Reg. No: 2023/51/B/HS2/02217; Principal Investigator: dr Marcin Wągiel

Collectives and singulatives across languages

Human beings perceive things in the world and talk about things in the world in various manners. One of the ways in which we view external entities is that we intuitively conceptualize certain objects (either simplex or complex) as counting as 'one'. This mechanism is called INDIVIDUATION and human languages differ in how they express it. In this project, we will investigate two types of linguistic expressions that encode different modes of individuation. The first class is called COLLECTIVES and includes expressions such as *a committee of women* and *a pile of dishes*. They describe groups as units. The second one is called SINGULATIVES and resembles English phrases such as *a grain of rice* and *a drop of water*. They designate units among a group viewed as a homogeneous collection of objects.

Linguistic expression of mechanisms of individuation, and thus the form and meaning of collectives and singulatives are not uniform since, undeniably, natural languages vary a lot. However, this variation is not entirely arbitrary. To the contrary, decades of cross-linguistic research provided good evidence demonstrating that it is systematic and to a great extent predictable. It is not uncommon that languages from different language families and different parts of the world share multiple formal and semantic features. One of the main aims of linguistics is the search for non-trivial cross-linguistic patterns and discovering how they result from underlying universal properties of language faculty. Such universals can relate both to the form and to the meaning of a linguistic expression and our aim is to search for them by investigating the way collectives and singulatives look and behave across languages.

Our project is intended to bridge two different linguistic traditions: on the one hand, typological descriptive linguistics, which searches for parameters of variation across languages and discovers robust cross-linguistic patterns, and, on the other hand, theoretical linguistics, which develops formal mathematical models in order to explain the structure and meaning of natural-language expressions. Our goal is to equip linguists with new tools to pursue comprehensive and theory-driven cross-linguistic research and to provide a novel perspective on the role of collectives and singulatives in grammar.

We will focus on the structure and interpretation of collectives and singulatives, their different functions and interactions with other categories such as grammatical number and gender. For this purpose, we will employ an innovative collection of methods. In order to uncover non-trivial dependencies, we will utilize the Terraling database, which is a cutting-edge digital resource for cross-linguistic data collection. We will focus mainly on the relevant constructions in Slavic (with an emphasis on East Slavic, especially Ukrainian), Germanic (English, German and Dutch), Celtic (mostly Welsh and Breton), Romance (especially Romanian and Spanish), Semitic (mainly various dialects of Arabic and Maltese) and Nilo-Saharan (mostly selected Nilotic languages). The choice of the languages is not random. They have been carefully selected since they are characterized by certain similarities and differences that will allow us to identify underlying universal properties of collective and singulative expressions.

Once the linguistic patterns are discovered and described, we will use formal mathematical tools to model the smallest identified meaning components and the way in which they compose into bigger structures. In particular, we will combine Nanosyntax with compositional semantics. Nanosyntax is a theory of how complex linguistic forms are generated, whereas compositional semantics describes how meanings arise. The basic conceptual structures will be captured within mereotopology, which is a formal theory of parts and wholes that captures both the part-whole relation and spatial configurations of entities. As a result of our research, open and publicly accessible linguistic data sets in the Terraling database will be created. Moreover, we will propose a formal linguistic model that will help us to understand structural and semantic properties of collective and singulative expressions across languages of the world.