

OLIF2 Consortium Organizational Meeting

April 6, 2000
SAP
Walldorf, Germany

MINUTES

Revised May 5, 2000 - SM

The meeting came to order at 9:30 AM.

Present were:

Gregor Thurmair, Sail Labs
Johannes Ritzke, Sail Labs
Alex Murzaku, Logos
Pierre-Yves Foucou, Systran
Yves Mahé, Xerox
Paulo Martins, EC
Chris Pyne, L10NBRIDGE
Jörgen Danielsen, L10NBRIDGE
Nils van der Laan, Trados
Peter Quartier, Lotus
Pierre-Yves Foucou, Systran
Ulrike Irmeler, Microsoft
Daniel Grasmick, SAP
Susan McCormick, SAP
Jennifer Brundage, SAP
Christian Lieske, SAP
Christoph Pahlke-Lerch, SAP

- Participants introduced themselves and gave brief overviews of the translation tools their companies/organizations are currently developing, using and/or supporting:
 - Trados is supporting TMX for TWB and TBX for MultiTerm.
 - Lotus is using the OTERM format.
 - EU has 6 million entries in an Oracle DB; covers total of 17 languages with 5 new ones; has 3 separate DB's (Euterpe, Eurodicautom, TIS); 80,000 term requests per day in Web version of Eurodicautom; will support TMX.
 - Xerox is working on term management for Asian languages; moving from Access to SQL server; tools include term mining and term checking; integrating TMX.
 - Microsoft covers 25 languages in bidirectional DB's.
 - L10NBRIDGE has Foreign Desk and Nortel tools for TM; supporting TMX.
 - Logos has moved from Unix to NT; has Java front end with Oracle-DB back end; currently has 40,000 English entries; Term Search tool for terminology mining.
 - SAIL Labs is working on speech technology with audio retrieval and a conversational system; content technology includes information retrieval from the Web; MT system is Comprendium, term DB is LexTor with upwards of 800,000 entries.
 - Systran has an SGML-based term format; has specific dictionaries for EU and other applications; main coverage is for European languages, but Asian languages are in development.
 - SAP is currently using Logos and SAIL Labs MT tools; SAPterm is the internal term DB; MultiTerm is also used; TM systems include TWB and SAP-internal tools; SAP will double its terminology volume by the end of 2000 and add additional NLP tools needing term data
- Daniel Grasmick stated the purpose of the consortium as upgrading "the current OLIF standard so that it can be supported by tool vendors and applied by users in 2001." OLIF1 was a prototype and is not usable as is. The SALT project will use OLIF as part of its XLT standard, but will not edit it

for content. LISA's TBX will be based on SALT XLT, and thus on OLIF. Since none of the other formats available supports MT requirements, upgrading OLIF is desirable.

- SAP suggested the following time frame for OLIF2 work:
 - Phase I: Specification
 - Working groups make recommendations for changes to OLIF format by May 31, 2000
 - Specifications for OLIF2 complete by September, 2000
 - Phase II: Implementation
 - Tool vendors support new format in 2001
 - Maintenance tools developed by end of 2000/beginning of 2001
- Gregor Thurmair presented background information on the design of OLIF and spoke on the current status of OLIF at SAIL Labs.
 - Rationale for OLIF: users process same text with a number of different tools and do redundant maintenance on lexicons and term DB's. Migrating from one lexicon to another is problematical. Common format would streamline work.
 - OLIF married requirements for MT and terminology.
 - Desire to keep OLIF simple and based on actual data from participating systems.
 - Format should support unicode and be XML-compliant.
 - Files have headers and bodies; relatively flat structure.
 - Entries follow concept-orientation, are monolingual with links.
 - Lex/term information categorized as part of the central entry, linguistics, terminology, transfer, or cross-reference.
 - Reviewed some of the linguistic features, including morphology, syntax, semantics, and terminology features such as definitions, comments, and administrative information.
 - Current status at SAIL: Implementation of a central DB, OLIF parser and generator; different versions of OLIF for OTELO and Aventinus projects; have their own formalism in addition to formalisms for SGML and XML. OLIF is supported by Logos, T1, SAPterm and DanTerm.

Paulo Martins suggested that we look into further factoring out repetitive information from the entries and incorporating it in an entry header. Mr. Martins also suggested that we look into having the format support allowing multiple <MONO>'s in an entry with no associated <XFR> links, so that entries could be represented without the high level of repetition that occurs if transfers are explicitly stated in each case.

- Susan McCormick reported on suggestions for changes to OLIF:
 - Three areas were discussed: technical, linguistic, terminology
 - Technical changes included XML-compliance, support for unicode, review of the central-entry base, and a distinction between monolingual and bilingual links.
 - Linguistic changes concentrated on morphology, where coverage for many languages must be added, as well as syntactic frames/types, semantic types, transfer conditions/actions, and conventions for canonical form representation. Systran requirements have to be reviewed and covered.
 - Changes to the terminology sector were less obvious; a review of current coverage was suggested; addition of a field for a unique entry identifier was proposed.
 - The question of a unique entry identifier generated a discussion on the purpose and need for this functionality. SAP uses the GUID to keep track of its term entries. The suggestion for OLIF is not for a sequential number or formal support for a GUID – rather an open field for the user to define and use as desired.
 - Current OLIF supports only English, German, and Danish. After general discussion, it was decided that OLIF2 development would target further language coverage based on priority:
 - Priority 1: EN, DE, DA, FR, ES, PT, JA
 - Priority 2: RU, IT, NL
 - Other: EL, HU, ZH, ZF, KO, AR
 - Gregor Thurmair suggested that we consult the EAGLES project work to expand on the morphology codes. Gregor agreed to follow up on this

- On the level of support that OLIF now enjoys:
 - Logos has conversion in place in both directions
 - SAIL will support import of OLIF files by the end of May, 2000
 - SAP has conversion of SAPterm to OLIF
- Susan McCormick reviewed some of the relevant interchange formats/projects, including MARTIF, its extension X-MARTIF, OSCAR, Geneter, and SALT. Current plans are for OLIF to be integral to SALT's XLT standard. Interaction with the SALT project was discussed. Consensus was that we should work independently, but ensure that our standard is compatible with SALT.
- Several action items were defined:
 - Action item 1: Susan will prepare a written description of the relationship between the OLIF2 consortium and the SALT project.
 - Action item 2: Peter Quartier will set up an OLIF2 web site. SAP will populate the site with information on progress within the project.
 - Action item 3: Participants will, where possible, provide data samples with format and content description.
 - Action item 4: Papers provided by IBM and Trados on their work in the OTELO project should be located to check for relevance to OLIF2 work.
- Task groups were set up; purpose of each group and tasks associated with the group were delineated:

Technical group: Goal is to provide formal structure of the format

Tasks:

Review for XML compliance
 Redundancy
 Links representation
 Definition of the header
 Incorporation of user-defined fields

= Output: OLIF DTD

Task group members:

Trados
 SAIL Labs
 Systran
 SAP
 Lotus
 (Microsoft – possibly)
 (L10NBRIDGE – possibly)

Group coordinator: Christian Lieske, SAP

Linguistic group: Goal is to provide a revised list of feature-value pairs for the linguistic component

Tasks:

- Canonical form formulation
- Morphology, syntax and semantics
- Transfer conditions and transformations
- Cross-references (based on ISO)

Task group members:

- SAIL Labs
- Logos
- Systran
- SAP
- CST (EAGLES)
- (IBM – possibly)
- (Xerox – possibly)

Group coordinator: Susan McCormick, SAP

Terminology group: Goal is to provide a revised list of feature-value pairs for terminology

Tasks:

- Concordance with other standards
- Administrative information
- Review domain handling in OLIF

Task group members:

- Microsoft
- Trados
- L10NBRIDGE
- IBM
- Xerox

Group coordinator: Christian Lieske, SAP (temporary assignment)

- The meeting was adjourned at approximately 4:15 PM.

Susan McCormick
Consultant to SAP
Baltimore, MD, USA
April 11, 2000