

LIRICS WP3 – Morpho-Syntactic and Syntactic Annotations

Thierry Declerck

declerck@dfki.de

DFKI GmbH

+ *LIRICS partners*

8th October 2007

Summary

- ✦ Main Objectives
- ✦ Overview of WP3 activities
- ✦ External links
- ✦ Outstanding results
 - ✦ Deliverables
 - ✦ Synergies with other WPs, ISO meetings
- ✦ Dissemination and Impact of WP3
- ✦ Future activities

WP3 - main 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 - overview of activities

- ✦ Accepted submission of MAF for ISO CD Ballot. Submission for DIS Ballot expected end of 2007
- ✦ Submission of SynAF for ISO CD Ballot, September 2007. New submission for Ballot expected in March 2008
- ✦ DCS for MAF Annotation available. DCS for SynAF annotation proposal being discussed, within ISO TC37/SC4 TDG 4.
- ✦ Embedding of SynAF and MAF annotation meta-models in the on-going LAF (Linguistic Annotation Framework). A graph model supporting multi-layered (linguistic) annotation
- ✦ In the second period of LIRICS, we focussed on a close collaboration with the ISO WD SemAF.

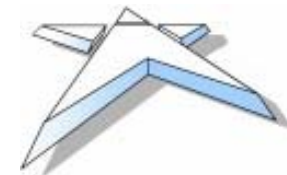
Relations with external bodies and initiatives

- ✦ DFKI leading member of DIN
- ✦ ISO TC 37/SC 4
- ✦ ISO TC 37/SC 4/TDG2 **Morphosyntax**
- ✦ ISO TC 37/SC 4/TDG4 **Syntax**
- ✦ ISO TC 37/SC 4/TDG3 **Semantic**

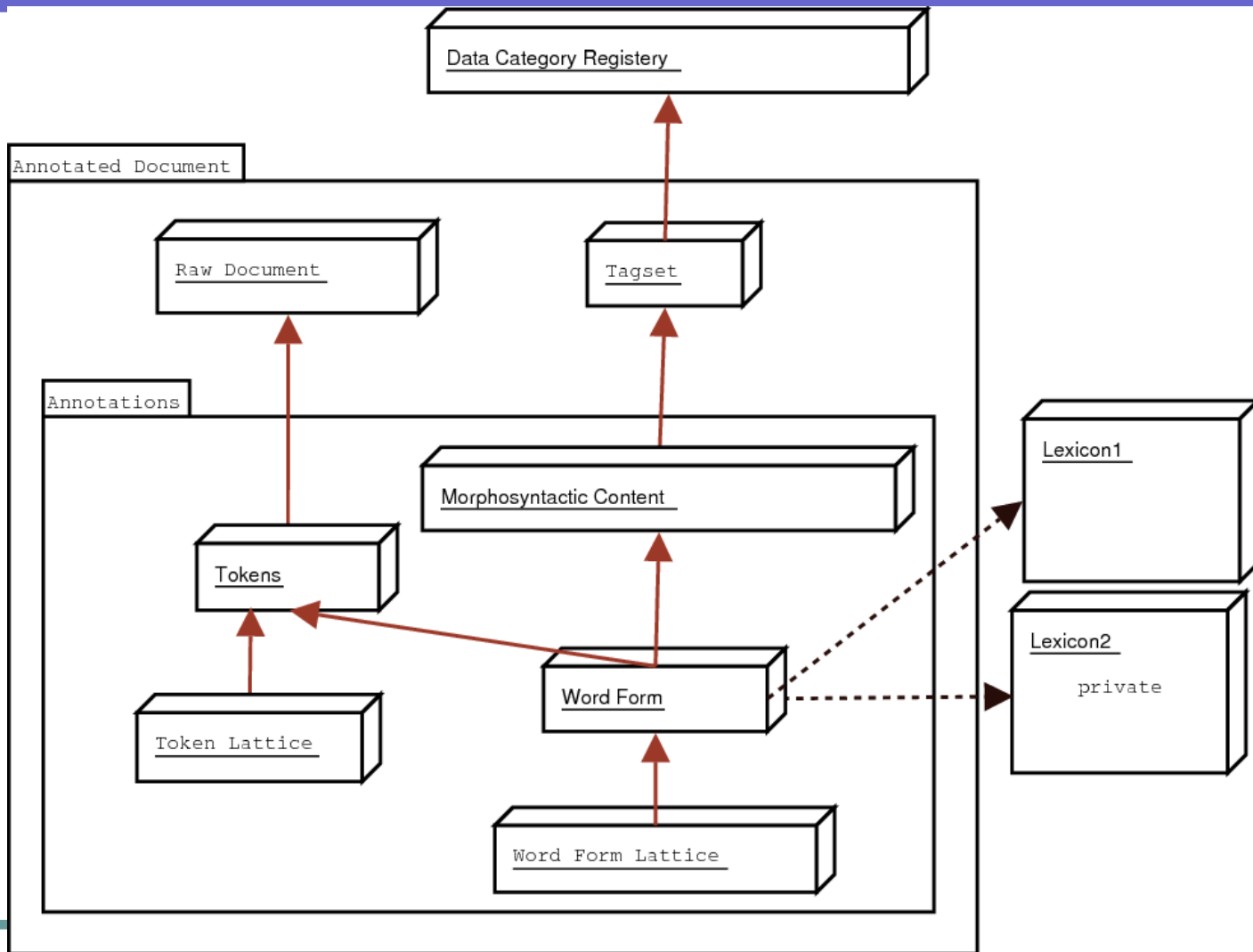
- ✦ Language-Grid (NICT, Japan)
 - ✦ Small additional funding for including Lirics/ISO works in Language Grid project

- ✦ W3C (joint Workshop in April 2007)

- ✦ OMG (on representation of language resources in ontologies), with the SIG: Semantics of Business Vocabulary and Business Rules (SBVR)



WP3- results: MAF Metamodel



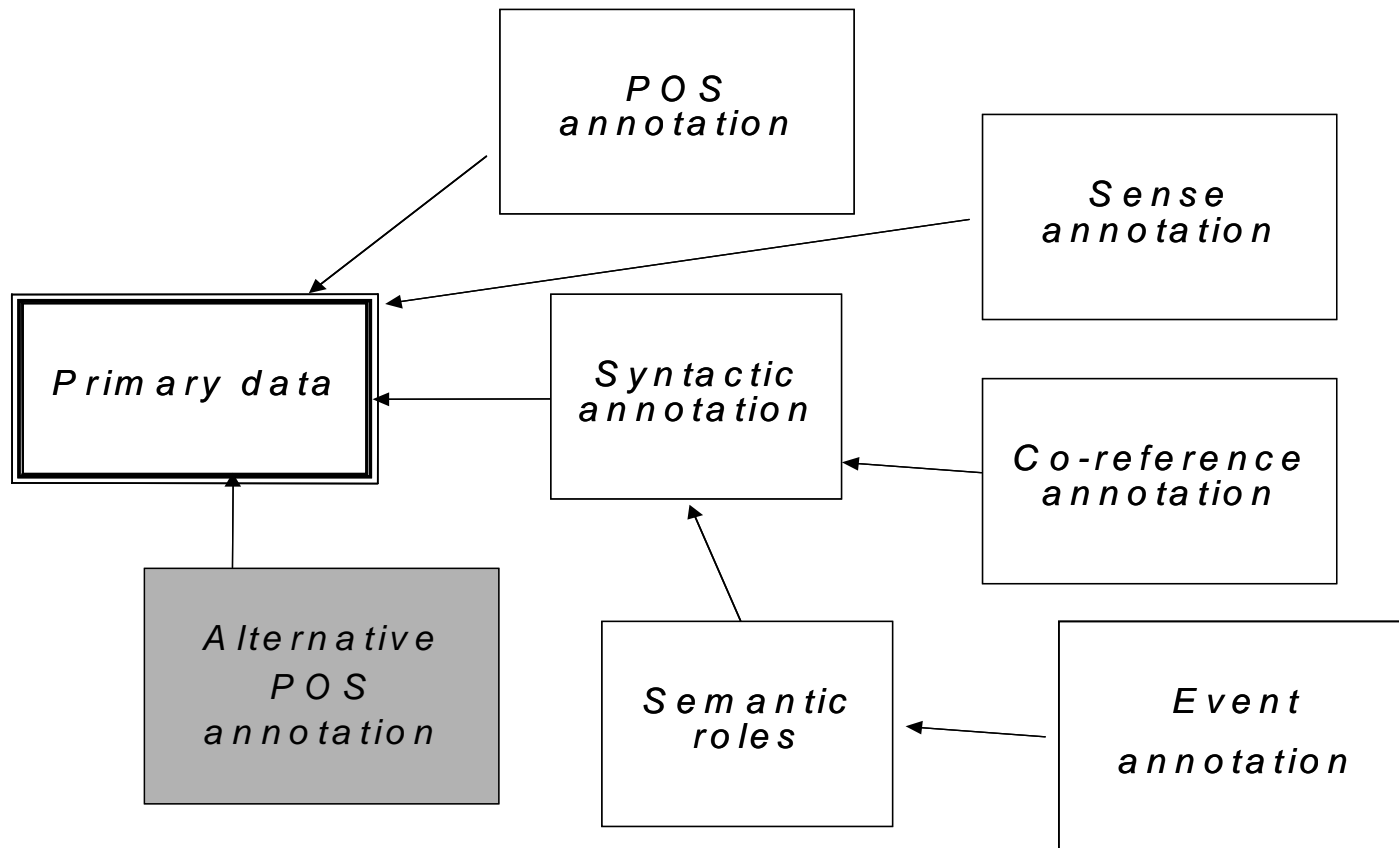
WP3- results: MAF DatCats

- ✦ Developed within LIRICS, in a cooperative effort between WP2 and WP3.
- ✦ References are in the syntax tools at Loria, to be ported to the new DCR implementation at MPI.
- ✦ Included in the MAF testsuites built by LIRICS partners on 6 languages
 - ✦ Examples: indefiniteDeterminer, subordinatingConjunction, fusedPreposition, declarativePunctuation, infinitive etc.

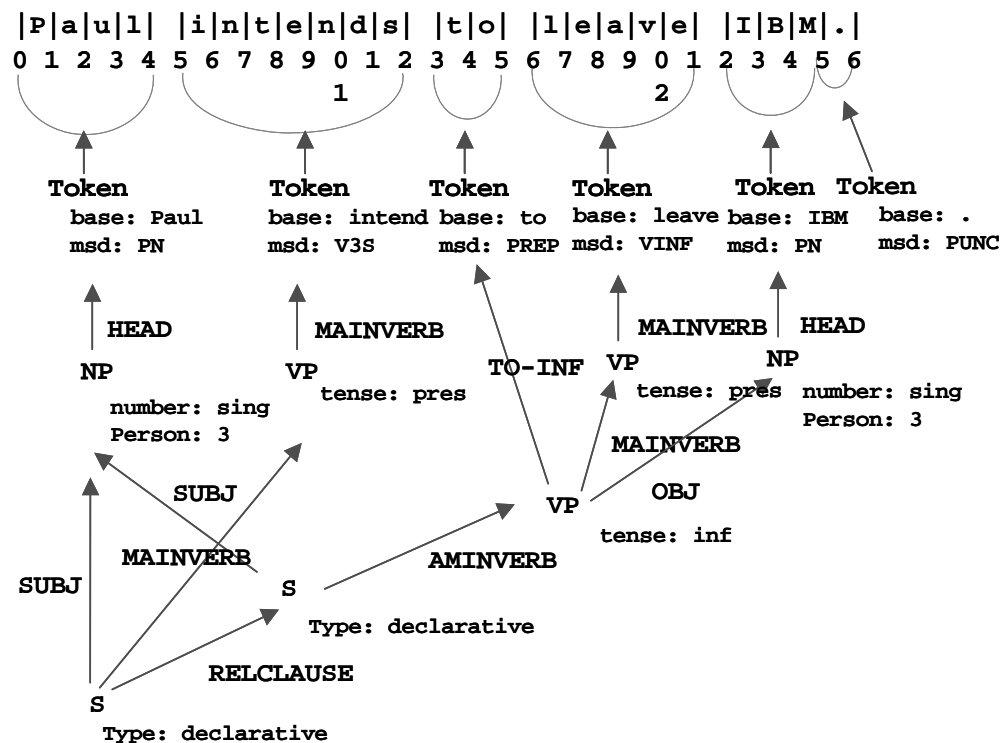
WP 3 - results: SynAF DatCats

- ✦ In development within the ISO TC37/SC4 /TDG4: **Syntactic profile** (*leader: Thierry Declerck*)
 - ✦ To be included as informal annex to the submission of SynAF to the DIS Ballot
 - ✦ Examples: nominalPhrase, adjectivePhrase, subordinatedClause, sentence, subject, directObject, modifier, specifier, head

WP 3 – Cooperation with LAF, Multi-LayerAnnotation



WP 3 – Cooperation with LAF, Multi-LayerAnnotation, Example



Segmentation

Morpho-syntactic layer

Syntactic layer

WP3 – results: Deliverables

- ✦ D3.1: Evaluation of initiatives for morpho-syntactic and syntactic annotation
- ✦ D3.2A: Interim Report: WD of morpho-syntactic annotation standard for CD ballot
- ✦ D3.2B: CD of morpho-syntactic annotation standard for ISO DIS ballot
- ✦ D3.3A: Interim Report: WD of syntactic annotation standard CD ballot
- ✦ D3.3B: CD of syntactic annotation standard for ISO DIS ballot
- ✦ D3.4: Test suite of ISO-conformant morph-syntactic and syntactic annotation.

WP3 – results: Dissemination

- ✦ Many talks, among them some invited talks, like at the 1st Symposium on Universal Communication June 14 (Thu) - 15 (Fri), 2007, Kyoto, Japan, and the German Workshop on Diachronic Linguistics (June 2007)
- ✦ Organization of the „The First Workshop on Automated Syntactic Annotations for Interoperable Language Resources“, which will be held at the First International Conference on Interoperability for Language Resources, City University of Hong Kong, Hong Kong, 9-11 January 2008. SynAF is a building block of this workshop

Future Work

- ✦ Achieve the next (mostly formal) steps in the standardization procedures of MAF (FDIS) and SynAF (DIS and FDIS)
- ✦ Cooperation with the Language Grid Initiative
- ✦ Include MAF and SynAF initiatives for the representation of language resources in Ontologies (already on the way for LMF)
- ✦ Use wide base of standardised syntactic annotations for supporting the automated building of semantic resources
- ✦ Cooperate with Multimedia Semantics Annotation initiatives (like in the NoE „K-Space“)