

LIRICS

Linguistic Infrastructure for Interoperable Resources and Systems

▶ **WP3**

▶ **Presented by Thierry Declerck (DFKI GmbH,
Saarbrücken, Germany)**

WP3: Overview

- Duration: M3 – M30
- Title: Morpho-Syntactic and Syntactic Annotations
- Partners: DFKI, INRIA, UFSD, CNR-ILC, UW, UTiL, IULA-UPF

WP3: Objectives

- A report on emerging morpho-syntactic and syntactic standards. Their strengths and weaknesses
- A Morpho-syntactic annotation meta-model standard, including a Data Category Selection (DCS) standard as an additional part to the 12620 series
- A Syntactic annotation meta-model standard, including a Data Category Selection
- Test suites for morpho-syntactic and syntactic annotation, that will build a small reference corpus of morpho-syntactically and syntactically annotated text and dialogues

WP3: Strategies

- Look first at current standardisation initiatives (Eagles, Multext-East) well known annotation strategies (TreeBanks), on the basis of which more abstract models of morpho-syntactic and syntactic annotation will be described.
- Interleaved work with ISO TC37/SC4 initiatives on Morpho-Syntax (mainly the Morpho-Syntactic Annotation Framework, MAF). Extend this work also to Syntax.

WP3: Main issues in the Standardization Work

- Interaction with Lexical MarkUp Frame (WP2)
 - * referring lexicon entries in MAF annotations
 - * being coherent on representing morphological content
 - * sharing common terminology
 - * sharing common Tag Sets
- Segmentation issues
 - * Asian languages (not central in LIRICS)
 - * Difficult phenomena in some languages (compounding, agglutination, ...)

WP3: Main issues in the Standardization Work

- Interaction with Data Category Registry (DCR), transversal to WP2,3 and 4 of LIRICS, managed in WP1)
 - * capturing and/or defining data categories in Tag Sets
 - * extending current data category registry with MAF terminology

WP3: Main issues in the Standardization Work

- **Extend MAF to Syntax and Parsing**
 - * requirements of Parsing community about MAF seen as input data for parsers (Tree Bank / Dependency Banks)
- **Interaction with WP4 on semantic content**
 - * Differentiate purely syntactic constituents, that can bear particular semantic content, from the semantic content itself. Interface syntax/semantic.

WP3: Main issues in the Standardization Work

- Interaction with related ISO initiatives, like for example the TC37/SC4 committee on Feature Structures (FSR/FSD)
 - * defining Tag Sets with FS libraries and with Typed Feature Declarations

WP3: Expected Risks

- Mainly the risks that are inherent with ISO initiatives: To get negative feedback from experts from a critical number of countries => negative ballots.
- Due to the number of experts involved in LIRICS and their actual work within national standardisation bodies, we expect this risk to be quite low.

WP3: Expected results (1)

- A report on current and emerging standards for morpho-syntax and syntax (M9)
- WD of morpho-syntactic annotation standard for CD ballot (M12)
- First selection of morpho-syntactically annotated samples for test suites no conformity with WD required (M15)
- CD of morpho-syntactic annotation standard for internal quality assessment (M18)
- CD of morpho-syntactic annotation standard for ISO DIS ballot (M21)

WP3: Expected results (2)

- WD of syntactic annotation standard for CD ballot (M18)
- First selection of syntactically annotated samples for test suites no conformity with WD required (M21)
- CD syntactically annotated standard for internal quality assessment (M24)
- CD syntactically annotated standard for ISO DIS ballot (M27)
- Final test suites of ISO conformant morph-syntactic and syntactic annotation (M30)

WP3: Work Plan for the next 3 Months

- **Start working on management and dissemination issues**
 - **Participation to the Fourth Meeting of the joint ACL/SIGSEM-ISO/TC 37/SC 4 “Working Group on the Representation of Multimodal Semantic” (10-11 Jan., 2005 in Tilburg).**
 - **Disseminate the LIRICS work at the DIN mirroring committee (already a paper on LIRICS related DIN activities submitted to the national GLDV conference in Mai 2005)**
 - **Preparation of participation at the Eighth International Open Forum on Metadata Registries (Open Forum 2005, <http://www.berlinopenforum.de/>), Berlin, Germany, 11 - 14 April 2005. And the TC 37-SC4 week in Poland in August 22-26.**