

LIRICS

Linguistic Infrastructure for Interoperable Resources and Systems

▶ **MAF-SynAF**

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

MAF

Morpho-Syntactic Annotations provide an important layer of linguistic information to a document. Large amount of corpora have been and are still manually annotated, while more and more annotations are now automatically produced by linguistic tools. Many NLP tasks (such as terminology extraction, information extraction, parsing, . . .) rely on these morpho-syntactic annotations

MAF: Goals

Towards the very ambitious goal of providing for a (unique) tagset organizing morpho-syntactic contents for all human language, MAF is more modestly dedicated to explore and propose a generic way to anchor, structure and organize annotations completed by mechanisms to specify comparable tagsets and annotation contents.

MAF: Format

As many recent standardization proposals, we favor the use of XML representations, because they ensure both human readability and easier machine processing. Still, these XML representations should rely on some consistent XML-independent model. In our case, the ISO proposal on feature structures for language data.

MAF: Format

As many recent standardization proposals, we favor the use of XML representations, because they ensure both human readability and easier machine processing. Still, these XML representations should rely on some consistent XML-independent model. In our case, the ISO proposal on feature structures for language data.

MAF: Terminology

The terminology or set of *categories* (types, features, and feature values) used in tagsets are described w.r.t. *registered data categories* whose meaning has been clearly stated. Feature structures and registered data categories provide a promising direction to build tagsets that may be automatically compared, even if only approximatively.

MAF: Terminology

Basic Units (words, tokens), problem of segmentation
Ambiguities: morphological, lexical ?how to represent them?

MAF: Metadata

Metadata are needed, for instance, for specifying the author (or tool) of a set of annotations, the date, the confidence,

.However, we do not plan to provide a specific mechanism to handle metadata but rather to rely on other proposals (IMDI probably a good candidate).

SynAF: A new work item

See the document

Dependency Structures seen as very important (both for Semantic Web applications as well as for Multimedia applications)

Call for contribution