

LREC 2012 ISOcat tutorial - exercises

Introduction

In the field of linguistic typology we're interested the diversity of the languages of the world. In many cases this means the creation of a typological database collecting information on how a linguistic phenomenon is expressed from a balanced sample of languages of the world. In this example we'll create such a database and annotate its schema with ISOcat data categories.

A database schema

The basic word order of a language is a very well-known linguistic phenomenon. A common encoding is based on subject, object and verb. This results in 6 possible orders:

1. subject-object-verb (sov)
2. subject-verb-object (svo)
3. object-subject-verb (osv)
4. object-verb-subject (ovs)
5. verb-object-subject (vos)
6. verb-subject-object (vso)

A simple database schema would consist of one table containing the following columns:

- language name
- language code (ISO-639-3)
- basic word order
- source
- informant

Hands-on: ISOcat basics

You might work in a team on this database schema (this is just the start ☺):

1. create an ISOcat group
2. invite your neighbor to this group

Hands-on: creating Data Category selections

ISOcat might already contain interesting data categories for this schema:

1. create a Data Category Selection (DCS)
2. share the DCS with the just created group
3. collect existing data categories matching with the semantics of the columns in the new DCS

Hands-on: creating Data Category specifications

Not all columns can be related to an existing data category, e.g., basic word order. Create

1. some simple data categories for some of the values for basic word order
2. a new closed data category for basic word order with these new simple data categories in its conceptual domain

Think about good definitions, data element names and name sections.

Take home: beyond ISOcat – data category references

How would you make the relations between our database (schema) and ISOcat data categories explicit?

Take home: beyond ISOcat – ontological relationships

The existence of object and subject is for some linguists debatable. In their view word order has three basic values:

1. predicate initial
2. predicate medial
3. predicate final

How would you express and store the relationship between a simple data category or concept for predicate initial and one of the simple data categories created in the hands-on session?