Monika Czerepowicka
Poland

Multi-word expressions in Polish

The aim of the poster is to present how multi-word expressions are defined in Polish linguistics. The main problem concerns the identification of a multi-word unit. It is also connected to another theoretical problem: how to define a unit of language?

It is Stanisław Skorupka who was the first Polish linguist attempting to define the notion of multi-word expression in Polish. He claimed that the phraseological unit is a combination of at least two words which have an established meaning in usage.

Skorupka divides phraseological units into three groups:

- idioms (związki stałe)
- collocations (związki łączliwe)
- loose compositional units (Skorupka 1967-68).

There is also another classification of Polish multi-word units (Lewicki 1986). Lewicki uses formal criteria and distinguishes the following classes:

- expressions – that have a noun or an adjective as the head (pomroczność jasna 'black-out' or zła jak osa 'angry like a wasp');
- phrase – which head is a verb form (wciskać kit 'to bullshit sb', zbijać bąki 'to fool around');
- sentence – which is a sentence (SVO) or an equivalent of a sentence (każdy kij ma dwa końce 'it cuts both ways', człowiek człowiekowi wilkiem 'man is a wolf to his fellow man').

Looking for a satisfying definition of multi-word expressions in Polish we find the Bogusławski's theory of language units (Bogusławski 1976). According to his definition, the unit of language is a sequence of letters and spaces (or other spelling marks) which meets two criteria:

1. units of language are mutually disjoint – there is a formal and functional symmetry between them. This condition can be enclosed in a schema:

$$a, b, c, d - \text{sequences of diacritics}$$
$$ac : ad = bc : bd$$

To make it clear let's take an example:

a: ta 'this'   b: inna 'another'   c: książka 'book'   d: sztywna okładka 'stiff cover/ hardback'

ta książka (ac) : ta sztywna okładka (ad)   = inna książka (bc)   : inna sztywna okładka (bd)
'this book'   : 'this stiff cover'   = 'another book'   : 'another stiff cover'

2. at least one of segments (a, b, c, d) belongs to an open class. The open class is a collection whose elements can be characterized in a general way (not given by enumerating).

The Bogusławski's theory develops and influences another theory. M. Grochowski suggests to define a unit of language as follows:
 "A lexical unit is a sequence of diacritic elements that has global meaning, which means it is a sequence semantically indivisible into sub-sequences which are elements of a non-open class" (Grochowski 1982).

It seems that the last theory could be a starting point in finding a satisfying definition of a multi-word expression in natural language processing. However, because of technical reasons it is reasonable to exchange elements of non-open (close) class for distribution condition which seems to be more readable. Consequently, we may define the multi-word unit as a connection of at least two words (sequence of diacritics) with a global meaning and in which at least one of the segments appears in a specific distribution. On the one hand, it makes possible to cover a wider range of cases. On the other hand, a semantical element is not the dominant one.

Concerning the classifications of phraseological units, none of the mentioned above seems to be satisfactory. Therefore, as the starting point we can accept a morphosyntactic classification of lexemes to multi-words expressions, however, provided expanding it to syntactic criteria. Still this point needs extra research.

References:

Bogusławski A. 1976: O zasadach rejestracji jednostek języka, Poradnik Językowy, v. 8, p. 356–364.

Bogusławski A. 1978: Towards an Operational Grammar, Studia Semiotyczne.

Grochowski M. 1982: Zarys leksykologii i leksykografii. Toruń.

Lewicki A. M. 1986: Składnia związków frazeologicznych, Bulletin de la Societe Polonaise de Linguistique XL, p. 75–82.

Skorupka S. 1967-68: Słownik frazeologiczny języka polskiego, Wiedza Powszechna, Warszawa.