Cross-linking Experience of Estonian WordNet

Neeme Kahusk, Heili Orav, Kadri Vare
University of Tartu

Introduction

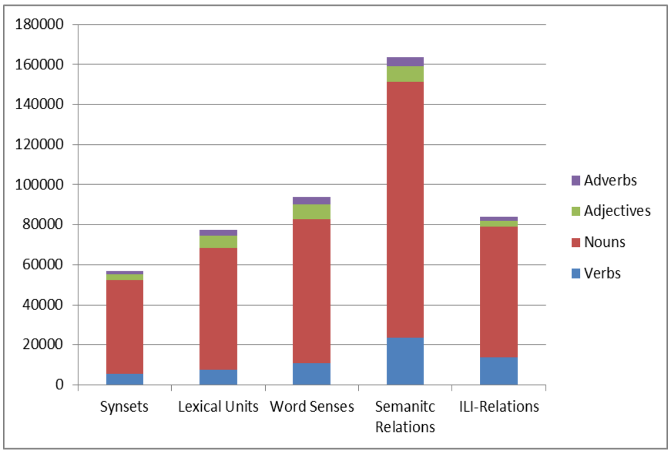
Our paper describes work we have done for Estonian WordNet according to META-NORD project tasks. We discuss the linking process of Estonian WordNet and CoreWordNet from linguistic, lexicographical and technical point of view.

Estonian Wordnet

Estonian Wordnet (EstWN) contains at present (October 2012) approximately 57 500 concepts. EstWN includes nouns, verbs, adjectives and adverbs, also there are some multiword units present and the main vocabulary of Estonian language is mostly covered.

EstWN has been compiled mostly manually but there are some endeavors for automatical additives. Automatically we have included an amount of words which have been derived via suffixes.

Secondly, our approach so far has been also domain-specific - we have added concepts from semantic fields like architecture, medicine, philosophy and so on. Thirdly, we use frequency lists of corpora and results of sense disambiguation. Besides quantity we are very concerned with quality control and try out different methods like algorithm of visualization (diagnostic tool), which indicates to possible errors as non suitable concepts, semantic relations etc (Lohk 2012).



META-NORD project

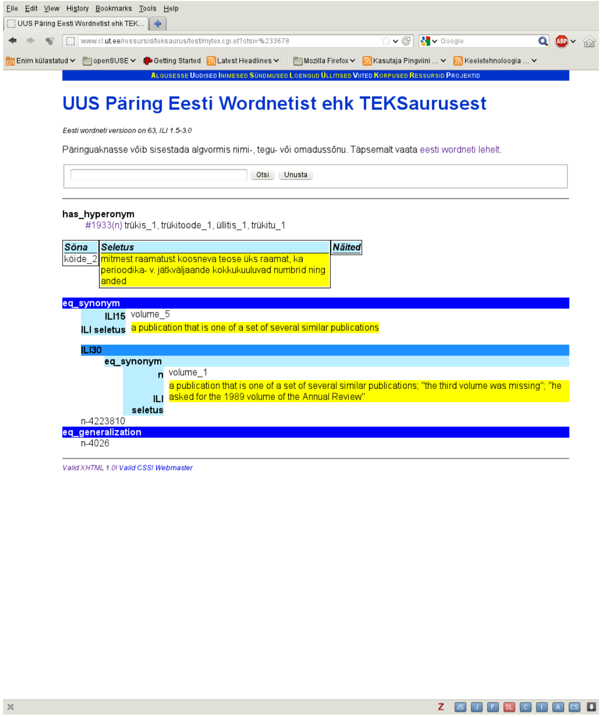
META-NORD is an EC project closely related to the META-NET network, whose aim is to take care of technological support to multilingual European information society. One of the key activities of META-NET is diminish high fragmentation and lack of unified access to language resources that hinder European innovation potential in language technology development and research.

The META-NORD project aims to establish an open linguistic infrastructure in the Baltic and Nordic countries. The project will focus on eight European languages — Danish, Estonian, Finnish, Icelandic, Latvian, Lithuanian, Norwegian and Swedish.

Besides general objectives META-NORD has several specific targets, and providing expertise in wordnets is one of them.

Linking with Core Wordnet

There is a subset of Princeton WordNet called core wordnet. This consists of about 5000 most frequently used English word senses, compiled semi-automatically. Core wordnet is part of WN ver. 3.0, and Estonian WordNet is by default linked to Princeton WordNet ver. 1.5 via many different semantic relations, so we had to map Estonian WordNet to Princeton WordNet ver. 3.0 at first, and later on adjust the results manually.


File Edit View History Bookmarks Tools Help
UUS Päring Eesti Wordnetist ehk TEK...
www.cl.ut.ee/ressursid/teksaurus/test/mytex.cgi.et?otsi=%233678
Google
Enim külastatud
openSUSE
Getting Started
Latest Headlines
Mozilla Firefox
Kasutaja Pingviini ...
Keeletehnoloogia ...
ALGUSESSE UUDISED INIMESED SÜNDMUSED LOENGUD ÜLLITISED VIITED KORPUSED RESSURSID PROJEKTID
Otsi
Unusta
Valid XHTML 1.0! Valid CSS! Webmaster

UUS Päring Eesti Wordnetist ehk TEKSaurusest

Eesti wordneti versioon on 63, ILI 1.5-3.0

Päringuaknasse võib sisestada algvormis nimi-, tegu- või omadussõnu. Täpsemalt vaata eesti wordneti lehelt.

has_hyperonym
#1933(n) trükis_1, trükitoode_1, üllitis_1, trükitu_1

Sõna	*Seletus*	*Näited*
köide_2	mitmest raamatust koosneva teose üks raamat, ka perioodika- v. jätkväljaande kokkukuuluvad numbrid ning anded	

eq_synonym
ILI15 volume_5
ILI seletus a publication that is one of a set of several similar publications

ILI30
eq_synonym
n volume_1
ILI seletus a publication that is one of a set of several similar publications; "the third volume was missing"; "he asked for the 1989 volume of the Annual Review"
n-4223810

eq_generalization
n-4026