

Deliverable 3.2B**Interim Report: CD of morpho-syntactic annotation standard
for ISO DIS ballot**

Project reference number	e-Content-22236-LIRICS
Project acronym	LIRICS
Project full title	Linguistic Infrastructure for Interoperable Resource and Systems
Project contact point	Laurent Romary, INRIA-Loria 615, rue du jardin botanique BP101. 54602 Villers lès Nancy (France) romary@loria.fr
Project web site	http://lirics.loria.fr
EC project officer	Erwin Valentini
Document title	Interim Report: CD of morpho-syntactic annotation standard for ISO DIS ballot
Deliverable ID	3.2B
Document type	Report
Dissemination level	Confidential
Contractual date of delivery	M21
Actual date of delivery	09.02.2006
Status & version	Draft, modified from report to interim report
Work package, task & deliverable responsible	DFKI
Author(s) & affiliation(s)	Thierry Declerck, Mirjam Kessler, Ulrich Krieger and Bernd Kiefer (DFKI)
Additional contributor(s)	Eric de la Clergerie (INRIA, Editor of the MAF document), and many experts from national standardisation bodies and ISO.
Keywords	Morpho-Syntax, Annotation, Standards, Tree-Banks

Document evolution

Version	date	version	date
0.9	30 th Nov. 2007		
1.0	31st Jan 2007		

Introduction

We presented in a former version of this deliverable (D3.2.A) the version of MAF (Morpho-Syntactic Annotation Framework) that has been submitted for an ISO CD ballot. Deliverable D3.2.B should be a report on the version of MAF to be submitted to the DIS ballot. As will be explained just below, some delays have to be observed with respect to the submission to the DIS ballot. This is the reason why we decided to modify the title of D3.2.B to “Interim report”, since we do not have for the time being the final version of MAF to be submitted to a DIS ballot. The editor of MAF being also not a member of the LIRICS consortium, we can not have direct influence of the date of the submission, but we are in regular contact, so that we know about the actual state and the work on other projects building partly on the results of MAF, like SynAF, can go on

Content:

**1 STATUS OF THE MORPHO-SYNTACTIC ANNOTATION FRAMEWORK (MAF), IN
THE PRE DIS STADIUM5**

2 MINUTES OF THE MEETING IN BEIJING (21-26 AUGUST 2006)6

1 Status of the Morpho-syntactic Annotation Framework (MAF), in the pre DIS stadium

As a summary of the previous deliverable, we would like to recorder that the general view on the MAF CD was positive, concerning the model, but negative concerning some linguistic issues, which were considered as not being enough elaborated and there was the request to revise a certain number of definitions. Many comments were issued on this and the editor of MAF decided to revise substantially the document.

At the plenary session of TC37 in Beijing (21-26 August 2006), further discussions with international experts, representing the ISO related national standardization bodies, took places and additionally to the comments already provided in the first part of the year, we identified the need to harmonize the terminology in use on the various ISO projects in SC4 (LAF, LMF, MAF, SynAF, SemAF...).

Also the editors of LMF decided to go for a new version of their document, at the DIS level, although it had been accepted by a majority of the standardisation bodies, in order to include some comments and also to go for the harmonization of terminology. The new deadlines for LMF have been set to November 2006 (LMF has been submitted last year on timeto the DIS ballot). MAF (and other projects) clearly had to wait for this new version of LMF in order to be completed.

In this version D.3.2.B, we restrict therefore ourselves to a very short interim report, explaining the status of MAF, and in the next section, we just add the minutes of the Beijing meeting on MAF. We will deliver the full version of D3.2.B as soon as we have the new version of MAF, to be submitted as a DIS ballot.

For more details on MAF we would like to refer to deliverable D3.2.A, in which the whole text of MAF, at the CD level, is included.

2 Minutes of the Meeting in Beijing (21-26 August 2006)

In the following we just include the minutes of the Beijing meeting dedicated to MAF, as an indication of the work that should still be done in MAF in the second half of 2006 (we expect a new version of MAF in March 2007):

- 1) MAF is at a pre DIS level (2 negative votes at CD ballot)
- 2) MAF Editor(s) should submit a form with the responses to the comments by experts during CD-Ballot, to be registered as DIS. Before end of year.
- 3) Preparation of at least for one tagset (for MULTEXT), as an instance of the meta-model, for the FDIS document
- 4) Scope of MAF.
 - A main comment: The full set of data categories is not provided by MAF, but MAF links to data categories (TDG2). MAF is basically a meta-model. The real standard is the meta-model, the data categories are to be included in a registry, evolving over time.
 - Meta-model also available as a UML diagram
 - Add annotated examples, present a tagset
 - LMF to be added to the normative references (or at some other relevant place in the MAF document). Relation to LMF also because of the relation of MAF to lexical entries.
- 5) Respond to the main comments, which are concerning the terminology. Cooperate with other SC4 projects on the terminology. Some comments are on the correctness of given examples. Provide for new examples, also in new languages. Contact experts for providing examples.
 - Possible relation to LMF for accessing lexical entries should be made explicit in the MAF document.
 - Reaction to the comment on empty word forms. Remove the empty string from MAF in case of “PRO”, but keep the possibility of having empty wordforms. Empty WF for “PRO” and similar phenomena (not realized subjects in some romance languages) to be described in SynAF.
 - Give more examples on the use of MAF for a wider set of documents (example of phonetic transcripts)
 - Give a differentiating definition of morphology and morpho-syntax. (3.21) and delete the corresponding NOTE.
 - Improve the definitions mentioned in the comments of the experts.
 - Given account of wordforms with possibly no lexemes available (Neologisms, date expressions like 2006-08-23).
 - Take a decision about the class of MWE MAF is dealing with. Introduce a definition for “expression”. Ask if some MWE are to be dealt at SynAF level.
 - Treatment of Affixes in MAF (we may consider them as possible PoS, and so as lexemes).
 - Check the coherence and/or consistency of definitions in LMF and MAF.
- 6) Resolution: Form an ad hoc coordination committee within SC4 consisting of PLs of MAF, LMF, SynAF, LAF etc. Identify a native speaker of English for proof reading.

