Curation Problems in Large Diverse Data Collections – The Case of the DoBeS Annotations

Alexander König, Sebastian Drude

Max Planck Institute for Psycholinguistics

The Language Archive: Tools, Projects, Language Data, Experts, Collaboration

Example: How to search for a pronoun across DoBeS corpora

Before (current state)

User has to enter complex search string himself to account for different notations

SEARCH FOR pron IN ps OR pn IN pos
OR pronoun IN part_of_speech OR ...

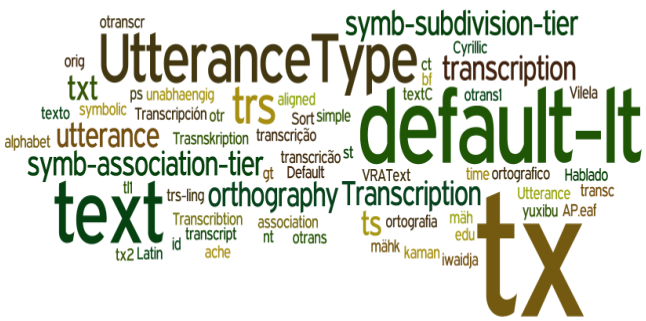
After (desired state)

Automatic translation to different notations via ISOcat DCR

SEARCH FOR pn IN pos

SEARCH FOR pron IN ps OR pn IN pos
OR pronoun IN part_of_speech OR ...

Example: Diversity of tiers of the type *transcription*



More information about

The DoBeS project: http://www.mpi.nl/dobes
ISOcat: http://www.isocat.org/
ELAN: http://tla.mpi.nl/tools/tla-tools/elan/

Contact & More Information:
Alexander.Koenig@mpi.nl
http://tla.mpi.nl/