

Aleksandra Zięba

Tytuł rozprawy doktorskiej: Wyodrębnianie polskich, rosyjskich i czeskich jednostek wielowyrazowych o funkcji fatycznej i ematywnej z użyciem narzędzi i zasobów korpusowych. Studium leksykograficzne.

SUMMARY

The aim of this dissertation is to develop methods for isolating, analyzing, and describing multi-word structures of spoken language with phatic and emotive functions in Polish, Russian, and Czech. The point of reference is the observation that traditional lexicography – both monolingual and bilingual – records complex forms characteristic of oral communication only to a limited extent, as it focuses primarily on the single word as the basic unit of lexicographic description.

This issue, long present in phraseological and lexicographic research, arises from the difficulties involved in defining the boundaries of such constructions, their variability, ambiguity, and the marginalisation of spoken language. In this dissertation, a multi-word unit is understood as a *reproduct*—following the definition proposed by Wojciech Chlebda—namely, a fixed linguistic unit extracted from texts on the basis of its regular reproducibility and functioning as a verbaliser of the speaker's communicative intention. The dissertation seeks to bridge this gap by describing and classifying phatic and emotive *reproducts* and by attempting to determine their translational equivalents using both corpus-based and non-corpus resources.

The research material was primarily drawn from the Polish National Corpus, the Russian National Corpus, the Czech National Corpus, the MoncoPL corpus, and the parallel corpus InterCorp. The use of these tools made it possible to empirically extract multi-word units in authentic contexts and to compare them across the three languages. A particularly important role in this process was played by parallel corpora, which made it possible to examine translational equivalents with regard to situational and pragmatic criteria. As a complement, non-corpus sources were also used, such as monolingual and bilingual dictionaries, as well as online materials (including discussion forums and language portals). The study adopts the

principle that the corpus serves as the foundation of the research and as a tool supporting the analytical process, but it does not function as the ultimate authority. Corpus data require confrontation with other sources and interpretation within a broader lexicographic context.

The dissertation is of a theoretical and analytical nature.

The first part (Chapters 1–5) provides a theoretical overview, a discussion of research tools, a review of studies on monolingual and bilingual lexicography in Poland, Russia, and the Czech Republic, as well as the fundamental assumptions concerning the understanding of multi-word units, reproducibility, and corpus linguistics.

The second part (Chapters 6–8) is of an empirical nature and presents methods for extracting multi-word units of spoken language (including the dictionary–collocational, synonymic, and translational methods), as well as methods for determining their translational equivalents using parallel corpora and audiovisual material.

The culmination of the dissertation is the concept of an original Polish–Russian and Polish–Czech dictionary, which includes a proposal for its macrostructure and the design of the dictionary entry. This work is the result of combining solutions developed in previous lexicographic practice with original and innovative approaches adapted to the specificity of spoken language and the possibilities of electronic processing. The results of the conducted research confirm that spoken language, although often marginalised in traditional lexicographic descriptions, constitutes a valuable source of fixed and regularly reproduced multi-word structures that form an integral part of everyday communication. The application of corpus-based methods made it possible to identify such units and to indicate their potential translational equivalents. It should be noted, however, that the resources of spoken texts in corpora are limited, and some of them consist of materials stylised as oral language. Nevertheless, corpora consistently remain a valuable tool supporting the process of identifying and verifying units, enabling the formulation of preliminary hypotheses concerning their functioning and equivalence in other languages.

The dissertation addresses a wide range of issues in translational lexicography, phrasematics, and corpus linguistics, proposing a new model for describing multi-word language units in an interlingual and translational perspective.

Słowa kluczowe: reproduct, translation lexicography, corpus linguistics, phrasematics, multi-word units, phatic expression, emotive expression