Multi-word expressions in Polish

MWE Classification in Polish Linguistics

a multi-word unit

- combination of at least two words which have an established meaning in usage (Skorupka 1976-68)

- according to different degree of connectivity:
  - idoms
  - collocations
  - loose compositional units

- according to grammatical criteria (Lewicki 1986):
  expressions - that have a noun or an adjective ia a head → *pomroczność jasna 'black-out', zły 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'*

a unit of language   (Bogusławski 1976)

sequence of letters and spaces (or other spelling marks) which meets two criteria:

- units of language are mutually disjoint; this can be enclosed to a schema:

  ac : ad = bc : bd

  a, b, c, d - sequence of diacrits
  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'*

- at least one of the segments (a, b, c, d) belongs to an open class which is a collection whose ellements can be characterized in a general way (not given by enumerating as in a close class)

a lexical unit   (Grochowski 1982)

is a sequence of diacritic elements that has global meaning, which means it is a sequence sem antically indivisible into sub-sequences which are elements of a non-open class

Multi-word expressions (MWEs)
by Baldwin and Kim (2010)

made up of multiple whitespace-delimited words (single-word MWE? - not in Polish!)

can be decomposed into multiple lexemes

display a few levels of idiomaticity:

lexical - at least one of the component is not a part of the conventional lexicon
*ciemek, *omacek in expressions: *po ciemku 'in the dark', po omacku 'blindfold'*

syntactic - syntax of MWE is not derived directly from the components
na domiar 'on top of it', do niczego 'good for nothing'

semantic - the meaning of MWE is not explicitly derived from the components
zbijać bąki 'to fool around', adwokat diabła 'devil's advocate', brać byka za rogi 'to take the bull by the horns'

pragmatic - MWE is associated with a fixed set of situations or a particular context
Dzień dobry! 'Good morning!'

statistical - a particular combinations of words occurs with markedly high frequency, relative to the component words or alternative phrasing of the same expression
w zanadrzu 'up the sleeve', bez liku 'plenty of', bez ustanku ' incessantly', po cichu 'quietly', po polsku 'Polish'

WG1

Monika Czerepowicka

Parseme Meeting
10-11 March 2014, Athens, Greece

References:
- Baldwin Timothy, Su Nam Kim (2010): *Handbook of Natural Language Processing*, Second Edition, CRC Press, Boca Raton, USA, p. 267-292.
- Bogusławski Andrzej (1976): *O zasadach rejestracji jednostek języka*, Poradnik Językowy, v. 8, p. 356–364.
- Grochowski Maciej (1982): *Zarys leksykologii i leksykografii*. Toruń.
- Lewicki Andrzej Maria (1986): *Składnia związków frazeologicznych*, Bulletin de la Societe Polonaise de Linguistique XL, p. 75–82.
- Skorupka Stanisław (1967-68): *Słownik frazeologiczny języka polskiego*, Wiedza Powszechna, Warszawa.

Types of MWE

nominal

1. **N + Adj** *pomroczność jasna 'black-out', kraina mlekiem i miodem płynąca 'land flowing with milk and honey'*
2. **Adj + N** *biały kruk 'rare avis'*
3. **N + N(Gen)** *adwokat diabła 'devil's advocate', metoda kija i marchewki 'carrot and stick method'*
4. **N + PrepN** *chłopcy z lasu, 'boys from the woods' (guerilla)*
5. **N + N(Nom)** *bieda-szyby 'bootleg mine'*

prepositional

preposition and a singular noun: *w zanadrzu 'up the sleeve', bez liku 'plenty of', w pobliżu 'in the vicinity', w zamian 'in exchange for', na jaw 'become known'*

complex prepositions: *ze względu na 'in view of'*

verbal

verb-noun constructions *zbijać bąki 'to fool around', wciskać kit 'to bullshit sb', odwracać kota ogonem '*
the noun complement tends to be fixed but verb inflected

→ main research problem, because verb inflection in Polish is complicated (some types of forms are excluded, for instance future tense forms, sometimes there are particular restrictions for particular verbs)

→ discontinuity problem - MWE occur in the joined (more often) or split word order (what kind of forms can split the MWE?)

→ word order problem - phrasal verbs and prepositional verbs are not supposed to be MWE in Polish (analytical government)

MWE Classification

Institutionalised Phrase   (statistically idiomatic)

w zanadrzu 'up the sleeve', do niczego 'good for nothing' , po polsku 'Polish' (prepositional expressions, syntactically adverbs)

Lexicalised Phrase

fixed - neither morphosyntactic variation nor internal modification
w zamian 'in return (for)', na wyrost 'over the top', bez liku 'plenty of'

semi-fixed - lexical variable, hard restriction on word order and composition (most of Polish MWE)
adwokat diabła 'devil's advocate', biały kruk 'rare avis', bieda-szyby 'bootleg mine', dawca obrączki 'wedding ring donor'

syntactically-flexible most of verbal MWE