

The aim of the project is to analyse changes in English phraseology between the earliest period in its history, i.e. Old English (5th-11th century) and the beginning of the 20th century. Phraseological investigations become increasingly popular thanks to the availability of large electronic collections of language data (the so-called corpora), which shows that native speakers of any language make regular use of various types of phraseological units on a large scale. It turns out that a great part of what we say and write is formulaic, i.e. it is not created from scratch by combining single words into larger units but based on prefabricated chunks (known as prefabs), which are multiple-word units of varying length, e.g. *and you know, no way, this is it*, stored by our brains in a ready-to-use form. Such phraseological units are typologically varied and include free collocations (semantically transparent combinations such as *play the piano* or *go home*), restricted collocations (largely transparent but with strict selection criteria for one of the elements, e.g. *throw a fit*), idioms (*spill the beans*), functionally specialised formulas (*best regards*) as well as the so-called lexical bundles, which are frequent word combinations which often do not form complete syntactic units, e.g. *I don't know if, so that they, and then he*. While we know quite a lot about the use and role of prefabricated language in contemporary languages, especially English, phraseological studies of historical languages are still rare. The main aim of our project to conduct a comprehensive diachronic analysis of English phraseology, and our research questions are the following:

1. what is the role of prefabrication in Old English (OE), Middle English (ME), Early Modern English (EME) and Late Modern English (LME)? are there any significant quantity and quality differences between the periods?
2. which types of phraseological units show the greatest dynamics and what are the reasons for their instability?
3. which specific phraseological units have been present in English throughout its history? what is their function in the language and which factors have contributed to their retention in the system?
4. which genres of English prose (medical texts, homilies, scientific essays, chronicle) have relied on prefabricated language to the greatest degree and has this tendency been stable in English?
5. which of the identified diachronic developments in English phraseology may have been triggered by extralinguistic factors (changes in society, cultural developments, historical events, new writing conventions), what is the mechanism of such an influence and what can this phenomenon tell us about the nature of phraseological change?

To this end, we are going to make use of modern corpus tools, including automatic lemmatizers based on NLP (natural language processing) methods, thanks to which it will be possible to align each wordform to its lemma, i.e. basic form (e.g. nominative singular for nouns or infinitive for verbs, in a standard spelling form). NLP algorithms, designed for the analysis of contemporary languages, after adequate training may be used to process historical data, and our team aims to take full advantage of this methodology. As a result, the lemmatization of corpora representing all periods in the history of English (between 1 and 2 million words per period) will allow us to create online dictionaries of collocations following the format of our VARIOE (<http://varioe.pelcra.pl/collocations>). It means that our project is a pioneering attempt at combining cutting-edge language technologies with the research agenda of historical linguistics, where studies are usually based on more traditional methods. Detected phraseological units will be subject to a multi-faceted analysis covering many aspects such as diachronic stability, function, contextual recurrence and genre specialization. Our goal is to identify the main quality and quantity differences between the periods and their interpretation in the context of social and cultural changes in the English-speaking community. The result will be a comprehensive and in-depth analysis of the dynamics of change in the area of English phraseology, allowing this discipline to gain adequate status within English historical linguistics.