

# Parameters of variation in the cross-linguistic analysis of phraseology

Koenraad Kuiper

## Abstract

The chapter proposes first that classifications should not be made in the case of multiword expressions (MWEs) but that they should be replaced by an empirically-determined set of properties which specify those areas of idiosyncrasy which would lead to a multiword sequence being lexicalised, i.e. find its way into a native speaker lexicon. Our reason for proposing this approach is that the properties conventionally used to classify MWEs cross classify and thus that the creation of classes of MWEs, is essentially futile.

Second we will examine all the areas of potential idiosyncrasy which MWEs might exhibit, i.e. the properties which lexical entries would have to list and which machine processes of parsing, production and translation need to take account of. These are examined as to whether they are digital or analogue, and more or less serious for machine applications. For example some anaphors have lexicalised antecedents as we shall suggest below and when they do, the antecedent is specified. This is thus a digital property. It is not serious for machine translation and parsing but it is for speech production since the algorithm would need to know this fact in each case. Semantic compositionality is an analogue property since the meanings of idioms can be more or less transparent. Machine approaches to speech have been seriously considering this property for some time.

The properties to be covered are as follows:

### Restricted Collocations

All MWEs are restricted collocations. Where the grammar and the semantics of an expression allow a number of words to function in a given syntactic position, in a restricted collocation only a subset is/are lexicalized, e.g. catch fire, #take fire, take flight, #catch flight.

### Polylexicalität

All MWEs have at least some constituents that are themselves lexemes. A lexicalized constituent of a MWE is one where the lexical content of the constituent is given in the lexical entry of the MWE, for example, in the imperative Shut up! both words are lexicalized constituents of the MWE. This criterion distinguishes some constructions which have no lexically specified items from MWEs which always do.

### Selection Sets

A MWE may have a lexicalized constituent which consists of a selection set where more than one constituent can be lexicalized in the same position (but the position is not open), e.g. in cross NP's palm/hand with silver, palm and hand function as alternatives.

### Slots

A MWE may have a slot(s). A slot is a position in the structure which requires to be 'filled' with other lexical items belonging to a particular constituent category but which is not lexically filled in the representation of the item in the lexicon. In most cases these are empty argument positions, e.g. in put NP into perspective, the NP is an obligatory complement of the verb which must be filled for the phrase to be used grammatically but the lexical content of the NP is not given in the

lexical entry of the MWE, i.e. is not a lexicalized constituent.

#### Slot restrictions

A MWE may have a slot restriction, an arbitrary constraint on what may fill a slot, e.g. in hold NP's liquor the NP has to be a human.

#### Optional constituents

A MWE may have an optional constituent. These are optional lexicalized constituents which may or may not be used, e.g. in catch NP's death (of cold), (of cold) is an optional constituent. This property is related to conventional deletability, e.g. keep your fingers crossed vs. fingers crossed.

#### Modifiability

A MWE may contain a lexicalized constituent which is modifiable, e.g. cut no ice, #cut no melting ice vs. cut NP's losses, cut NP's financial losses. This is sometimes attributed to the degree of compositionality of the expression and its constituents. The semantics of modification in MWEs is complex. Sometimes the modifier appears to have scope over the meaning of the whole MWE, at other times only of the constituent it would have scope over if the MWE were a freely generated expression, e.g. take careful note = 'take note carefully' c.f. cut NP's financial losses ≠ 'cut one's losses financially'.

#### Syntactic Structure

A MWE may have a syntactic structure which is flexible under movement, e.g. #NP's nose was cut off to spite NP's face vs the ground was cut out from under NP's feet. This property appears to be idiolectal and context dependent.

#### Lexical Antecedence

A MWE may have a pronominal with lexicalized antecedence where a pronoun in a MWE does not have its full range of possible antecedents, e.g. in dig NP's heels in the antecedent of the possessive must be the subject of dig. This phenomenon remains to be seriously studied.

#### Bound Words

Some MWEs contain bound words also termed cranberry words, (unikale Elementen), e.g. umbrage in remove wiggly underlining.

#### MWE Syntactic Constraint

There may be constraints on the phrase structure of MWE. While unlexicalized syntactic structures are potentially of infinite length, MWEs are finite and this may be because their syntax is restricted in various ways.

#### Idioms and Semantic Idiosyncrasy

MWEs may be semantically non-compositional. Such MWEs are normally termed idioms. There is a very large literature on this subject.

Koenraad Kuiper  
Professor Emeritus  
Linguistics Department  
School of Languages, Social and Political Sciences  
University of Canterbury  
Private Bag 4800  
Christchurch

and

Honorary Professor  
Department of Linguistics  
School of Letters, Arts, and Media

University of Sydney

NSW 2006