



ISO/TC 37/SC 4 **N 064**

ISO/TC 37/SC 4/WG 1 N28
2003-07-19 KY/DH

ISO/TC 37/SC 4
Language resource management

Charter for Joint ISO-TEI activity on Feature Structures

This document goes together with a draft of a New Work Item Proposal in ISO format for your information.

Draft by Laurent Romary; 28 November 2002

With comments from Lou, 8 Jan 03

Note: in ISO sense, the 'editor' is the person responsible for moving along a standard proposal up to its final stage. He or she acts as the head of an editorial group.

Background

The Text Encoding Initiative contains two complementary chapters on Feature Structure Representation (chapter 16) and Feature Structure Declaration Representation (chapter 26). Those two chapters originally drafted by Terry Langendoen and Gary Simons are now generally recognized as covering many needs in the field of linguistics.

Moreover, some Natural Language Processing applications have also provided implementations based on the proposed representations, which has further increased interest on this specific aspect of the TEI in this community. It should also be noted that these chapters are closely integrated with the rest of the TEI scheme, thus opening up the application of NLP techniques to a very wide community of users, while at the same time offering the NLP community access to a real-world range of different text types and applications

Proposal

It is thus suggested that the revision of the two chapters be conducted under the auspices of TC37/SC4, and that the corresponding editorial group be seen as a working group from the point of view of the TEI.

The most practical way to do so is simply that the TEI should be an official liaison to TC37/SC4 and would thus have access to all documents and debates of the group.

Note: the idea is to work in a slightly more coordinated way as what occurred with terminologies, where no continuous reporting to the TEI occurred.

Constraints resulting from being a TEI working group

Editing the FS and FSD documents in close connexion to the TEI consortium would imply that the editors pay extreme attention to maintain compatibility with general TEI mechanisms. In particular, the two following items should be considered:

- Clarifying the descriptions of alternatives, for example, there exist at present two possibilities to alternate between features and between values (ad hoc vAlt and fAlc vs. general alt mechanism): the work group should provide guidance on the circumstances in which either method might be preferable;
- Compatibility with the final proposals on the use of pointers and links from the stand-off markup TEI working group. It should be noted that, more generally, a close collaboration between TC37/SC4 and the TEI must be established on this topic;

Constraints resulting from becoming an ISO TC37/SC4 work item

Making the compilation/revision of the two chapters a TC37/SC4 work item would have the following consequences:

- It would be necessary to make the proposal more independent of the general TEI document architecture (feature structures would be the actual entry points, some general attributes may have to be dropped);
- It would oblige the document to be conformant to the general structure of an ISO standard (with scope, terms and definitions, etc.);
- It would give additional international visibility to the document, since it would be widely distributed to TC37/SC4 members, thus providing additional sources of comments and suggestions for modification;

Main action to be taken by the joint working group

The following items have been identified as requiring revision in the current documents:

- Preliminary formal description of feature structures additional to what is already provided
- Provision of a simplified representation (FS lite) describing the basic subset of FS representation without libraries;
- Provision of a re-entrance mechanism;
- Description of typed feature structure
- Simplification of feature value content by replacing some elements (<symp>, <num> etc.) by references to types (à la XML schemas)
- Provision of more NLP related examples (syntactic (e.g. à la HPSG), semantic information etc.)

Budget issues

No specific budget will be requested from the TEI to support those activities considering that they will be incorporated into the normal ISO TC37/SC4 activities (experts are sent on a national basis).

As a liaison, the TEI will send specific representatives to all editorial group meetings (expectedly one of the two editors).

In the case a proposal is made to an EU call to fund those activities, the TEI council would be informed.

Agenda

The following agenda is presented as an indication of how things could be managed within TC37/SC4:

- February 2003: a New Work Item Proposal (NP) together with a first draft is sent to TC37/SC4 members for approval and comments (a good opportunity to check that the liaison mechanism works)
- May 2003: result of voting and compilation of comments
- July 2003 (SC4 meeting in conjunction to ACL conference in Sapporo): study of comments and decision to proceed to Committee Draft (CD), or, if things are going well as a Draft of International Standard (DIS)
- November 2003: DIS document, incorporating Sapporo comments, is sent for vote to TC37 members
- Dec. 2004: Publication.

A key factor to timeliness is the immediate reactivity to incorporate comments and provide update documents after plenary SC4 meetings.

Participants

Editor:

Kiyong Lee klee@korea.ac.kr

Members (to be confirmed):

Syd Bauman Syd_Bauman@brown.edu

Lou Burnard lou.burnard@oucs.ox.ac.uk

Eric de la Clergerie Eric.De-La-Clergerie@inria.fr

Tomaz Erjavec Tomaz.Erjavec@ijs.si

Ulrich Heid Ulrich.Heid@IMS.Uni-Stuttgart.DE

Terry Langendoen langendt@U.Arizona.EDU

Laurent Romary Laurent.Romary@loria.fr

Gary Simons Gary_Simons@sil.org