Reg. No: 2022/47/B/HS2/02999; Principal Investigator: dr hab. Marcin Maria Kilarski

Semantic universals of classifier systems

The aim of the project is to examine classifiers, a type of noun categorization that is widespread in the world's languages. Classifiers exhibit a remarkable diversity in the meanings they convey and the morphological and syntactic contexts in which they occur. Several types of classifiers are distinguished, including the most well-known numeral classifiers which occur with numerals, as in Mandarin Chinese in the phrase *yí liàng chē* (one VEHICLE car) 'one car', where *liàng* is the classifier for vehicles. Other classifier types include noun, possessive, verbal, deictic, and locative classifiers. Classifiers are found in typologically diverse languages, ranging from the analytic languages of Southeast Asia to the polysynthetic languages of North America. While classifiers share several common features with grammatical gender, the other main type of noun categorization, their number in a language may be much larger, as illustrated by around 100 classifiers in Mandarin Chinese. A wide range of semantic categories serve as the basis for categorization in classifier systems, including animate vs. inanimate, male vs. female, and social status, together with categories found among inanimates, including physical properties such as shape and size as well as function and value. These properties tend to correlate with the type of classifier system. For example, possessive classifiers in Oceanic languages distinguish such categories as valuable, edible, and drinkable, while verbal classifiers in the Athabaskan languages of North America are used to classify objects as, e.g., long, round, and flat/flexible.

As part of the project, we will create a database of over 3000 languages based on data available from over 7000 grammars and other descriptions, and from open-access databases under a Creative Commons Licence. Automatic detection followed by manual checking will be used to identify the classifier types and semantic categories found in classifier languages. We will thus try to answer the following questions: Does a language have a specific type of classifier, e.g., a numeral classifier or a possessive classifier? Which semantic categories, e.g., animate, human, long, and round are found in the classifier system of that language? Based on the compiled data, we will examine fundamental aspects of classifier systems: 1) the diversity of types of classifiers found in the world's languages; 2) the diversity of semantic categories used for categorization of humans, animals, and inanimate objects; 3) the distinction between universal vs. language-specific features; and 4) correlations between classifier type and the meanings that are expressed. For example, with regard to the universality of semantic categories, our aim is to determine which categories are more common and which ones tend to occur in particular language families or areas. As regards correlations between classifier types and their semantics, we will establish whether such correlations are indeed found, or whether our knowledge of the meanings expressed by classifier systems is influenced by available descriptions of the most well-known languages representing particular types, e.g., of Mandarin Chinese or Japanese in the case of numeral classifiers.

Although classifiers constitute one of the prototypical examples of linguistic diversity and the capacity of the human mind for categorization, our knowledge of their nature is constrained by the absence of data. Classifiers attracted only marginal interest in linguistics up to the early 1970s, and most existing descriptions of classifier types and their semantics are based on either qualitative descriptions, small-scale surveys, or case studies that typically focus on 'big' and well-described languages. The project will thus make a significant contribution to the study of nominal classification systems, linguistic typology, and linguistics in general. For the first time it will allow typological, quantitative, and evolutionary analyses of all types of classifiers, thus contributing to a better understanding of the variation in their semantics and means of expression. Since classifiers offer a window onto the complex ways of experiencing the world that are encoded in languages, the database will constitute a foundation for descriptions of not only the universal, cognitive basis of categorization but also its cultural, social, and environmental correlates. In turn, these descriptions will contribute to ongoing discussions concerning key issues in linguistics and more generally in the humanities and social sciences. These include categorization, linguistic complexity, and the correlations between language structure and cognitive competence, culture as well as the construction of social reality.

The planned output of the project includes an online database of classifier types and semantic categories, six topical journal articles presenting the key findings, and a monograph that offers an exhaustive presentation of the results of the project.