

Study on the LMF definition section

Gil Francopoulo INRIA Loria

15 July 2005

1 Introduction

This document is an auxiliary working document on LMF.

The goal is to improve the content of the definition section by the means of FDIS 704. More precisely, the objectives are to detect the important groupings, the inconsistencies, the isolated concepts and the missing concepts.

This study is based on the LMF revision-7 dated from 14 July 2005. This is an internal version between the editors.

2 Conventions

FDIS 704 being based on its proper diagram notation, UML will be used instead.

By convention, in order to focus on important aspects, among synonyms, only one string will be used. This is the string that holds the definition in the LMF document.

3 Important groupings

We can classify the concepts in the follows sets:

- Concepts around form
- Concepts around word
- Concepts around data category
- Other concepts that are not connected.

Let's note that the concepts around form and the concepts around word are connected.

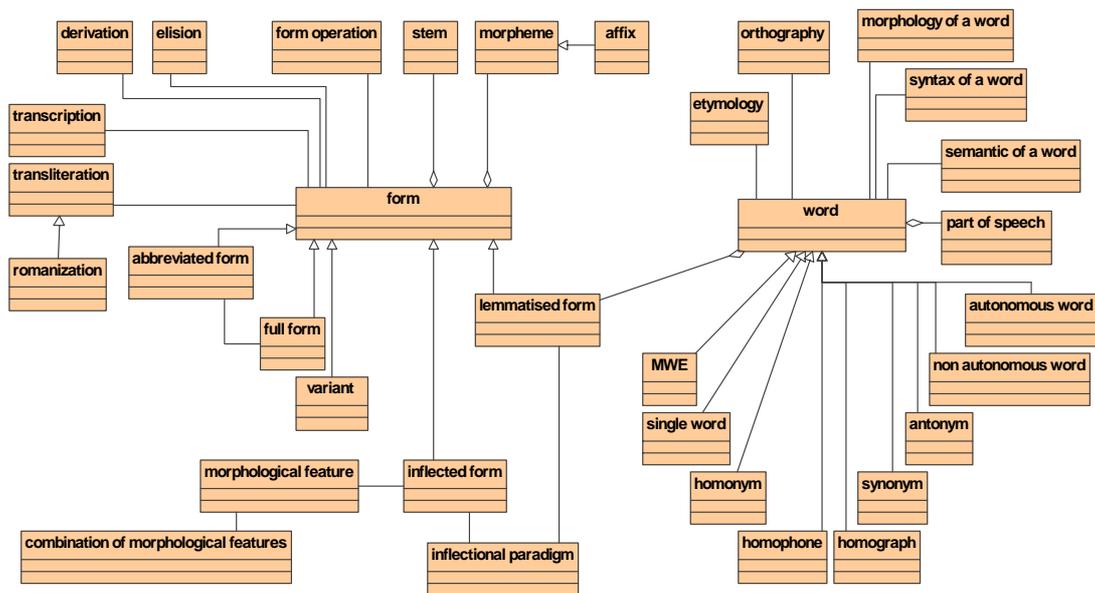
The **concepts around form** can be subcategorized as follows:

- Subclasses of form (e.g. full form)
- Components of form (e.g. morpheme)
- Operations on form (e.g. form operation)
- Result of operations (e.g. transcription)

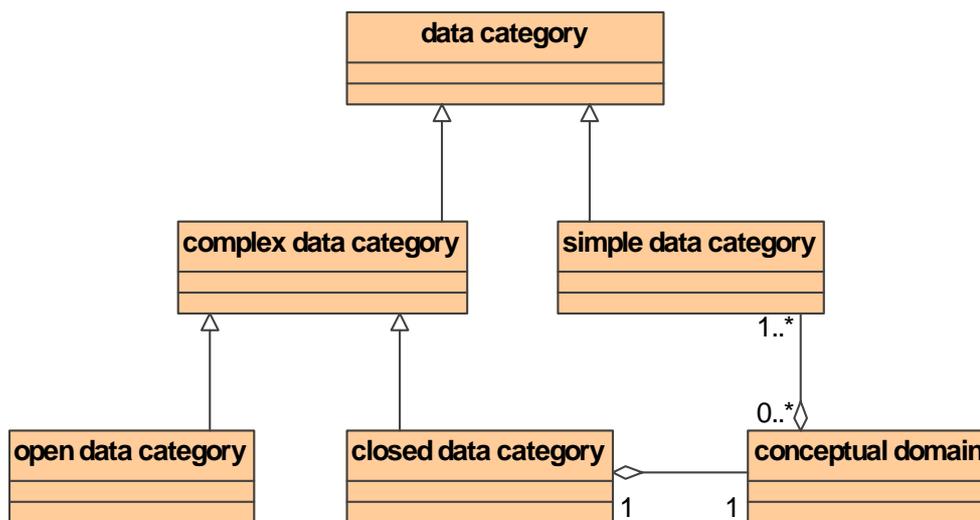
The **concepts around word** can be subcategorized as follows:

- Subclasses of word (e.g. MWE)
- Components of word (e.g. part of speech)
- Properties of a word (e.g. syntax of a word)

The diagram of these two concepts is as follows:



The other grouping is around data category and is as follows:



4 Isolated concepts

Some concepts are not connected. The list is as follows:

- Automaton
- Circonstant
- Citation
- Collocation
- Compiled program

- Electronic lexical resource
- Electronic terminological resource
- Ergative verb
- HLT
- Interlingua
- Intransitive verb
- Machine translation lexicons
- NLP
- Object language
- Splitting conditions
- Subcategorization frame
- Support verb
- Syntactic actant
- Transitive verb
- Usage note
- Word frequency

5 Missing concepts within the definition section

We can notice that the following concepts are not defined:

- Language (e.g. object language)
- Lexical resource (e.g. for machine translation resource)
- Verb (e.g. for transitive verb)
- Database collection (e.g. for ELR)

These concepts are used in definitions but are not defined within the definition section. These concepts are rather specialized: they are not part of the general English nomenclature.

One important concept is not defined. This is "lexicon" that is used a little bit everywhere.

6 Missing concepts regarding to the rest of the document

MRD is not defined.

7 Inconsistencies

The definition for “script” is rather fuzzy. It has not been possible to create a box for this concept.

8 Possible improvements for grouping

If “verb” were defined a certain amount of concepts could be connected like “transitive verb”, “support verb” etc.