STO –LMF documentation

Syntax

April 2013

Sussi Olsen, Lene Offersgaard, Dorte Haltrup Hansen

1 Introduction

In the META-NORD project[1], 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 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

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