Morphological Lexicon

Fri, 16/05/2008 - 17:09 — webmaster

Characteristics

All Neurolingo's language tools are based, in one or another way, on the *Morphological Lexicon*. It contains 90.000 lexical units. Each lexical unit is accompanied by a set of morphological rules that produce all its wordforms. The 90.000 lexical units expand to 1.200.000 wordforms, each of which carries information about:

- Orthography: the sequence of letters that represent the wordform at the graphemic level
- Syllabification: the syllables that constitute the word form
- Morphology: the morphemes (prefix, stem, infix, suffix) that constitute the wordform
- Morphosyntax: the basic form (headword) and the morphosyntactic attributes (Part of Speech, Gender, Number, Case, Voice, Tense, Mood, Person, etc.) of the wordform
- Style: the style attributes, if any, of the wordform, e.g. the wordform *δίνανε* is "informal" compared to the morphosyntactically equivalent wordform *έδιναν* which is not.
- Domain: the domain(s), if any, of wordform's usage, e.g. the wordform *αβιογένεση* is used in Biology.

Apart from common words, the Morphological Lexicon also includes words of special vocabularies, e.g. 10.000 Greek toponyms (i.e. names of Greek counties, municipalities, districts, towns, villages, etc.), and will expand to cover domain-specific vocabularies. This expansion has been initiated with the incorporation of a sub-lexicon with (currently 6.000) biomedical terms.

Try the Morphological Lexicon online. [1]

Applications

The Morphological Lexicon is a language resource utilized by all Neurolingo's language tools. Specifically:

- Hyphenator [2]'s ability to handle the phenomenon of synizesis [3] is based on knowledge extracted from Morphological Lexicon's syllabification information.
- Speller [4]'s functionality is based on Morphological Lexicon's orthographic information.
- Lemmatizer [5]'s functionality is based on an index that contains all Morphological Lexicon's word-forms, so as to be able to normalize any word-form to the corresponding lexical unit.
- Thesaurus Browser [6]'s ability to handle the morphological variation of search terms in user's queries is based on an index that contains all wordforms of each Thesaurus lemma. Moreover, Thesaurus utilizes the morphosyntactic information contained in the Morphological Lexicon so as to return synonym/antonym wordforms that have the same morphosyntactic attributes as the search term.

Source URL: http://www.neurolingo.gr/en/technology/lexica/morpholexicon.jsp

Links:
[1] http://www.neurolingo.gr/en/online_tools/lexiscope.htm
[2] http://www.neurolingo.gr/en/technology/application_tools/hyphenator.jsp

©2005-2010 Neurolingo L.P.

[3] http://www.neurolingo.gr/en/technology/application_tools/hyphenator/synizesis.jsp
[4] http://www.neurolingo.gr/en/technology/application_tools/speller.jsp
[5] http://www.neurolingo.gr/en/technology/application_tools/lemmatizer.jsp
[6] http://www.neurolingo.gr/en/technology/application_tools/thesaurus_browser.jsp

©2005-2010 Neurolingo L.P.

W3C XHTML 1.0