

ISOcat in Daily Life
Possible Uses of a Large Repository
of
Widely Used Linguistic Concepts

Monday morning, 21 May 2012

Presenters:

Matej Durco, Irina Nevskaya, Ineke Schuurman,
Franca Wesseling, Menzo Windhouwer, Sue Ellen Wright

Tutorial Overview

09:00 – 09:20 Welcome and introduction to ISOcat	Sue Ellen/Menzo
09:20 – 09:30 Prologue: the CGN showcase	Ineke
09:30 – 10:00 <i>Hands-on</i> : ISOcat basics	<i>all</i>
10:00 – 10:30 <i>Hands-on</i> : creating Data Category selections	<i>all</i>
10:30 – 11:00 coffee	
11:00 – 11:10 Data Category specifications	Menzo
11:10 – 11:40 <i>Hands-on</i> : creating Data Category specifications	<i>all</i>
11:40 – 11:55 Known problems (do and don't)	Ineke
11:55 – 12:05 Epilogue: the CGN showcase	Ineke
12:05 – 12:20 Tagset bridges in Edisyn	Franca
12:20 – 12:30 Beyond ISOcat	
12:30 – 12:45 The RELISH MDF and GOLD crosswalk	Irina
12:45 – 13:00 Semantic mapping in CMDI	Matej

Prerequisites for the *hands-on* sessions:

- a. Bring your own laptop (we'll use the LREC internet connection or a special purpose local wireless network)
- b. Register before Sunday 20 May at www.isocat.org (you'll need your credentials to login to the ISOcat "testbed")

Tutorial Description

Language resources are a very valuable asset, whether they come just with metadata, or with one or several types of annotation (PoS, syntax, semantics, ...). Not only now, where they form the basis for new scientific publications, but also in the future when new research might need to reassess previous findings. In order to relate annotation or metadata schemes, for example when comparing two syntactically annotated corpora, when adding a layer of annotation making use of a previous one, or when comparing various instantiations of the 'same' annotation scheme (which may have been adapted over time), one needs information on these schemes, preferably in relation to each other, to be able to perform such a task. ISOcat can be used for such a task, alone or in combination with RELcat and SCHEMAcat. They enable you to specify relations between (parts of) schemes.

ISOcat is a linguistic concept database developed by ISO TC 37 to provide reference semantics for annotation schemata (both features/attributes and values, all of which are designated as Data Categories (DCs) in the context of the repository). Some of the DC specifications in ISOcat will be standardized, meaning that their use is promoted. These items will mainly be descriptions that are general enough to be used by as many users as possible. In some cases, however, the theoretical background of an annotation scheme presupposes different definitions from those available, which necessitates the creation of new, perhaps closely related DCs. New DCs should be related to each other and to the existing definitions, which can be achieved using RELcat and SCHEMAcat, two other registries designed to express relations between specified DCs, which are referenced using the persistent identifiers for their respective ISOcat DCs.

This tutorial will teach how to deal with DCs in ISOcat. It will provide hands on experience with the ISOcat web interface (viewer and editor), show how to find existing DCs, how and when to create new ones, how to work with DC Selections (DCS; coherent, defined sets of DCs), etc.

Instructional sessions will be alternated with practical ones as well as with reports from experienced users, who will explain why they make use of ISOcat, and what are its benefits. After the tutorial the participants should be able to decide when an existing DC can be reused, when a new one is to be defined and how this new DC relates to the existing ones. They will also be able to construct such new DCSs as well as new DCs. During the tutorial the participants will make use of the ISOcat "testbed", which will enable them to experiment without doing any harm to the "real" ISOcat.