

Introducing the MuLexFoR : a Multilingual Lexeme Formation Rule Database

Cartoni Bruno

Many (bilingual or monolingual) dictionaries include morphological information in their list of entries, usually with the purpose of providing information about how to produce new words. In bilingual dictionaries, this kind of information is intended to help users to understand and coin new words in L2.

Representation of morphological processes in monolingual and bilingual dictionaries has often been criticised (Prcic 1999; Dardano, Frenguelli et al. 2006; ten Hacken, Abel et al. 2006; Cartoni 2008). All of these studies point out the inadequacy of using only affix representation, even though it is almost always the only possibility for including morphological elements into the dictionary list of entries. Main problematic issues come from the fact that many affixes participate in more than one word-formation process, and one affix can have more than one meaning.

In the lexematic approach to morphology (see (Fradin 2003) for the most recent studies in this framework), affixes are only the formal component of Lexeme-Formation Rules (hereafter LFRs), which contain other constructional operations (change of category, semantic function) and which, most importantly, are “semantically driven”.

In this paper, we present a project that aims to build an electronic database that gathers Lexeme-Formation processes in a multilingual perspective. The MuLexFoR database (Multilingual Lexeme-Formation Rule database) will provide a new solution for representing word-formation process in lexical resources.

On the lexicographic side¹, adopting the lexematic approach is very useful for building multilingual LFRs that match equivalent constructional processes from different languages, according to the meaning they coin. For example, in French and Italian there is one single LFR of “reiterativity” which builds verbs on verbal bases (LFR_reiter(v→v) and uses only one prefix (ri in Italian, re in French). The semantics of the rule is used as a “pivot” in the translation process, and consequently ri and re can be theoretically seen as the surface forms of one single LFR. Another advantage of this approach appears in cases where one single rule can have more than one affix. For example, the LFR for “unspecified plurality” contains three prefixes in both languages (IT: multi, pluri, poli; FR: multi, pluri, poly), although monolingual constraints can probably be applied to select the appropriate affix.

On the implementation side², the use of a multi-access and dynamic database allows the user to access morphological information through different modes and different languages. First of all, the user can browse the LFRs by semantic denomination (by

¹ Some interesting attempts to rationalise morphological information and to present it in a user-friendly way have already been implemented, but only in a monolingual perspective: E. Bernal’s DSVC, database for Catalan affixes (Bernal et DeCesaris 2008) is, to our knowledge, the first attempt to implement a rule-based approach. This project largely inspired the multilingual database presented in this article.

² The MuLexFoR database is implemented in PHP and will be soon available on the web.

choosing for example the “reiterativity rule”) thus showing the whole multilingual LFR, with the affixes and constraints present for that type in the different languages of the database. Obviously, this option requires a high level of competence in morphology, therefore, two other access modes are provided.

First, the user can choose the affix s/he wants to look for in the affix index and thus have access to all the rules in which it is involved, together with a complete description of each rule and the equivalent affixes in a target language, with their usage constraints. For example, if an Italian user wants to know how to express the prefixation in multi in French, s/he can browse the affix index and select the Italian prefix multi. MuLexFoR then provides all the rule(s) that involve(s) this Italian prefix, together with a complete description of each rule, where the user will find the French equivalent affixes (multi, pluri and poly), the constraints to use them, and some examples.

The other access mode is through the lexical index, which gives access to the rule that coined the chosen word. In the future, this lexical index can simply constitute a “bridge” between the constructed lexemes of a bilingual dictionary and the MuLexFoR database.

For the moment, the database contains more than 130 LFRs that imply the 54 most productive prefixes in French and Italian. The implementation of other languages and other word-formation elements is currently under work. Some evaluations of the use of this database are also planned. First results show some interesting issues about the coverage of the database and the information that are expected by the users.

References

- Bernal, E. et J. DeCesaris (2008) A Digital Dictionary of Catalan Derivational Affixes. Proceedings of *Euralex 2008*, Barcelone, Spain, IULA
- Cartoni, B. (2008) La place de la morphologie constructionnelle dans les dictionnaires bilingues: étude de cas. Proceedings of *Euralex 2008*, Barcelone, Spain, IULA: 813-820.
- Dardano, M., G. Frenguelli et G. Colella (2006) What Lexicographers Do with Word Formation. Proceedings of *Euralex 2006*, Torino: 1115-1127.
- Fradin, B. (2003) *Nouvelles approches en morphologie*. Paris, Puf.
- Prcic, T. (1999) The treatment of affixes in the "big four" EFL dictionaries. *International Journal of Lexicography* 12(4): 263-279.
- ten Hacken, P., A. Abel et J. Knapp (2006) Word Formation in an Electronic Learners' Dictionary', *International Journal of Lexicography*.