

OLIF2 Consortium Review Meeting

April 4, 2001
SAP
Walldorf, Germany

MINUTES

The meeting came to order at 9:15 AM.

Present were:

Gregor Thurmair, Sail Labs
Mike Dillinger, Logos
Jean Senellart, Systran
Carlo Mergen, EC
Michael Wetzel, Trados
Daniel Grasmick, SAP
Susan McCormick, SAP
Christian Lieske, SAP
Jennifer Brundage, SAP
Matthew Chermiside, SAP
V. Srinivasan, SAP
Joerg Schuetz, IAI
Hubert Lehmann, Linguattec

- Daniel welcomed the participants; Susan reviewed the agenda and stated that the purpose of the meeting was to:
 - review OLIF v.2 DTD proposal as published on www.olif.net
 - discuss testing, implementation, certification plans for v.2
 - discuss future direction for the OLIF2 consortium

- Christian reviewed the current proposal for the OLIF2 DTD, as well as the supporting documentation:
 - Choice of DTD rather than schema – easier to generate schema from DTD, rather than other way around
 - Can generate OLIF entries with EMACS structure editor; contact Christian for help
 - Presented graphical view of DTD (see ppt slides for specifics) – participants had following comments/questions:
 - Concern about length of tag names for processing purposes – Srini pointed out that numbers can be used for processing and then replaced
 - Question on use of entryID vs. monoId – possible overkill, but may want to use both for search and organizational purposes
 - Gregor brought up question of possibility of overgeneration of features – consensus that this should be an issue for OLIF validation
 - Hubert raised issue of referent gender; Susan will check
 - Point made that, for user extensions, users can attach hierarchies as well as lists
 - Mike D. asked whether DCS's are DTDs – Christian said that this is still open, could be any type of file.
 - Question on workflow information – needs to be defined, can contain information on how DCS info is to be processed
 - Question on replacement specification in header – can replace only values? Or tag names?
 - General discussion on whether certain header features should be obligatory or not; decision to leave them as they are now specified
 - Note that DCS's should have an overall view and XML format
 - Question on whether current list of possible DCS's should be extensible – point that current list simply reflects items we couldn't easily decide values for

- Mike D. pointed out that the Part-of-speech DCS may be unnecessary – is already a defined set of values and can be extended with user extension, rather than replaced with DCS
- Question from Joerg on whether DCS can be used informationally, i.e., to provide explanation of specific features being used
- Question on langIdUse – Gregor suggests change of values to *language*, *regional variant*, *default*, *marked exception*
- Question of whether defaults for morphological and semantic features should be renamed and/or extended
- Discussion of inflection handling: need to add explicit alternation patterns to inflection values tables
- xml:lang – should we use it? Idea to leave the *language* tag, but also incorporate xml:lang as keyDC attribute
- Question on necessity for orthographic variant type
- Question whether *transfer action* is appropriately named; perhaps we should add *context* to the action part of the transfer condition
- Joerg raised question about a public ID for the OLIF DTD
- Discussion of how to address shared resources (see action items below)

Action Items:

1. **Participants are to send ideas via e-mail to Christian/Susan for definition of workflow process attributes**
 2. **Christian will write proposal for content/structure of a DCS**
 3. **Rework the values for langIdUse based on the consortium meeting**
 4. **morphDefaults to be renamed to inflectionDefaults**
 5. **semDefaults to be bifurcated to semTypeDefaults and semReadingDefaults**
 6. **Susan to add alternation patterns to inflects-like class descriptions**
 7. **Part-of-speech coverage will be left in DCS registry; values list will be reviewed after testing**
 8. ***language* tag will be maintained; xml:lang will be permitted as an attribute for keyDC element**
 9. **Susan/Christian will review the *orthVariantType* element**
 10. **Change name of *transfer action* (*trAct*) element**
 11. **Establish namespace for public identifier for OLIF DTD – persistent URL where server can change, but address will be the same**
 12. **Christian/Joerg will find out what SALT does for representing shared resources**
- Susan led a discussion of testing, implementation, and certification plans for OLIF v.2
 - When asked plans for testing/implementation, participants responded as follows:
 - Linguattec will try to implement
 - Logos wants to implement – asks if we have OLIF1-to-OLIF2 conversion
 - EC will support
 - Trados cannot support until end of May this year; says import will be more difficult due to required part-of-speech feature
 - IAI interested in using for CL
 - Systran will work with OLIF in 2 EU projects
 - Sail Labs will develop converter for MT and multilingual content
 - SAP is currently working on SAPterm-OLIF converter
 - Jean Senellart of Systran volunteered to be the test manager for OLIF v.2 testing
 - Members are to provide data for testing
 - Members will test OLIF representation of data and report back results in June
 - Issue of validation discussed
 - Several levels of validation, e.g., allowable values, validity of canonical form, validity of entry structure, implicational relationships
 - Suggestion to work on validation via SALT collaboration
 - Question of certification – general agreement to wait until we have software support before we have concrete discussions about certification

- Susan briefly discussed OLIF-SALT cooperative work
 - OLIF-SALT working group consists of Christian Lieske, Gregor Thurmair, Joerg Schuetz, and Laurent Romary
 - Areas of convergence for the two projects are:
 - data category registry (in RDF)
 - TMF representation
 - RDF for header information
 - canonical form guidelines – working group within OSCAR using OLIF suggestions as base for guidelines
 - domain hierarchy – suggestion to provide top-level hierarchy with way for user to generate user-specific domains; several suggestions for base: use Library of Congress schema; Carlo showed EC schema with app. 30 general domains that we could use to start
 - validation – see comments above re validation
- Issue of semantic reading raised:
 - Mike D. suggested we adopt the scheme in Roget's Thesaurus to indicate semantic readings; Gregor suggested looking at WordNet for word sense representation
- Jean has been working on specs for the *synStruct* and will distribute his comments
- General outline of priorities for upcoming work:
 1. Present OLIF work at MT Summit
 2. Collect/test data
 3. Develop OLIF converters

Action Items:

13. Jean will distribute his work on *synStruct* element
14. Jean will take over organization of testing
15. Members will provide test data to Jean ASAP
16. Test data will be represented in OLIF form with target date for results = end of May 2001
17. Work on OLIF structure/content/XML representation should be presented at MT Summit in September; paper to be submitted by April 15, 2001
18. Information will be posted on our web site:
 - a. ppt slides from this meeting
 - b. instructions on using the structure editor
 - c. test data – raw and in OLIF format
 - d. current draft of guidelines for canonical form formulation

Our next meeting will take place mid- to end of June, 2001 to discuss problems with test data representation

- The meeting was adjourned at 4:15 PM

Susan McCormick
April 17, 2001