

THE STATUS OF SEMANTICS IN MULTILEX

Nicole Modiano
modiano@capsogeti.fr

CAP GEMINI INNOVATION
86/90, rue Thiers
92513 Boulogne-Billancourt Cedex - France

(Position Paper)

From: AAAI Technical Report SS-93-02. Compilation copyright © 1993, AAAI (www.aaai.org). All rights reserved.

MULTILEX¹ is an ESPRIT project which is to define standards for multilingual and multifunctional dictionaries for the 9 languages of the EEC.

A key task is to identify an object which could play the role of anchoring to allow for equivalence relations over the languages.

Word forms cannot be used as basic units for translation because

- they are bound to specific language character sets and phonemic systems,
- they result from monolingual morphological mechanisms
- they are semantically ambiguous.

Meaning, in the contrary, is a discrete unit and it exists in all the languages.

The MULTILEX linguistic architecture clearly distinguishes the meanings from the signs of these meanings in specific written and oral codes.

Words are described at two levels :

1. The surface level in which orthographic, phonologic and morphological information is organized in distinct blocks of functionally coherent features. These blocks are called GPMUs (Grapho-Phono-Morphological Unit).
2. The semantic level in which the semantic and/or terminological, syntactic and cross-reference information is attached to the lexical unit. The lexical unit (LU) is defined as an identifier of a single meaning. The meaning of "meaning" can be specified by a semantic theory (if there is one available) or it can refer to the intuitive notion used by traditional lexicography to separate the meanings listed under canonical forms.

Bi-directional links relate surface level behaviours (GPMUs) and semantic units (LUs).

Uni-directional links relate an LU from a dictionary of language 1 (L1) to LUs in other dictionaries of language 2 (L2), language 3 (L3), etc. An LU of L1 can be related to one LU of L2 or to several LUs if this meaning of L1 corresponds to several meanings in L2, as in English the word "river" corresponds to two words in French, "rivière" and "fleuve".

The previous example shows one case of partial equivalence in which meanings of a given language cover "pieces of reality" which are not defined in the same way in other languages. Hence, the translation

equivalence in MULTILEX is labeled according to the degree and the kind of translation which can be established.

Two different approaches of semantic description have been proposed in the MULTILEX project:

1. Logically regimented explanation of ordinary language entries from existing dictionaries (developed at the University of Bochum, Germany).
2. Metalinguistic constructs based on a set of semantic features (developed at the University of Pisa, Italy).

Both these theories are compatible with the MULTILEX standard, and they provide valuable guidance to the lexicon builders. However, the objective is not to fix and impose a semantic approach to lexicon builders because:

1. The standard is seen as an open structure in which to express different views. Its aim is to help identify, clarify and compare pieces of lexical information in order to facilitate lexicon building, usage and re-use. Choosing one semantic theory would be to exclude all - or nearly all - the others.
2. At present, there is no available multilingual lexical resource large enough to demonstrate universal validity of a single semantic theory. There is still an enormous amount of work to do before some uncontroversial results are achieved.

Considering the present state of the art, MULTILEX has to face this paradox : meaning is the key object because of its anchoring function for the multilingual dictionary, but its formal definition is still to be clarified and strengthened. This is not to hinder development of dictionaries according to the MULTILEX standard if one considers that the LU is, above all, a functional object.

1. The partners of this project are Cap Gemini Innovation (France) and TA Triumph Adler (Germany) as coordinators and ASSTRIL, GETA (France), University of Bochum, University of Münster, Siemens Nixdorf Inc. (Germany), L-CUBE (Greece), Lexicon, University of Pisa (Italy), Philips, Vrije Universiteit Amsterdam (Netherlands), Siemens Nixdorf Inc. (Spain), University of Surrey, UMIST (Great Britain). The content of this paper is Cap Gemini Innovation's responsibility.