

The main aim of the project is to answer the question: **how many lexemes – or lexical units – TO are there?** The form *to* is one of the most frequent words in Polish and, at the same time, one of the most troublesome. Linguistic analyses devoted to such forms come to different conclusions, and dictionaries sometimes significantly diverge in the information about their grammatical status. Moreover, this form has many syntactic and semantics functions and is involved in interesting linguistic phenomena: it acts as a copula in sentences such as (1) and is involved in organisation of the thematic-rhematic structure of sentences (see (2)–(3)). It means that in (2), the form *To* indicates the rheme (focus) – new information, represented in the sentence by the accented word *Janek*. In (3), instead, *to* emphasises the thematic status of the word *Janek* (which may be paraphrased as *Jeśli chodzi o Janka...* ‘As for Janek...’).

- | | |
|---|--|
| (1) Janek to świetny lekarz.
‘Janek is a great doctor.’ | (2) To Janek kupił ten prezent.
‘It is Janek who bought this gift.’ |
| (3) Janek to pracuje w szkole.
‘As for Janek, he works in a school.’ | |

Sentences (1)–(3) are only a part of the uses of *to* known in linguistics. It also acts, among others, as an ordinary demonstrative pronoun and a conjunction. Despite numerous analyses devoted to the forms *to*, it is still unclear how many lexemes TO should be distinguished, what their syntactic and semantic properties are, and where the boundaries between them lie. By means of an empirical (corpus and survey), syntactic and semantic analysis – with particular emphasis on thematic-rhematic phenomena – I will try to fill this research gap. I assume that thorough empirical research will lead to the reevaluation of the previous analyses, which are largely based on examples constructed by linguists. Furthermore, an exploratory corpus investigation will make it possible to notice many different uses of *to*, possibly unnoticed in previous works. The verification of the previous analyses and of new hypotheses and generalisations will allow for a formalisation of the syntax of different constructions with *to*. Basing on a reliably established syntax, a formal semantic analysis will be conducted, taking into account the thematic-rhematic structure. Such a comprehensive study will make it possible to propose a well-motivated classification of the lexemes TO, which will be a significant contribution to Polish linguistics and a potential reference point for lexicographers.

One of the innovative features of the project will be the use of methodological triangulation, a modern approach to empirical research involving a combination of different methodologies – in the case of linguistics: corpus and survey research. Linguistic corpora, despite their significant impact on the quality of linguistic analyses, due to their limited size are not able to guarantee the full range of occurrence of particular language units. Only the combination of such investigation with carefully designed survey research makes it possible to establish a solid empirical basis for further linguistic analysis.

Besides solving the problem of number, range, and the syntactic and semantic properties of various TO lexemes, the project will try to contribute to general linguistic research on two important issues. The first concerns the types of copular constructions in natural languages. While syntactic research on such constructions with *to* has already been carried out, there is no analysis focused on their semantics. The use of modern formal semantic tools and representation methods for the analysis of copular constructions with *to* may result in interesting theoretical conclusions. The second issue important for general linguistics concerns the possibility of formal semantic modelling of phenomena related to the thematic-rhematic structure. Constructions with *to*, such as (2)–(3), are a promising research area that may allow for the formulation of new generalisations relevant to the field.

It is worth noting that the result of the empirical part of the project will be a database of theoretical statements regarding the grammatical status and the number of the lexemes TO found in the scientific literature, containing the information whether a given statement is covered by survey and corpus research. This database is to be a significant resource for researchers of Polish. The results of the syntactic and semantic analysis will be submitted in the form of scientific articles to renowned Polish and foreign linguistic journals (general, Slavic and semantic) and will be presented at Polish and international scientific conferences.