

The Cornetto Database



RU | STEVIN

Overview

1. Project background information
2. Alignment of lexical resources
3. Database design
4. Database demo
5. Strategies of work

Goal

- <http://www.let.vu.nl/onderzoek/projectsites/cornetto/start.htm>
- Stevin tender project (STE05039) to develop a lexical semantic database for Dutch:
 - 40K Entries
 - Generic and central part of the language
- Data:
 - Combination of WordNet and FrameNet
 - Vertical and horizontal semantic relations
 - Combinatorial lexical constraints
 - Aligned with the English Wordnet
 - Extended with an ontology
- Automatic acquisition toolkit
- Started April 2006, ends March 2008
- Licensed from TST-centrale, Nederlandse Taalunie

Consortium

- Vrije Universiteit Amsterdam, Faculteit der Letteren, General Linguistics Department, Onderzoeksgroep Lexicologie/Terminologie:
 - Willy Martin, Isa Maks, Hennie vd Vliet, Roxane Segers, Piek Vossen
- Universiteit van Amsterdam, Instituut voor Informatica:
 - Maarten de Rijke, Erik Tjong Kim Sang, Katja Hofmann
- Katholieke Universiteit Leuven, Interdisciplinair Centrum voor Recht en Informatica (ICRI):
 - Sien Moens, Jan de Beer
- Irion Technologies BV:
 - Joop van Gent, Hetty van Zutphen, Piek Vossen

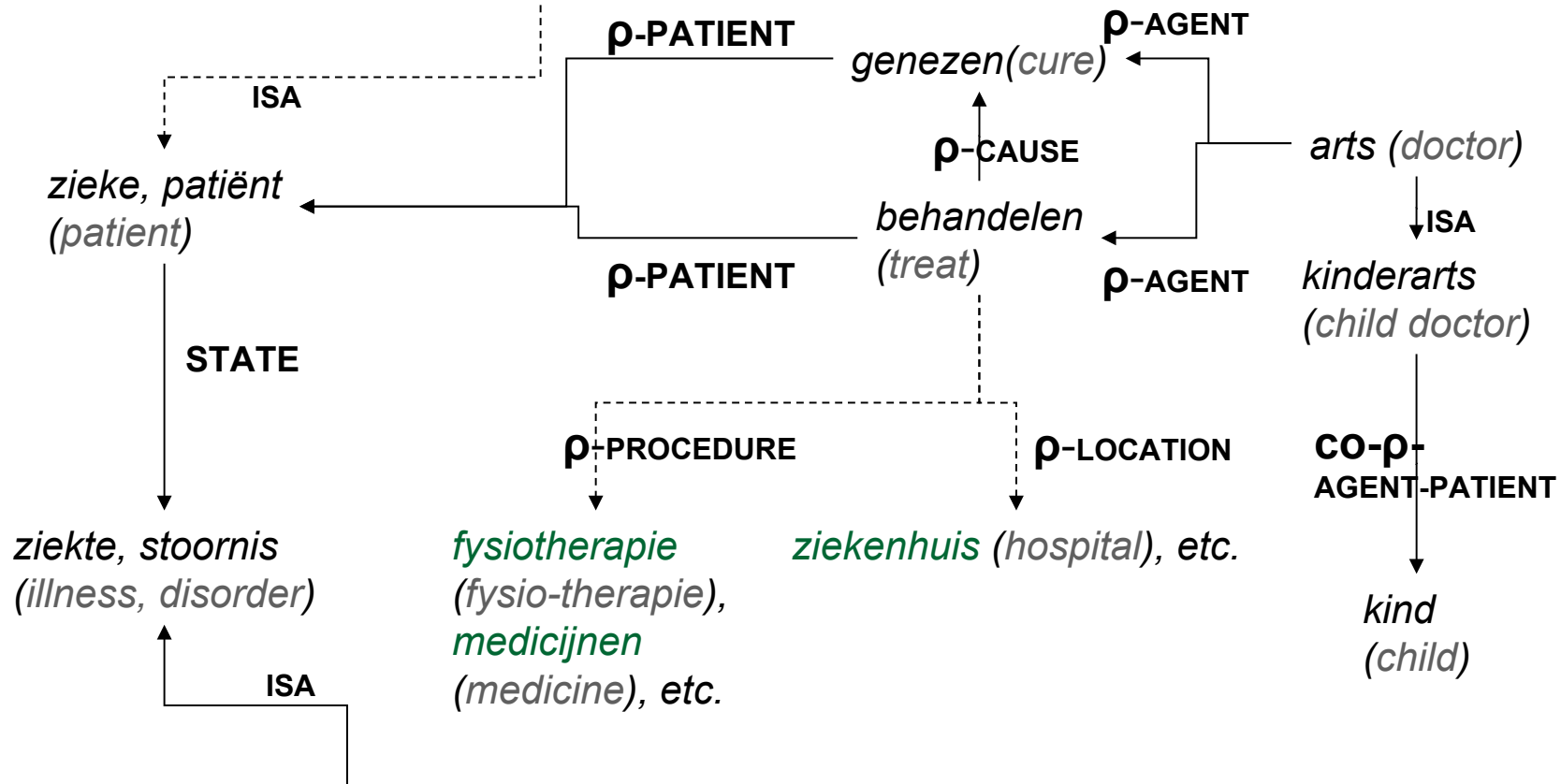
Other partners

- User-group:
 - ❑ Polderland
 - ❑ Knowledge Concepts
 - ❑ LibRT
 - ❑ Irion Technologies
 - ❑ Van Dale Lexicografie
 - ❑ Larcier-De Boeck
 - ❑ Rik Schutz
 - ❑ Gridline

- Ontology-group:
 - ❑ Dr. W. Ceusters, Office Line Engineering nv
 - ❑ Prof. F. van Harmelen, Vrije Universiteit Amsterdam
 - ❑ Dr. P. Buitelaar, DFKI
 - ❑ Dr. P. Monachesi, Universiteit van Utrecht

Horizontal & vertical semantic relations

chronisch zieke (chronical patient), *langdurig zieke* (long-term patient),
psychisch/geestelijk zieke (mental patient)



maagaandoening (stomach disorder)
nieraandoening (kidney disorder), *keelpijn* (sour throat).

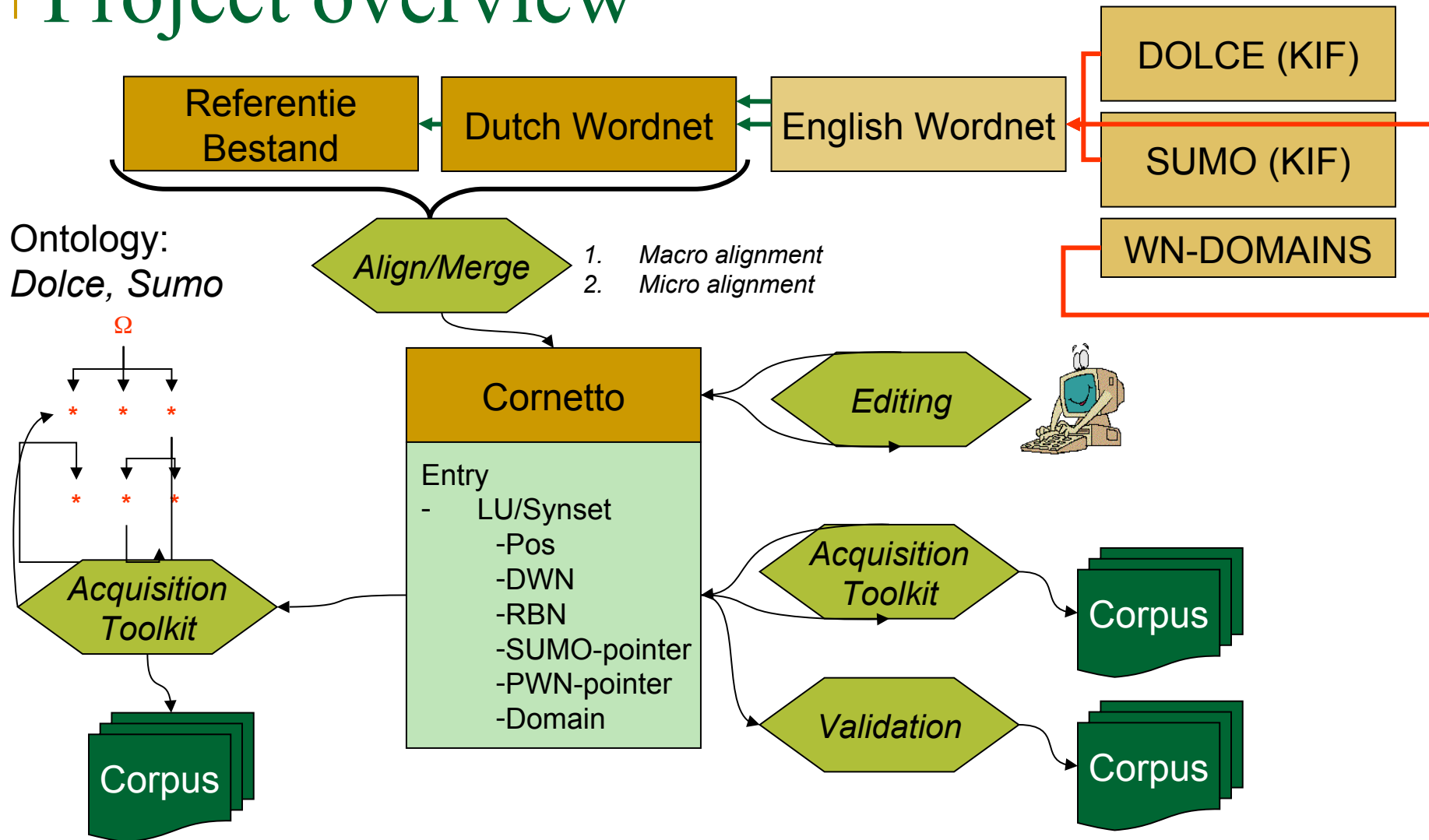
Combinatorics

slots	fillers (lex/conc)	fillers (coll)
action	<i>behandelen</i> (someone treat)	<i>iem. behandelen</i>
theme	<i>patiënt</i> (a patient treat)	<i>een patiënt behandelen</i>
state	<i>ziekte</i>	<i>iem. behandelen <u>voor</u> een ziekte</i> (someone treat for a disease)
	<i>behandelen</i>	<i>iem. <u>aan</u> zijn verwondingen</i> (somene at his injuries treat)
	<i>treat)</i>	<i>een ziekte behandelen (a disease</i>

Approach

- Combine the information from two existing Dutch lexical resources:
 - The Dutch wordnet: synsets and lexical semantic relations
 - The Referentiebestand Nederlands: morpho-syntactic information, semantic information, pragmatic information, frame structures, lexical functions and combinatorics
- Macro level alignment
- Micro level alignment
- Populate with an ontology

Project overview



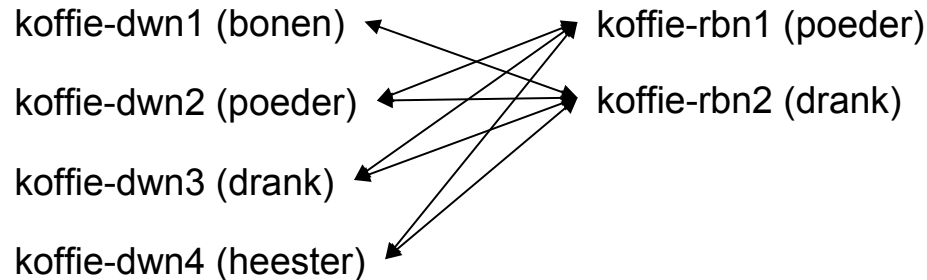
Global planning

- Two year project:
 - Month 1-6: design and database
 - Month 1-6: automatically aligned data
 - Month 7-10: ontology assignment
 - Month 7-22: editing
 - Month 7-15: acquisition
 - Month 16-17, 23-24: task-based evaluation

Alignment of lexical resources

Alignment

- Generate all weighted combinations:



- Produce merged output with mappings above probability threshold:

- New structure of word meanings

- koffie-cbn1(bonen) (source dwn1)
- koffie-cbn2 (poeder) (source dwn2, rbn1)
- koffie-cbn3 (drank) (source dwn3, rbn2)
- koffie-cbn4 (heester) (source dwn4)

Macro Alignment RBN - DWN

SYNSET	DEFINITION	DIFFERENTIAE	DOMAIN
baspartij_1 bas_1		die de bas zingt of speelt	#MUZ
bas_2 basstem_1	laagste mannenstem	laag, bij mannen	#MUZ
bas_3	baszanger, basspeler	met de basstem	#MUZ
contrabas_1 bas_4 basviool_1	het grootste en diepst gaande strijkinstrument	grootste en laagst klinkend	#MUZ

bas (noun) (bassen)

- 1 (count; nondynamic) [<gen-muz> zangstem] ⊢ <laagste> zangstem (BVD)
- 2 (count; human) [<gen-muz> zanger] ⊢ man met de stem van een bas (AA)
- 3 (count; artefact) [<gen-muz> contrabas] ⊢ strijkinstrument dat het grootst is en dat het laagste speelt (AA): *contrabas*
- 4 (count; artefact) [<gen-muz> basgitaar] ⊢ basgitaar (BVD-1)

6 Senses out of 8 candidates

- 1: lowest singing voice, RBN-1 & DWN-2
- 2: man with the voice of a bass, RBN-1 & DWN-3
- 3: biggest and lowest string instrument, RBN-3 & DWN-4
- 4: bass guitar, RBN-4
- 5: part of the music for the bass, DWN-1
- 6: bass singer or player, DWN-3

Strategies for the macro-alignment

	Conf.	Dev.	Factor	LINKS	
1: 1 RBN & 1 DWN meaning, no synonyms	97.1	4,9	3	9936	8,1%
2: 1 RBN & 1 DWN meaning	88.5	8,6	3	25366	20,8%
3: 1 RBN & >1 DWN meaning	53.9	8,1	1	22892	18,7%
4: >1 RBN & 1 DWN meaning	68.2	17,2	1	1357	1,1%
5: overlapping hyperonym word	85.3	23,3	2	7305	6,0%
6: overlapping hyponyms	74.6	22,1	2	21691	17,7%
7: overlapping domain-clusters	70.2	15,5	2	11008	9,0%
8: overlapping definition words	91.6	7,8	3	22664	18,5%

- 8 reviewers
- 100 random links per strategy
- nouns, verbs, adjectives, adverbs
- single confidence score per link based on all weighted strategies

Results of the macro-alignment

	LUS	LINKED		NOT-LINKED
RBN	66.024	47.250	72%	18.774
VLIS	106.504	46.924	44%	59.580
RBN-VLIS LINKS		58.053		

Cornetto Mapping Record

- CID unique pointer to bind them all, assigned by IRION
- C_LU_ID LU id to be assigned to each LU in CDB
- C_SY_ID SYNSET id to be assigned to each synset in CDB
- C_FORM lexical form
- C_SEQ_NR sequence number in CDB
- R_LU_ID LU id currently used in RBN
- R_SEQ_NR sequence number currently used in RBN
- D_LU_ID LU id currently used in DWN (original Vlis ID)
- D_SEQ_NR sequence number currently used in DWN
- D_SY_ID synset id currently used in DWN
- Score confidence score assigned by algorithm
- Status manually confirmed
- Name editor

Creation of Cornetto LUs and Synsets

- No mapping for a LU in RBN to a synonym in DWN:
 - create unique LU in Cornetto based on RBN LU. We do not create a synset for the LU in Cornetto;
- No mapping for a synonym in DWN to an LU in RBN:
 - create unique synonym in a unique synset in Cornetto
 - create corresponding Cornetto LU with the information from DWN;
- If there is a best scoring mapping between an LU in RBN and a synonym in DWN:
 - create single unique LU and a single unique synonym in Cornetto that point to each other and to both RBN and DWN;
- All remaining mappings:
 - do not create LUs and/or synsets;
 - stored as additional mappings (as weighted alternatives);

Micro-alignment

- ❑ Separate layers with co-indexing:
 - DWN:
 - ❑ gitarist<0> -co_agent_instrument-> gitaar<1>
 - RBN:
 - ❑ gitarist <0> speelt **op** een gitaar <1>
- ❑ Unified CBN structure:
 - Event structure:
 - ❑ E: gitaarspelen <e0>
 - ❑ A1: gitarist <a1>
 - ❑ A2: gitaar <a2>
- ❑ Conceptual information shared by all synonyms
- ❑ Lexical information unique per synonym

Database design

Lexical Unit & Synsets

- Lexical Unit = form-meaning relation, such that:
 - form = abstract representation of certain realizations;
 - part-of-speech is the same;
 - meaning is the same, where meaning is defined by a reference to a unique Synset;
- Synset = Set of synonyms (LUs) that refer to the same entities in most contexts.
 - Defined by lexical semantic relations;
 - Defined by reference to ontology Terms or KIF expressions involving Terms from the ontology;

Data structure overview

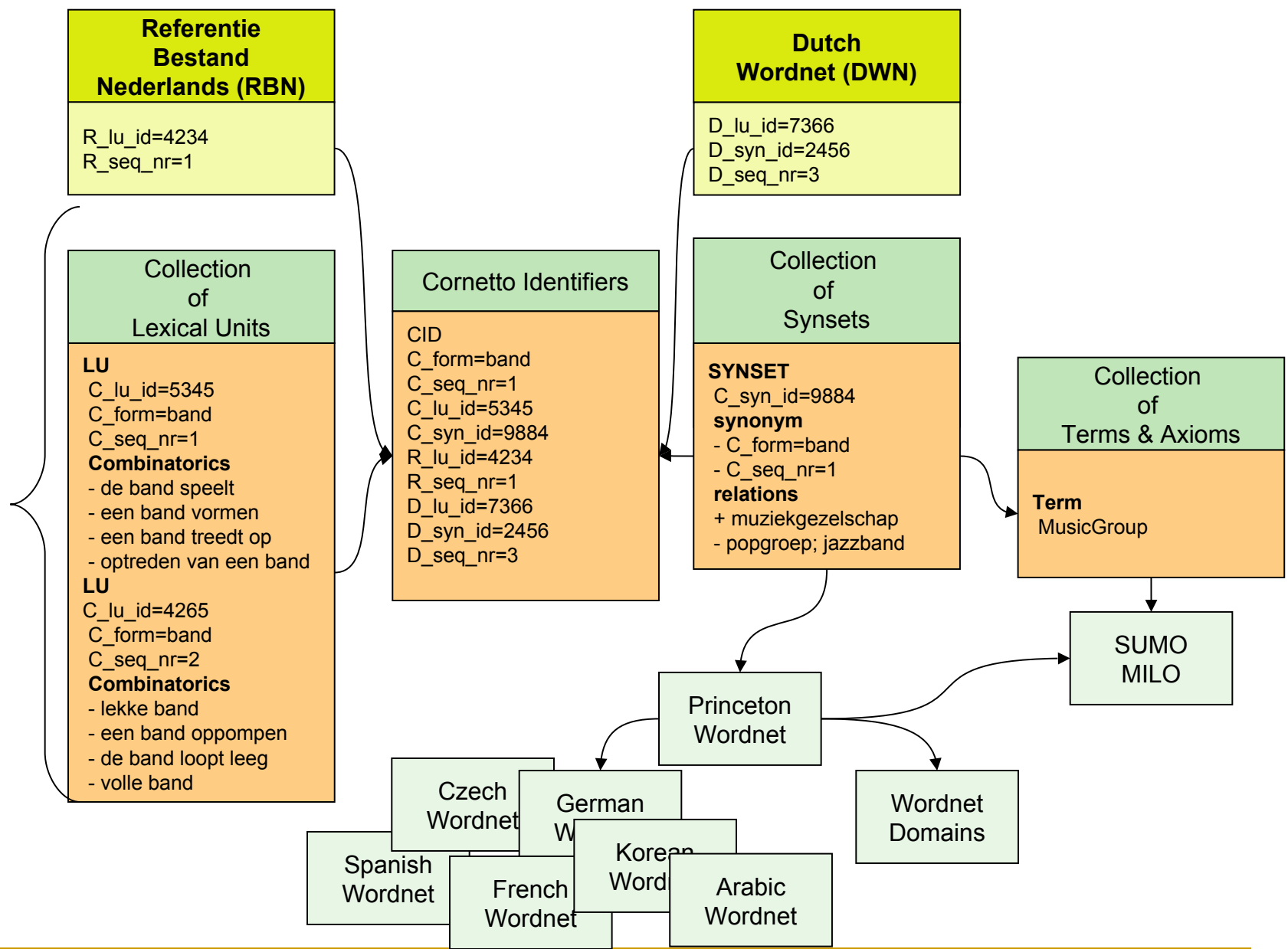
■ Collections:

- ❑ Lexical units (LU): -> mainly derived from RBN
- ❑ Synsets (SY): -> mainly derived from DWN
- ❑ Terms (TE): -> based on SUMO/MILO, linked to PWN
- ❑ Domains (DM): -> based on Wordnet domains

■ Mappings:

- ❑ LU<-> SY
- ❑ SY <-> SY (within Dutch and from Dutch to English)
- ❑ SY <-> TE
- ❑ SY <-> DM

Cornetto Database (CDB)



Lexical Unit: form variants

- Inflectional variants, *appel, appels, appelen*
- Spelling variants
 - Meaning is identical
 - Pronunciation is mostly identical (*droppel, druppel*)
 - Spelling is different but the morphology is mostly the same: spelling variation can be systematic or incidental but it is in both cases conventional.
- Shortening:
 - Meaning is identical
 - Pragmatics is usually different
 - Pronunciation and spelling are different
 - Reduction in length for efficiency:
 - short forms (bus vs autobus)
 - abbreviation
 - contractions
 - acronyms
- sms language



Lexical Unit: subtypes

- ❑ Different data structures for nouns, verbs & adjectives
- ❑ Single word unit:
 - simple morphological structure
 - complex morphological structure (compounds, derivations)
- ❑ Multiword unit (MWU, 85.000):
 - fixed (41 000)
 - free combinations (44 000): e.g. *morgen is er een vergadering over het bestemmingsplan van het industrieterrein* (tomorrow's meeting will be about the zoning plan for the industrial park))
- ❑ Excluded are **productive, compositional** derivations, compounds and multiword expressions => possibly included as examples

Fixed expressions

- **Grammatical collocations** (10.000): frequently co-occurring words; noun, verb or adjective with a function word.(e.g. *met betrekking tot* (*with regard to*), *te voet* (*by foot*), *uit woede* (*out of anger*)).
- **Lexical collocations** with lexical function (15.500): frequently co-occurring lexical words; set of lexical functions (loosely based upon Melcuk).(e.g. *zware roker* (*heavy smoker*); *een vergadering schorsen* (*adjourn a meeting*))
- **Lexical collocations** – not specified (4.500): cannot be expressed with a Melcuk function; reduced semantic transparency & ontological type often differs from the type of the head of the phrase, i.e. it is subsumed by the head of the phrase.(*systematische catalogus* (*systematic catalogue*), *open breuk* (*compound fracture*), *enkelvoudige breuk* (*simple fracture*))

Fixed expressions

- **Pragmatic formula** (1500): highly fixed, fully compositional expressions; fixed communicative situation. (*goedemorgen* (good morning), *geen paniek!* (no panic), *adres onbekend* (address unknown))
- **Idioms** (9500): expressions with less semantic transparency (e.g. *de kaarten zijn geschud*, *stoken in een goed huwelijk*, *een heilige koe*, *een vette kluif*)
- **Proverbs** (350): completely frozen sentences (e.g. *Pas als het kalf verdronken is, dempt men de put*)

Lexical Unit: polysemy

- Metonymy and metaphors:
 - lexical rules, encoded as shifts in RBN: school (institute) -> school (place)
- Complex Types (Pustejovsky):
 - Book incorporates different meanings layers: physical object & content
- Not a single LU but possible super-entries that group LUs that have a specific semantic relation (shifts in RBN)
- Motivation:
 - Semi-productive:
 - World-knowledge/culture-bound

walnut: refer to a wide range of objects and products that are related and derived from the walnut tree: fruit, wood, colour, taste, ice-cream, furniture, etc.

chestnut. similar object but less productive in our culture and therefore has less conventionalized meanings, even though anybody could productively create them at any time.

Lexical Unit: Meaning variants

- Roles, including male/female variants, e.g.:
 - theoloog/theologe
 - leraar/lerares
 - secretaresse vs. mannelijke secretaresse
 - kleuterleidster vs. ?
- Criteria for distinguishing different concepts:
 - If defined exhaustively as a role that is neutral with respect to male/female
AND
 - If the male/female form can be derived with a regular and compositional derivation
THEN 1 LU for the abstract neutral form.
 - In all other cases: separate LUs, possibly related to different ontology terms or a KIF expressions, depending on the ontology.
- Consequences:
 - Single LU for *theoloog*:
 - neutral form (possibly zero-derivation) for *man*.
 - neutral form or apply a derivational rule to create *theologe* when applied to *woman*
 - Separate LUs for *leraar* (male teacher) and *lerares* (female teacher), there is no neutral form. The same applies to *verpleger* (male nurse) and *verpleegster* (female nurse).

Combining resources

Combinatoriek

in een band spelen

een band oprichten

de band speelt

Combinatoriek

de band oppompen

een band plakken

een lekke band

de band springt

Combinatoriek

de band starten

op de band opnemen

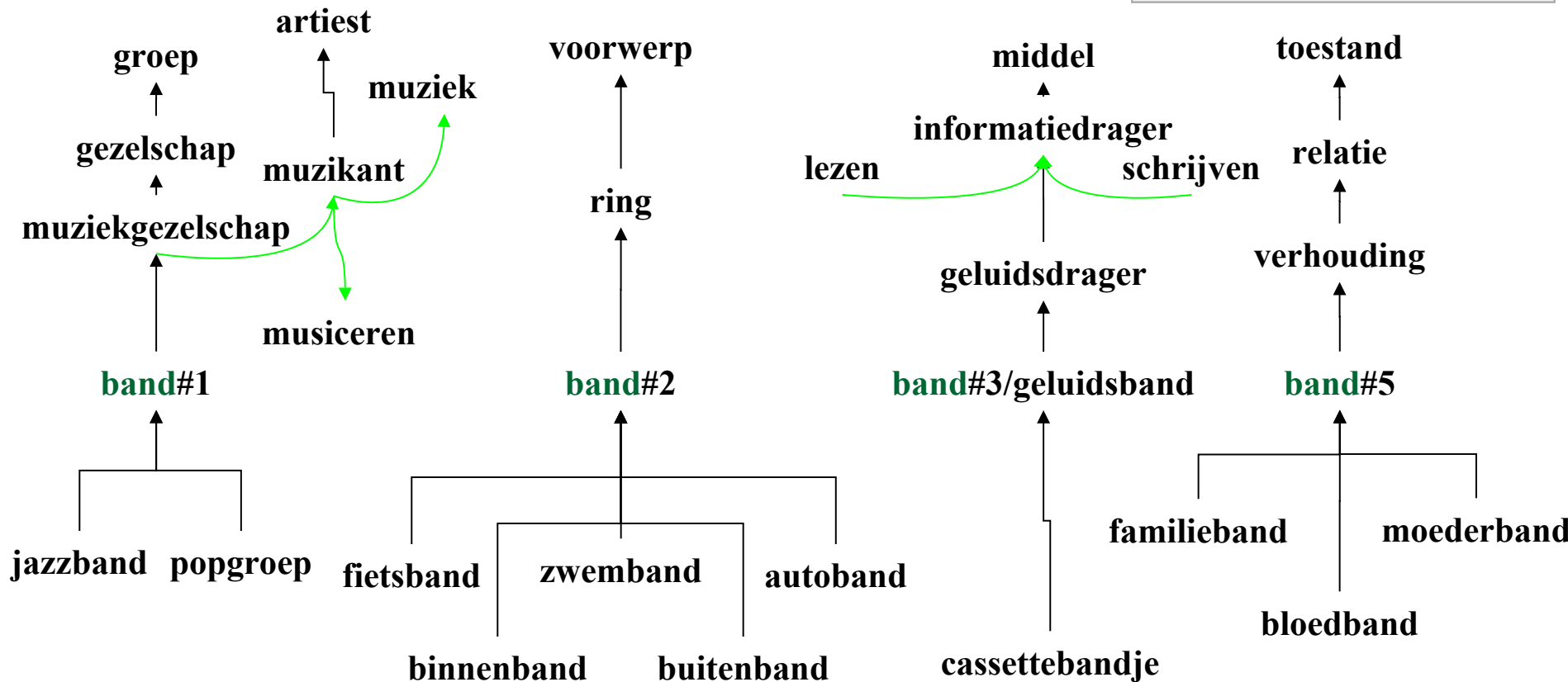
de band afspelen

Combinatoriek

een goede/sterke band

de banden verbreken

een band hebben met iemand



Disambiguating combinations

slag & leveren (fight)
slag & vrije (free style swimming)

punt & delen (share points)
punt & pakken (take a point)
punt & scoren (score a point)
punt & hard (firm point)

trap & geven (to kick)
trap & krijgen (be kicked)
trap & vrij (free kick)
trap & hoge (high stairs)

band & lekke (flat tire)
band & lopende (converter belt)
band & losse (loose relation)
band & spelen (music band plays)
band & starten (start a tape or belt)
band & sterke (strong relation)

bui & flinke (heavy shower)
bui & plaatselijk (local shower)
bui & vallen (there can be shower)
bui & slechte (bad mood)

Prototypical contexts

functie & bekleden (take a position)
functie & hoge (high position)
functie & neerleggen (put down position)
functie & ontheffen (discharge from position)
functie & openen (open up a position)
functie & treden (start a position)
functie & uitoefenen (carry out position)
functie & vervullen (fulfill position)

werk & aannemen (take on work)
werk & bieden (offer work)
werk & gaan & naar (go to work)
werk & half (incomplete work)
werk & staken (strike)
werk & zetten & aan (to put to work)
werk & zoeken (search for work)
werk & zwaar (heavy work)

Semantics for frame structures

- Event structure for verbs from RBN:
 - ❑ E: behandelen <e0> action
 - ❑ A1: <?a1> pers
 - ❑ A2: <?a2>
 - ❑ C3: <?c3> prep
 - ❑ iemand aan [zijn verwondingen] behandelen
 - ❑ een patiënt voor [een nieraandoening/puistje/keelpijn] behandelen
 - ❑ iemand met [fysiotherapie/medicijnen]_{Instrument} behandelen
- DWN:
 - ❑ [causes] [v] genezen:2, beteren:1, herstellen:1
 - ❑ [involved_agent] [n] arts:1; dokter:1 <?a1>
 - ❑ [involved_patient] [n] zieke:1; patiënt:1 <?a2>
 - ❑ [involved_instrument] [n] hart-longmachine:1 <?c3>
 - ❑ [involved_instrument] [n] mitella:1, draagdoek:1 <?c3>
 - ❑ [involved_instrument] [n] geneesmiddel:1; medicijn:1 <?c3>
 - ❑ etc...

Advantage of combining RBN & DWN

- Missing information can be derived from the other resource;
- Argument structures & frames of RBN can be enriched with semantics => WSD;
- Add hierarchy to combinatorics;
- Synsets in DWN are enriched with contextual information: => WSD;
- Synsets in DWN are enriched with lexical & grammatical selection criteria: => text generation;

Ontologize Cornetto

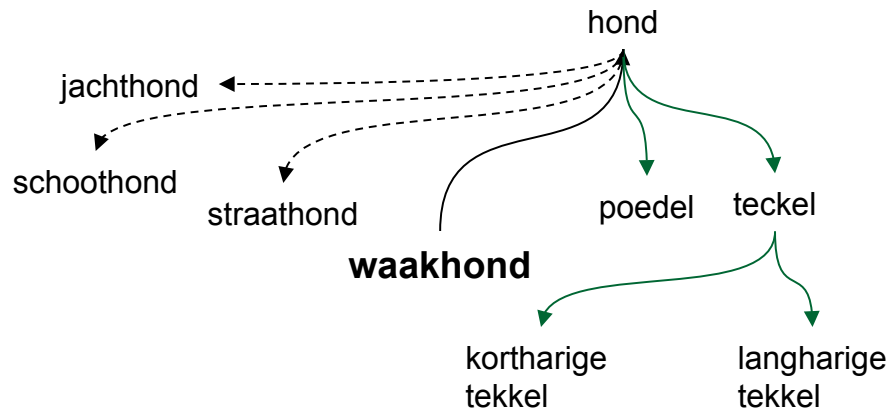
- Identity criteria OntoClean (Guarino & Welty 2002), :
 - **rigidity**: to what extent are properties true for entities in all worlds? You are always a *human*, but you can be a *student* for a short while.
 - **essence**: what properties are essential for an entity? *Shape* is essential for a *statue* but not for the *clay* it is made of.
 - **unicity**: what represents a whole and what entities are parts of these wholes? An *ocean* is a whole but the *water* it contains is not.
- Hyponyms of *hond* (dog) in DWN:
 - bokser; corgi; loboer; mopshond; pekinees; pointer; spaniël;
 - pup; reu; teef
 - bastaard; straathond; blindengeleidehond; bullebijter; diensthond; gashond; jachthond (hunting dog); lawinehond; schoothondje (lap dog); waakhond (watch dog)

Identity criteria applied to DWN

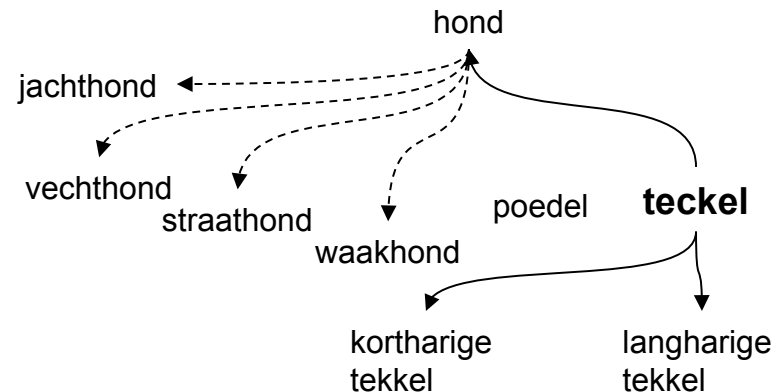
- (Semi-)rigid type hierarchy in the ontology:
 - Canine => PoodleDog; NewfoundlandDog; DalmatianDog, etc.
- Wordnet consists of names for (semi-)rigid dog-types and other words for dogs with roles:
 - poedel = PoodleDog
 - jachthond (?CAN)
 - ⇔ (and
(instance ?CAN Canine)
(instance ?EV Hunting)
(agent ?CAN ?EV)))
- Type hierarchy remains compact and pure

Expansion with pure hyponymy relations

Expansie van rol naar types



Expansie van types naar rollen



Advantage of ontologizing Cornetto?

- More explicit definition of hierarchical structures & roles: => **better expansion**;
- Apply semantic implications through relations results in more exact database;
- Enables inferencing;
- Enables the development of domain specific database: => **required for semantic web applications**;

Demo

Cornetto Synset

bas

Search Search in all

d_n-10187 [noun] bas:1, basstem:1 MUZ laagste mannenstem
d_n-15805 [noun] bas:2 MUZ baszanger, basspeler
d_n-10194 [noun] bas:3, baspartij:1 MUZ

Cornetto LU

bas

Search Search in all

r_n-6013 noun, bas:1, zangstem
r_n-6014 noun, bas:2, zanger
r_n-6015 noun, bas:3, contrabas
r_n-6016 noun, bas:4, basgitaar
r_n-6019 noun, bas:5

Cornetto Synset

basgitaar

Search Search in all

d_n-10189 [noun] basgitaar:1 MUZ als bas gestemde gitaar

Concept Relations Tree Revtree Edit Xml

C_SY_ID: d_n-10189
Synonyms: basgitaar:1/r_n-6026

PoS specific: NOUN_DE
Definition: als bas gestemde gitaar
Differentiae: laag, viersnarig
VLIS domains: MUZ
SUMO: (+, , Guitar)

-->> [HAS_HYPERONYM] basinstrument:1/d_n-21425
-->> [HAS_HYPERONYM] gitaar:1/d_n-28225
-->> [CO_INSTRUMENT_AGENT] basgitarist:1/d_n-15807
-->> [HAS_MERO_PART] bassnaar:1/d_n-27044

WordNet domains: music

Querying a dictionary is complete. Item(s): 1

Examples Edit Xml

LU_ID: r_n-6016 C_SEQ_NR: 4
[an] bas length: shortening

Phonology:

forms: bassen

countability: count type: artefact

(true)

as stemmen
excol

as:(Roland Knoppe)
racol

basinstrument

[*] d_n-21425 [noun] basinstrument:1 MUZ

Concept Relations Tree Revtree Edit Xml

C_SY_ID: d_n-21425
 Synonyms: basinstrument:1/d_n-40434

PoS specific: NOUN_HET
 Differentiae: voor de baspartij
 VLIS domains: MUZ
 SUMO: (+, , MusicalInstrument)

-->> [HAS_HYPERONYM] muziekinstrument:1, instrument:2/d_n-40434
 <<-- [HAS_HYPONYM] basbazuin:1/d_n-38269
 <<-- [HAS_HYPONYM] basklarinet:1/d_n-38275
 <<-- [HAS_HYPONYM] bombardon:1/d_n-21665
 <<-- [HAS_HYPONYM] contrabas:1, bas:5, basviool:1/d_n-3317
 <<-- [HAS_HYPONYM] fagot:1/d_n-22341
 <<-- [HAS_HYPONYM] basblokfluit:1/d_n-10188
 <<-- [HAS_HYPONYM] basgitaar:1/d_n-10189
 <<-- [CO_PATIENT_INSTRUMENT] basversterker:1/d_n-2143
 -->> [ROLE_INSTRUMENT] baspartij:1, bas:3/d_n-10194
 -->> [ROLE_INSTRUMENT] bassen:2/d_v-970

WordNet domains: music

bassen

d_v-6541 [verb] bassen:1, snauwen:1, bekken:1, blaffen:5, grauwen:2, katten:1, sneren:1
 d_v-970 [verb] bassen:2 MUZ een basinstrument bespelen

Concept Relations Tree Revtree Edit Xml

C_SY_ID: d_v-970
 Synonyms: bassen:2/d_v-22196

PoS specific: VERB_INTRANSITIVE
 Definition: een basinstrument bespelen
 Differentiae: een basgitaar
 VLIS domains: MUZ
 SUMO: (=, , InstrumentalMusic)

-->> [HAS_HYPERONYM] bespelen:1/d_v-1252
 -->> [INVOLVED_AGENT] basgitarist:1/d_n-15807
 -->> [INVOLVED_AGENT] bassist:1/d_n-21429
 -->> [INVOLVED_INSTRUMENT] basinstrument:1/d_n-21425
 -->> [INVOLVED_INSTRUMENT] basversterker:1/d_n-21431

WordNet domains: music

bas

d_n-10187 [noun] bas1, bas1toen1

d_n-15805 [noun] bas2, bas2toen2

d_n-10194 [noun] bas3, bas3toen3

d_n-33176 [noun] bas4, bas4toen4

Concept

Relations

Tree

C_SY_ID

Synonyms

PoS specific

Definition

Different

VLIS do

SUMO:

-->> [HAS_HYP

-->> [HAS_HYP

-->> [CO_INSTR

-->> [HAS_MER

WordNet

Querying a d

basgitaar

d_n-10189 [noun] basgit

Concept

Relations

Tree

C_SY_ID: d_n-10

Synonyms: basgita

PoS specific: NOU

Definition: als bas g

Differentiae: laag, v

VLIS domains: MU

SUMO: (+, , Guita

-->> [HAS_HYP

-->> [HAS_HYP

-->> [CO_INSTR

-->> [HAS_MER

WordNet domains

Querying a dictionary is complete.

DEBVisDic English WordNet

SYNSET.ID=ENG20-02705597-n

[n] bass:7

Preview

Tree

Revtree

Edit

Query

Xml

POS: n STAMP: ID: ENG20-0

Synonyms: bass:7

Definition: the member with the

instruments

Domain: music

SUMO/MILO: + MusicalInstru

-->> [hypernym] musical instrum

<<-- [hyponym] bass fiddle:1, b

bass:1, contrabass:1, string bass

<<-- [hyponym] bass guitar:1

<<-- [hyponym] bass horn:1, so

<<-- [hyponym] bombardon:2,

Querying a dictionary is complete.

DEBVisDic English Wor

bass guitar

[n] bass guitar:1

Preview

Tree

Revtree

Edit

Query

Xml

POS: n STAMP: ID: E

Synonyms: bass guitar:1

Definition: the lowest si

-->> [hypernym] bass:

-->> [hypernym] guitar

Querying a dictionary is complete.

Cornetto SUMO ontology

MusicalInstrument

Search

Search in all

MusicalInstrument

Preview

Tree

Revtree

Xml

Item

Physical

Object

SelfConnectedObject

CorpuscularObject

Artifact

Device

MusicalInstrument

HornInstrument

PercussionInstrument

StringInstrument

Guitar

Piano

Violin

WindInstrument

Querying a dictionary is complete. Found 0 item(s). Item(s): 1

Cornetto

Capita Selecta, INL Leiden, 22 mei 2007

Cornetto LU

bas Search Search in all

r_n-6013 noun, bas:1, zangstem
 r_n-6014 noun, bas:2, zanger
 r_n-6015 noun, bas:3, contrabas
 r_n-6016 noun, bas:4, basgitaar
 d_n-21942 noun, bas:5

Form Examples Edit Xml

C_LU_ID: r_n-6016 C_SEQ_NR: 4
 Examples:
 Example: 37756 [+]
 canonical form: een bas stemmen
 type: fixed subtype: lexcol

Example: 37757 [+]
 canonical form: op bas:(Roland Knoppe)
 type: fixed subtype: gracol

Example: 37758 [+]
 canonical form: bas spelen
 type: fixed subtype: lexcol

Querying a dictionary is complete. Item(s): 5

Cornetto LU

behandelen

r_v-1170 verb, behandelen:1, omgaan met iets of iemand

r_v-1171 verb, behandelen:2, bespreken

r_v-1172 verb, behandelen:3, trachten te genezen

d_v-24304 verb, behandelen:4

d_v-24305 verb, behandelen:5

d_v-520197 verb, behandelen:6

d_v-24306 verb, behandelen:7

d_v-24302 verb, behandelen:8

d_v-24303 verb, behandelen:9

Form

Examples

Edit

Xml

C_LU_ID:r_v-1170 C_SEQ_NR

Examples:

Example: 855 [+]

canonical form: als een normaal me

type: fixed subtype: pragma

Example: 856 [+]

canonical form: het personeel goed

type: free

Example: 857 [+]

canonical form: een klacht behande

type: free

Example: 858 [+]

canonical form: met zorg behande

type: fixed

Example: 859 [+]

canonical form: met respect behandelen

type: fixed

Cornetto LU

behandelen

r_v-1170 verb, behandelen:1, omgaan met iets of iemand

r_v-1171 verb, behandelen:2, bespreken

r_v-1172 verb, behandelen:3, trachten te genezen

d_v-24304 verb, behandelen:4

d_v-24305 verb, behandelen:5

Form

Examples

Edit

Xml

C_LU_ID:r_v-1171 C_SEQ_NR: 2

Examples:

Example: 1132 [+]

canonical form: de slag bij Warns behandelen in de geschiedenisles

type: free

Example: 1133 [+]

canonical form: uitputtend/grondig/uitgebreid behandelen

type: fixed subtype: lexcol

Querying a dictionary is complete.

Item(s): 9

Cornetto LU

behandelen

r_v-1170 verb, behandelen:1, omgaan met iets of iemand

r_v-1171 verb, behandelen:2, bespreken

r_v-1172 verb, behandelen:3, trachten te genezen

d_v-24304 verb, behandelen:4

d_v-24305 verb, behandelen:5

Form

Examples

Edit

Xml

C_LU_ID:r_v-1171 C_SEQ_NR: 2

Examples:

Example: 1132 [+]

canonical form: de slag bij Warns behandelen in de geschiedenisles

type: free

Example: 1133 [+]

canonical form: uitputtend/grondig/uitgebreid behandelen

type: fixed subtype: lexcol

Querying a dictionary is complete.

Item(s): 9

Querying a dictionary is complete.

Item(s): 9

ru | STEVIN

46

Cornetto LU

behandelen

- r_v-1170 verb, behandelen:1, omgaan met iets
- r_v-1171 verb, behandelen:2, bespreken
- r_v-1172 verb, behandelen:3, trachten te genezen
- d_v-24304 verb, behandelen:4
- d_v-24305 verb, behandelen:5
- d_v-520197 verb, behandelen:6
- d_v-24306 verb, behandelen:7
- d_v-24302 verb, behandelen:8
- d_v-24303 verb, behandelen:9

Form Examples Edit Xml

C_LU_ID: r_v-1172 C_SEQ_NR: 3
[verb] behandelen

Morphology:

type: derivation flex conjugation: type
flex mode: inf flex tense: ntense flex

Syntax:

trans: tran separ: onsch class: main
complementation: np, pp [...] np
iemand voor /met iets behandelen

Semantics:

type: action
resume: trachten te genezen
definition: XVD: als arts een patiënt
hypernyms: verrichten

Querying a dictionary is complete.

Cornetto LU

behandelen

- r_v-1170 verb, behandelen:1, omgaan met iets of iemand
- r_v-1171 verb, behandelen:2, bespreken
- r_v-1172 verb, behandelen:3, trachten te genezen
- d_v-24304 verb, behandelen:4
- d_v-24305 verb, behandelen:5
- d_v-520197 verb, behandelen:6
- d_v-24306 verb, behandelen:7
- d_v-24302 verb, behandelen:8
- d_v-24303 verb, behandelen:9

Form Examples Edit Xml

```
- <cdb_lu c_seq_nr="3" type="swu" is_complete="true" c_lu_id="r_v-1172">
  <form form-cat="verb" form-spelling="behandelen"/>
  <morphology_verb>
    <morpho-type>derivation</morpho-type>
    <flex-conjugation>
      <flex-conjugationtype>regular</flex-conjugationtype>
    </flex-conjugation>
    <flex-mode>inf</flex-mode>
    <flex-tense>ntense</flex-tense>
    <flex-number>nnumber</flex-number>
    <flex-person>nperson</flex-person>
  </morphology_verb>
  <syntax_verb>
    <sy-trans>tran</sy-trans>
    <sy-separ>onsch</sy-separ>
    <sy-class>main</sy-class>
    <sy-peraux>h</sy-peraux>
    <sy-valency>tri</sy-valency>
    <sy-reflexiv>nrefl</sy-reflexiv>
    <sy-subject>pers</sy-subject>
    <sy-complementation>
```

The image displays three overlapping windows of the Cornetto Synset application. The top window shows a search for 'n_n-503673 [noun] bergwater:1'. The middle window shows a search for 'bluswater' with results for 'd_n-16013 [noun]'. The bottom window shows a search for 'drinkwater' with detailed results for 'd_n-16556 [noun] drinkwater:1'.

Cornetto Synset
 cdb_synset.@c_sy_id=n_n-503673
 n_n-503673 [noun] bergwater:1 AARDR

Cornetto Synset
 bluswater
 d_n-16013 [noun]

Cornetto Synset
 drinkwater
 d_n-16556 [noun] drinkwater:1 CUL water dat geschikt is om gedronken te worden

Concept Relations
 C_SY_ID: n
 Synonyms: be
 PoS specific:
 Differentiae: u
 VLIS domain:
 SUMO: (insta
 -->> [HAS_H
 -->> [HAS_H
 Querying a dictiona

Concept Relations Tree Revtree Edit Xml
 C_SY_ID: d_n-16556
 Synonyms: drinkwater:1/r_n-11588
 PoS specific: NOUN_HET
 Definition: water dat geschikt is om gedronken te worden
 Differentiae: geschikt voor consumptie
 VLIS domains: CUL
 SUMO: (+, , Water) , (instance, 0, Water) , (instance, 1, Drinking) , (patient, 0, 1)
 -->> [HAS_HYPERONYM] water:1, majem:1/d_n-20623
 WordNet domains: chemistry
 Querying a dictionary is complete. Item(s): 1

Strategies of work

Editing strategies

- Macro Alignment: LUs vs Synsets
- Top-down classification of ontology types
- Lexical units:
 - examples and frames
- Synsets
 - add relations
 - ontological types
- Violation of semantic constraints
- Domain clustering

Alignment: select critical subsets

- ❑ LUs with low score;
- ❑ LUs without synsets:
 - with synonyms
 - without synonyms
- ❑ LUs with many alternative mappings to Synsets
- ❑ LUs of words with many meanings
- ❑ LUs with type-shift label
- ❑ LUs with information that conflicts information of correlated synsets:
 - LUs with synonyms that do not match correlated synsets
 - LUs with hyperonym that does not match hyperonym of correlated synsets
 - LUs with mismatching domain
 - etc..

Ontology

- ❑ Traverse synsets through ontology hierarchy top-down
- ❑ Synsets with conflicting ontology relations
- ❑ Synsets with conflicting ontology relations after inheritance
- ❑ Synsets with many or multiple ontology terms
- ❑ Synsets without ontology mapping and that occupy important positions in the hyperonym chain

Lexical Units

- ❑ Select:
 - LUs with specific morpho-syntactic features, e.g. di-transitive verbs, verbs with certain complementation constraints, collective nouns, etc.
 - LUs lacking certain information types
 - LUs with many examples
 - LUs without examples
 - Multiword units
- ❑ Fix examples and frames

Synsets

❑ Select:

- Synsets without hyperonyms
- Synsets with many synonyms
- Synsets with many meronyms
- Cyclic synset relations
- Synsets at certain levels of hyperonyms: high, low

❑ Fix:

- semantic relations
- ontology mapping

Constraint violations

- ❑ LUs/Synsets that violate constraints directly
 - Within the same information type
 - Across information types
- ❑ LUs/Synsets that violate constraints through inheritance
 - hyponymy
 - antonymy
 - meronymy
 - ontology: types & roles; instances

Domains

 Select:

- Synsets without Domains
- Synsets with Factotum domain
- Synsets with many domains

 Fix:

- horizontal relations
- frames and examples





The end....