

How to create a MWE lexical entry

Aleksandar Petrovski

International Slavic University, Sveti Nikole
a.petrovski.sise@gmail.com

Katerina Zdravkova

University Sts Cyril and Methodius, Skopje
katerina.zdravkova@finki.ukim.mk

WG1

Classification presented by Baldwin and Kim has been used. MWEs are classified by three main properties: flexibility, idiomaticity and syntactic structure.

Property	Value	Code
Fixedness/Flexibility	Fixed	FixF
	Semi-fixed	SFixF
	Flexible	FlxF
Idiomaticity	Lexical	LexI
	Syntactic	SynI
	Semantic	SemI
	Pragmatic	PrgI
	Statistical	StatI
Syntactic structure	Nominal	NomS
	Verbal	VerS
	Adjectival	AdjS
	Adverbial	AdvS

This is an example of a verbal flexible MWE (eng. spread rumours). Inflectional class V2 will comprise all inflectional forms of the lexical entry. But, this MWE could appear in different word orders, e.g. гласини се шират – reflexive form, or гласини се ширени – passive form.

Internal modifications (тој шири опасни гласини eng. he spreads dangerous rumours) and extractions (какви гласини тој шири? eng. what kind of rumours does he spread?) are also possible. Combinations of e.g. internal modification and passivisation make the situation even worse.

Some of these forms might be captured by inflectional paradigms (e.g. passive forms), but generally speaking, using a NooJ MWE lexicon as a single resource is not possible in these cases.

The only way to recognise all forms of flexible MWEs is building syntactical grammars. One grammar should be built for each flexible MWE, which will comprise passivisation, internal modification, and extraction.

A possible lexical entry in a NooJ MWE lexicon would look like:

кабелска телевизија, N+FLX=N2 +SFxF+PrgI+Stat+NomS

where кабелска телевизија is a MWE (eng. cable television), N stands for noun - grammatical category of the MWE, N2 is the inflectional class code that describes the inflectional paradigm, SFxF, PrgI, Stat and NomS are properties' codes.

Since N2 describes the lexical entry's inflectional behaviour, NooJ is able to recognise all of its 9 inflectional forms.

NooJ formalism allows lexicalisation of fixed and semi-fixed types of MWEs, since all inflectional forms can be recognised.

But, what could be done with flexible MWEs? A possible lexicon entry of a flexible verbal MWE: шири гласини, V+FLX=V2+FlxF+SemI+VerS

MWEs with non verbal syntactic structures could also be flexible:

црн како јаглен, A+FLX=A2+FlxF+SemI+AdjS

This is an example of an adjectival MWE (eng. as black as coal). It is coded as flexible, since internal modifications are possible (црн како најцрн јаглен eng. as black as blackest coal).

Conclusions:

NooJ lexicons are very good when dealing with fixed and semi-flexed MWEs. For flexible MWEs, due to possible change of word order, lexicons can't help much. The only way is to build a syntactical grammar for each MWE.

It would be interesting to analyse how other NLP tools deal with flexible MWEs and to compare them with NooJ.