THE RESOURCE

IDION http://idion.ilsp.gr:
• a lexical DB
• encodes multiple aspects of information on MWEs: origins, semantics, morphosyntactic information, corpus examples, syntactic properties, diathesis alternation phenomena, lexical relations, usage examples
• addresses both the human user and the machine

THE ISSUE

We discuss the transcription of IDION contents to a lexical resource for parsing; as a case study we use an LFG/XLE lexicon that contains fixed sequences of fixed elements [4], known as Words With Spaces (WWS) [3], marked with underscores.

(1) Ανάβω ολα τα λαμπάκια
turn-on.1SG all.ACC the light-bulbs.ACC to smbd
‘I make someone furious.’

Syntactic tests such as word order permutations and XP interpolation show that the MWE in (1) contains the WWS ολα_τα_λαμπάκια (all the light-bulbs).

MORPHOSYNTACTIC INFO IN IDION IS (almost) THEORY NEUTRAL

A theory neutral morphosyntactic encoding is cost effective:
- it is user-friendly for a larger number of encoders
- it reduces the number of errors
- it uses standard morphological encoding (PAROLE)
- it uses syntactic notions of wide acceptance (rather than notions specific to a particular theory)

The price to pay for the theory neutral encoding is that IDION cannot be used as a lexical resource directly by any syntactic formalism; some work of syntactic interpretation/translation is left for an application that reads information off IDION and feeds the parser.

But this is the case for all encodings, whether bound to a theory or not.

Encoding in IDION
- provides a fully fledged morphological description
- provides a rather flat phrase description of MWEs where only free constituents may have a phrasal status (2) while fixed constituents are listed as words
- avoids to assign syntactic functional information, for instance whether the fixed parts have some syntactic function.

(2) The IDION vocabulary of phrasal category symbols:

NP-NOM/NP-NOM-anim/NP-NOM-nonanim
NP-GEN/NP-GEN-anim/NP-GEN-nonanim
NP-ACC/NP-ACC-anim/NP-ACC-nonanim
VP

Rules

(3) ολα_τα_λαμπάκια NWWS * (^ PRED) = ‘όλα_τα_λαμπάκια’

Step 2. PPs detection

(R2) If there is a LEMMA token marked as preposition followed by an NP-ACC/NP-GEN create a new phrasal constituent named PP which replaces the previous two and bears the constraint (^PFORM)=preposition_type’. (R2) yields (4).

(4) ο λαμπάκια NWWS * (^ PRED) = ‘ο λαμπάκια’

Step 3. Select Vhead and check for verb arguments

V heads are marked so (PAROLE tag). Using LEMMA, the algorithm incrementally creates the verb lexical entry (5).

(5) ανάβω ολοί (* PRED) = ανάβω...’

Step 4. Semantic information

Add semantics to the VbHead lexical entry: (^SEM)=’string_of_the_MWE_meaning’

REFERENCES


The XLE entries produced correspond to the idiomatic parts of the MWE while the non-idiomatic parts are received from the lexicon of the general language.

ανάβω ολοί (* PRED) = ανάβω...’

Step 5. Semantic information

Add semantics to the VbHead lexical entry: (^SEM)=’string_of_the_MWE_meaning’

Table 1. Morphosyntactic encoding of (1) in IDION