

# **STO –LMF documentation**

## **Syntax**

April 2013

**Sussi Olsen, Lene Offersgaard, Dorte Haltrup Hansen**

## 1 Introduction

In the META-NORD project<sup>1</sup>, an EU project which aimed at developing and documenting methodologies for building language resources for the under-resourced languages in the Baltic and Nordic countries, UCPH, Centre for Language Technology undertook various language resource initiatives including the upgrade of the STO export format to Lexical Markup Framework (LMF).

The Lexical Markup Language is an internationally well-known and accepted XML format and the ISO standard for Natural Language Processing (NLP) lexicons. See [www.lexicalmarkupframework.org](http://www.lexicalmarkupframework.org) for more information on LMF.

The export format for STO used to be a flat, comma-separated text format for the morphological part and an XML format developed at UCPH for the syntactic part. The advantage of using an XML format common to various other lexical resources and widely accepted all over the world is obvious.

The following documentation is meant as an introduction to the STO LMF format. After the introduction of the LMF structure we describe the structure of the STO entries when converted to LMF and explain some of the choices and adjustments we have made. Finally, we show a table of data categories (STO and LMF) that have been changed in the conversion – i.e. a list of old and new data category names. As an appendix we add the lexical entry of circus as a good example of a complex noun entry.

Please note that this documentation does NOT document the general content and the linguistic aspects of STO. For an introduction and documentation of these parts, you may consult the [STO Sprogteknologisk Ordbase, Danish Monolingual lexicon, Documentation, version 2](#).

## 2 Table of the STO data categories and the corresponding LMF categories

---

<sup>1</sup> META-NORD was funded by the DG INFSO of the European Commission through the ICT PSP Programme, Grant Agreement: No 270899

STO LMF syntax

LMF structure	LMF features		IsoCat (for attributes)	STO xml	
	Attribute	Value (isoCat no)		Attribute	Example
<b>LexicalEntry</b>				Morph_Syn_Units	
<b>Lemma</b>	id	(string)	1845	Mu_Synu	
<b>FormRepresentation</b>	morphologicalUnitId	(string)	5282	Mu_Id	HÆFTE,3
	writtenForm	(string)	1836	spelling	hefte/hæfte
<b>SyntacticBehaviour</b>	officiallyApproved	yes(1904), no(1905)	5284	ro_approved	YES/NO
	synuld	(string)	5588	Synu_Description	
<b>subcategorisationFrames</b>	example	(string)	1958	Synu_id	SYNU_HÆFTE_3_1
		(string)		Description_Id	Dv2N
<b>SubcategorizationFrame</b>				Description	
	naming	(string)	5589	Naming	divalent: NP, obligatory, NP
	example	(string)	1958		
<b>id</b>	constructionId	(string)	5590	Construction_Id	Cv2N
<b>LexemeProperty</b>		(string)	1845	Description_Id	Dv2N
	partOfSpeech	adjective(1230), adverb(1232), noun(1333), verb(1424)	396	Self	
	selfId	(string)	5591	Cat	VERB
	reflexiveVerb	yes(1904), no(1905), unspecified(1908)	5592	Self / id	have_NO_NO
	takesParticle	(string)	5593	Reflexive	NO
	takesAuxiliary	(string)	5594	Particle	NO
	passiveVerb	yes(1904), no(1905), unspecified(1908)	3840	Auxiliary	have
	modal	yes(1904), no(1905)	1329	Passive	NO
	auxiliary	yes(1904), no(1905), unspecified(1908)	1244	Modal	NO
	adjectivalFunction	attributiveFunction(5287), predicativeFunction(5288)		Aux	NO
<b>SyntacticArgument</b>	positionNumber	(string)	5595	Adj_Func	
	syntacticFunctionType	subject(1391), indirectObject(1310), directObject(1274), nominalComplement, adverbialComplement(4639), clausalComplement, externalComplement, formalComplement, formalSubject, objectComplement(4625), prepositionalComplement(4638), relationalGenitive, somPrepComplement, specifierNoun, subjectComplement(4624)		Construction	
	optional	yes(1904), no(1905)	5596	Position / number	1,2,3,4
	syntacticConstituentLabel	NP(2256), PP(2257), clause(2295)	5597	Function	
	syntacticConstituentPhraseld	(string)	5598	Phrases	Optional NP, PP, Clause
	case	nominativeCase(1331), accusativeCase(1226), genitiveCase(1293), unspecified(1908)	5601		Id Casus
	reflexiveVoice	yes(1904), no(1905), unspecified(1908)	1840		Reflexive
	expletive	(string)	3842		Eksplitive
	definiteness	indefinite(2005), unspecified(1908)	1283		Definite
	npIndex	(string)	1926		NPIndex
	introducer	(string)	5603		Introducer
	ppComplementLabel	NP(2256), infinitiveWithoutControl, infinitiveSubjectControl, infinitiveObjectControl, infinitiveIndirectObjectControl, infinitivePrepositionalComplementControl, interrogativeClause(2299), whType(2299)	2245		Clause_Type / NP_type
	controlType	subjectControl(4187), objectControl(4189), indirectObjectControl, prepositionalComplementControl, subjectRaising(4188), withoutControl	5610		Control
	coreferenceRelation	I, J, K, N, withoutCoreference	5604		Coref
	finite	yes(1904), no(1905)	2974		nil
	clauseType	thatType(4623), whType(2296), interrogativeClause(2299), infinitiveNoIntro, infinitive(1312)	1287		Clause_Type
			5602		