - Russian National Corpus: <http://www.ruscorpora.ru/>
- RRuDi - Regensburg Russian Diachronic Corpus: <https://www.slawistik.hu-berlin.de/de/member/meyerrol/subjekte/rrudi>
- AnCora – Spanish and Catalan corpus: <http://clic.ub.edu/corpus/en>

⁶ <http://www.ruscorpora.ru/>