The "Digital Grammar" Project

Integrating the Wiki/CMS approach with Language Archiving Technology and TEI

Sebastian.Drude@gmail.com

1. Background

- Project (submitted to German DFG)
- Context: "A discourse-based multimedia grammar of Aweti" (Dilthey fellowship)
- Wider context: DOBES-programme
- "Digital Grammars" should be an authoring system, useful to descriptive linguists, typologists etc.
- Community-based project
- Building on previous relevant work (some being presented in this colloquium)

2. "Digital Grammars"

Key features:
- Descriptive texts as digital documents
- Different from office-software or PDFs
- True hypertext documents with individual but interlinked pages
- Logical/content mark-up with functionalities instead of visual formatting
- Questions: Authoring/editing software?
  - Archiving, working, display data formats?
  - Different uses (humans / machine)?

3. Language Archiving Technology

- Developed at the Max-Planck-Institute for Psycholinguistics since ca. 2000
- Context: language documentation (DOBES)
- Language corpora with IMDI-metadata
- ELAN and ANNEX for creating / viewing annotated multimedia-data
- LEXUS for online lexical databases
- ISOcat data catalogue for linguistic terms
- No component for scientific meta-texts such as typological work or grammars

4. The Wiki/CMS approach

Content Management Systems & Wikis:
- Online-collaboration
- User-management
- Version control
- Updates etc. are developed by others

Challenges:
- Serial & hierarchical ordering of pages
- Specialized markup (extendable)
- Special functionalities

2011-02-12
4. The Wiki/CMS approach

5. Text Encoding Initiative (XML)

- XML promises to be a lasting standard
- Human and machine readable
- The TEI "recommendations" build on XML
- TEI-XML should be one format (archival, interchange) of the digital grammar

**Challenges:**
- How to enter (XML) text & markup?
- No specific TEI module for linguistic descriptive / typological work yet

6. Versioning and publication

- DGs are not static but "living documents"
- Version control (automatic in a Wiki/CMS)
- Printing the grammar should be possible
- and producing digital "snapshot" distributions
- Versions and distributions must be citable
- Ideally, it would be possible to work on an offline copy (e.g., in the field)
- This posits complex issues of synchronization
- and, again, the question of a suitable editor

7. Conclusion

- The project needs first to decide on some basic questions
- ... and to establish realistic goals and priorities
- This is work which needs input from and exchange with a community of experts
- Group and individual meetings planned
- I hope that this colloquium helps to form such a community
- Comments are welcome! Thank you!