

The Cornetto Database

=> *Vrije Universiteit Amsterdam, Faculteit der Letteren*

=> *Universiteit van Amsterdam, CWI*

=> *Irion Technologies*



CLIN-17, 12 January 2007, Leuven

nū | STEVIN

Overview

1. Project background information
2. Alignment of lexical resources
3. Database design

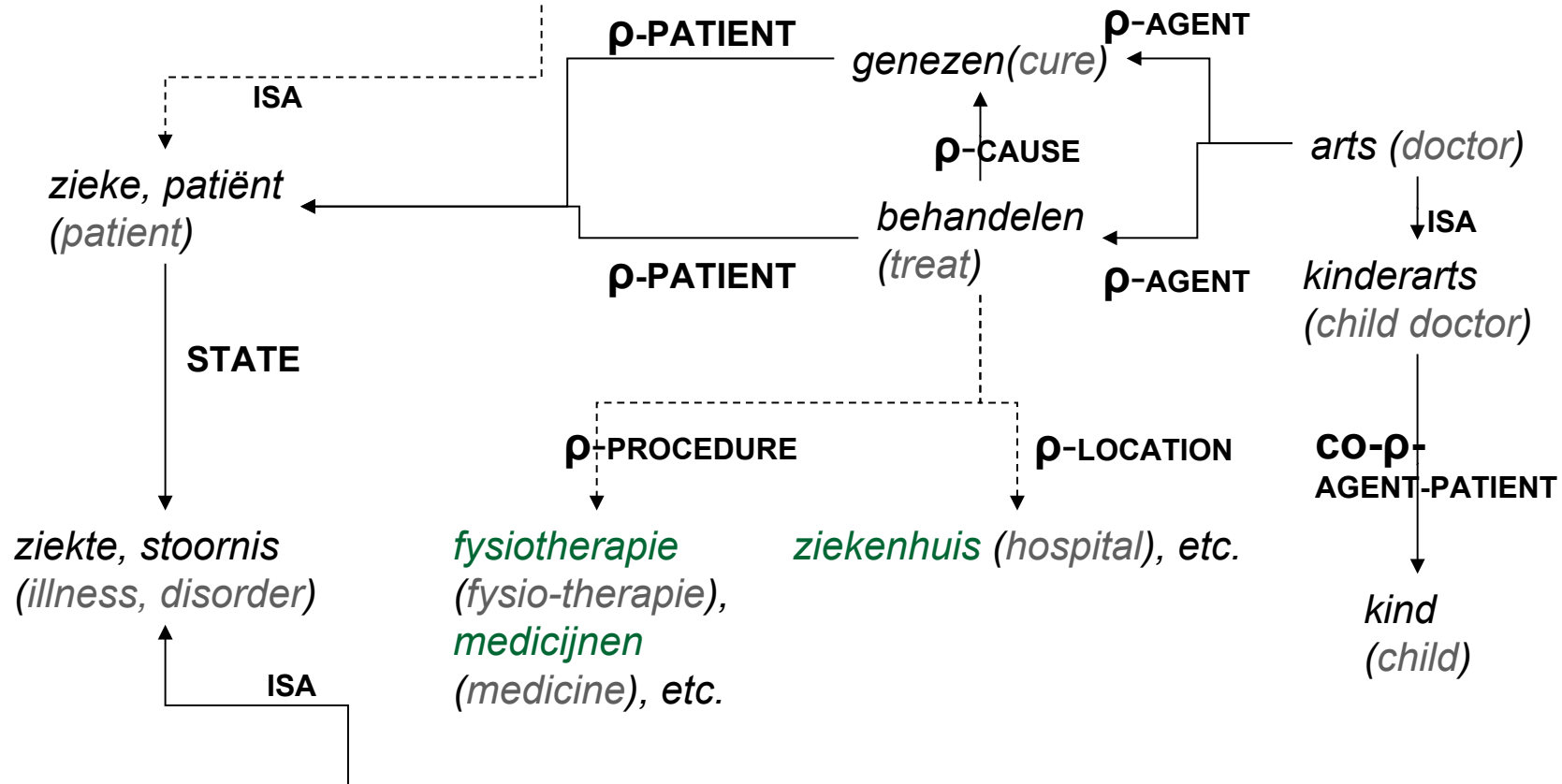


Cornetto background

- Stevin tender project 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
- Consortium: Vrije Universiteit Amsterdam, Universiteit Amsterdam, Universiteit Leuven, Irion Technologies
- Started April 2006, ends March 2008
- Licensed from TST-centrale, Nederlandse Taalunie
- <http://www.let.vu.nl/onderzoek/projectsites/cornetto/start.htm>

Horizontal & vertical semantic relations

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

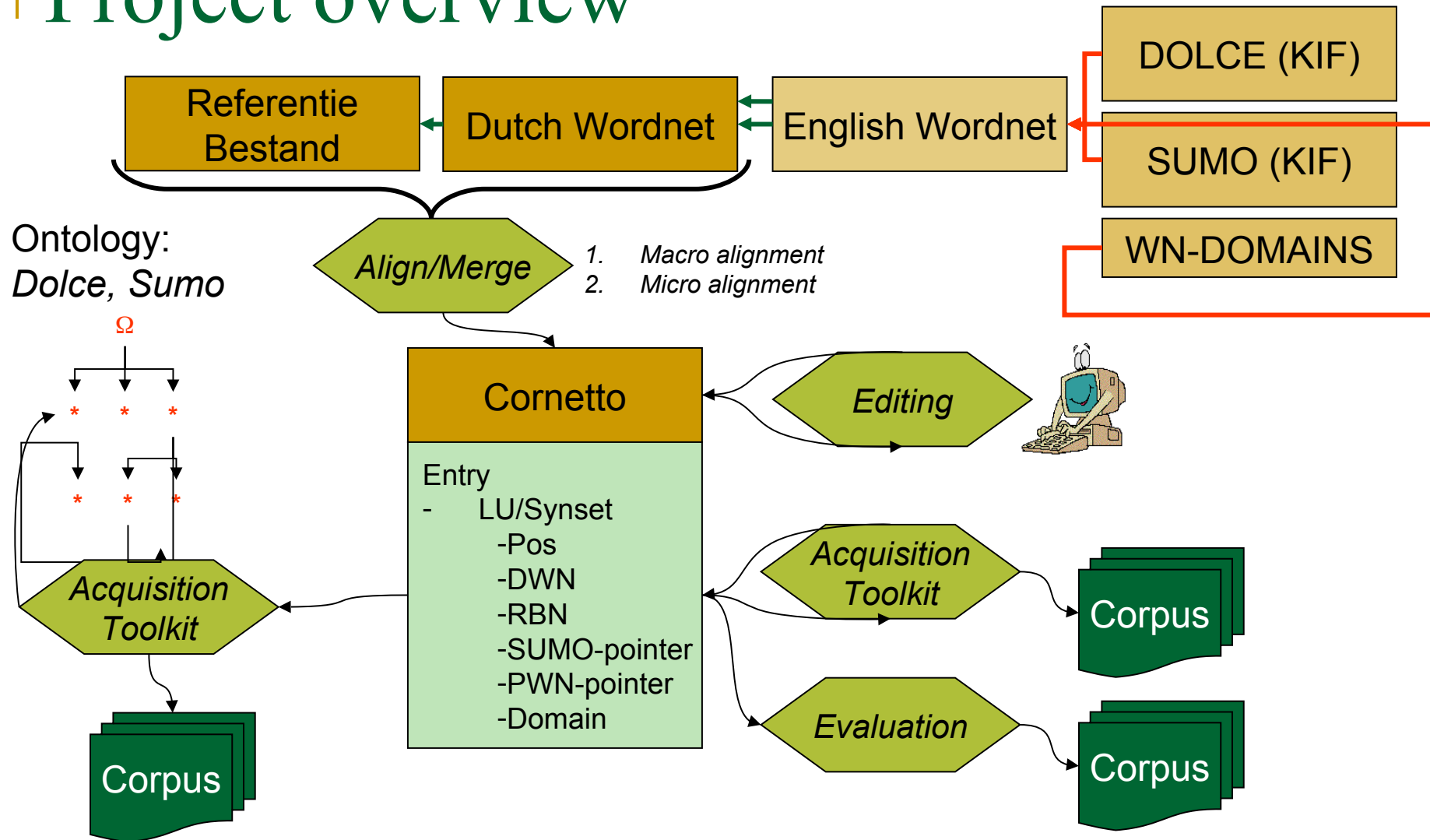


maagaandoening (stomach disorder)
nieraandoening (kidney disorder), *keelpijn* (sour throat).

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>een ziekte behandelen (a disease treat)</i>

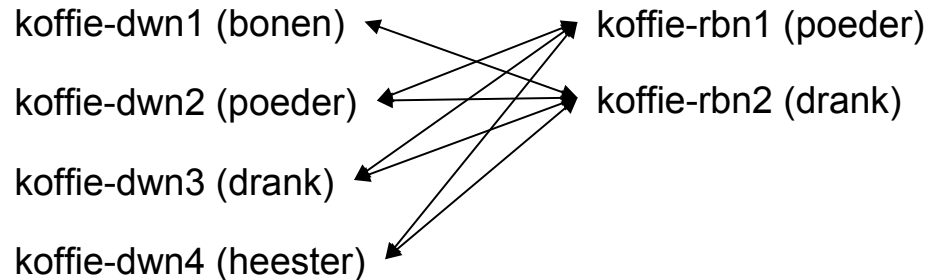
Project overview



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)

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		

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;

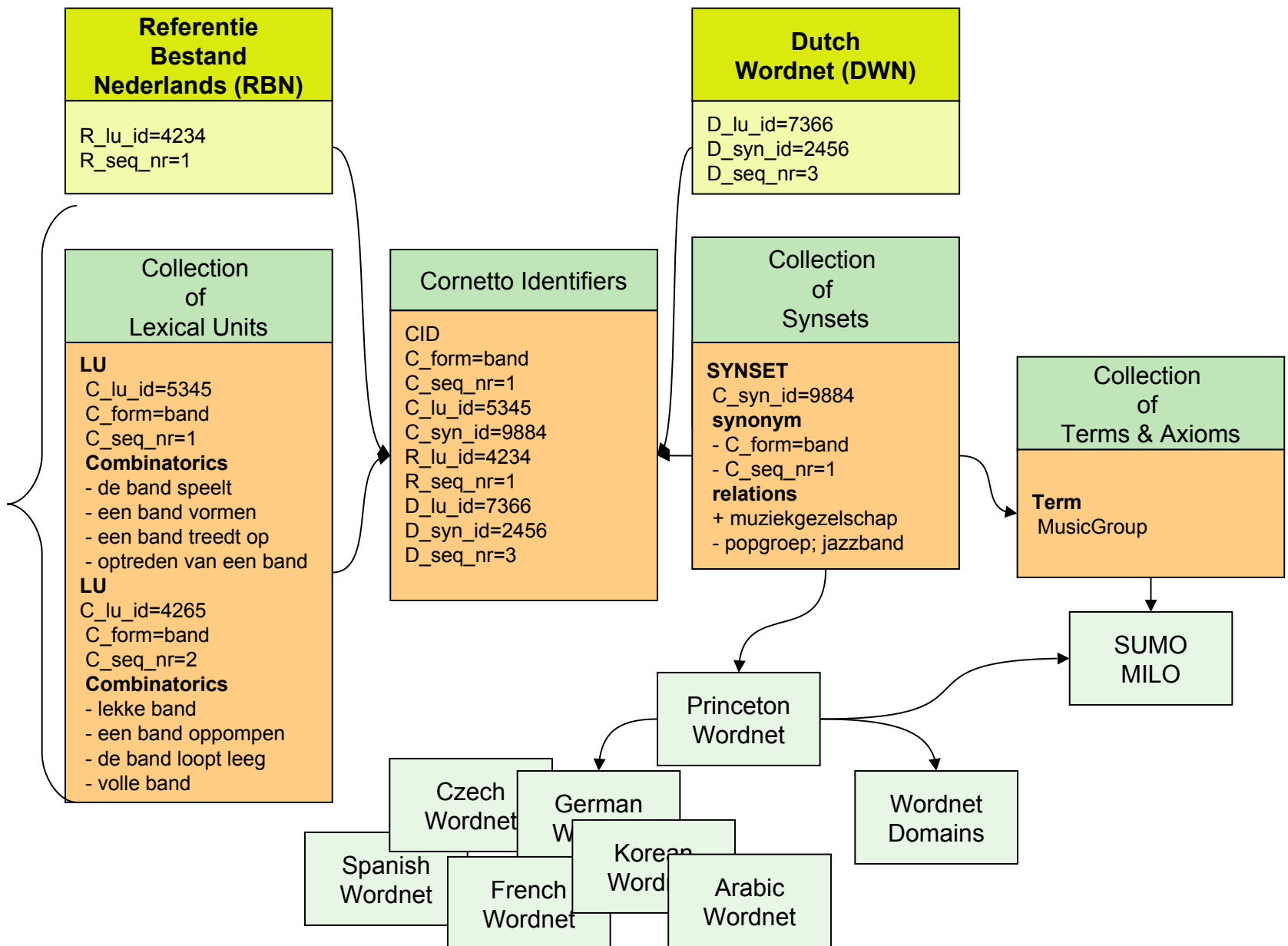
Data structure overview

■ 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



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

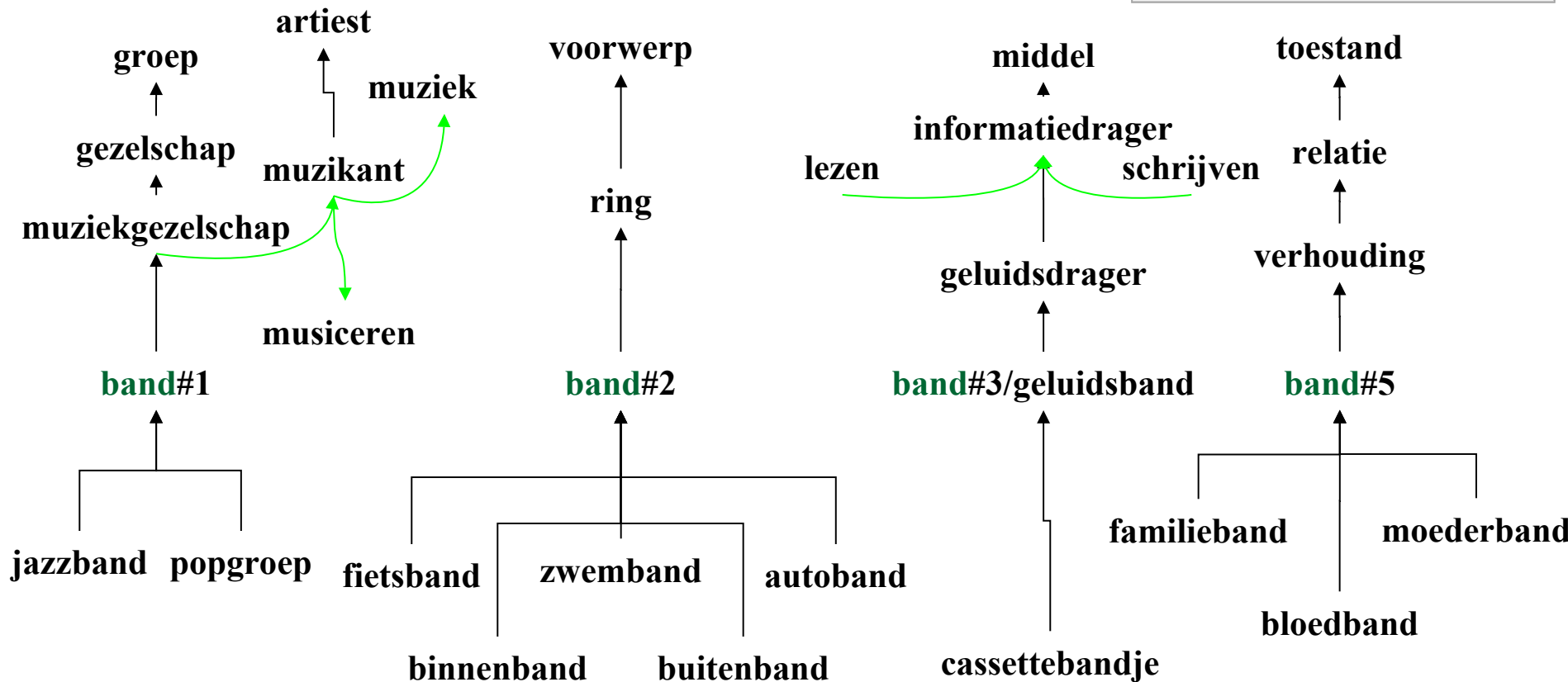
de band afspelen

Combinatoriek

een goede/sterke band

de banden verbreken

een band hebben met iemand



Semantics for frame structures

- Event structure for verbs from RBN:
 - ❑ E: behandelen <e0> action
 - ❑ A1: <a1> pers
 - ❑ A2: <a2> pers
 - ❑ 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...

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; loboor; 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)
 - ⇔ (exists (?CAN ?EV)
(and
(instance ?CAN Canine)
(instance ?EV Hunting)
(agent ?CAN ?EV)))
- Type hierarchy remains compact and pure

Cornetto LU

behandelen

Search

Search in all

DEBVisDic English WordNet

treat

Search

Search in all

POS: v STAMP: ID: ENG20-00076458-v BCS: 3

Synonyms: treat:3, care for:3

Definition: provide treatment for

Usage: The doctor treated my broken leg

Usage: The nurses cared for the bomb victims

Usage: The patient must be treated right away or she will die

Usage: Treat the infection with antibiotics

Domain: medicine

SUMO/MILO: = TherapeuticProcess

--> [eng_derivative] treatment:1

<--- [hyponym] hyperventilate:1

<--- [hyponym] massage:2

<--- [hyponym] purge:7

<--- [hyponym] correct:8

<--- [hyponym] insufflate:2

<--- [hyponym] detoxify:2

<--- [hyponym] irrigate:2

<--- [hyponym] iodize:2, iodise:2

<--- [hyponym] doctor:2

<--- [hyponym] vet:3

<--- [hyponym] nurse:1

<--- [hyponym] manipulate:6

<--- [hyponym] administer:4, dispense:3

<--- [subevent] bring around:2, cure:1, heal:3

<--- [hyponym] remedy:2, relieve:8

<--- [hyponym] dress:14

<--- [hyponym] splint:1

<--- [hyponym] operate on:1, operate:7

Cornetto Synset

behandelen

Search

Search in all

verb] behandelen:2 bespreken of beschrijven

verb] behandelen:3 op de vereiste wijze bewerken

verb] behandelen:4, bejegenen:1 op een bepaalde wijze omgaan met (iem.)

MED.T (een patiënt) als arts trachten te genezen van een ziekte

behandelen:1, omgaan:7

behandelen:6 pleiten voor

Tree

Revtree

Edit

Xml

72; 83.1

638; 83.1

72; 64.6

88; 64.6

72; 83.8333

72; 79.95

72; 91.6

72; 64.6

88; 64.6

72; 64.6

88; 64.6

+ d_980|beademen:1, verb C=0/0 D=false

+ d_1243|besnijden:2, verb C=0/0 D=false

+ d_1282|bestralen:2, verb C=0/0 D=false

+ d_1728|castreren:1, ontmannen:1, snijden:7, verb C=1/1 D=false

+ d_1856|curetteren:1, verb C=0/0 D=false

+ d_1899|defibrilleren:1, verb C=0/0 D=false

+ d_1964|dialyseren:1, verb C=0/0 D=false

+ d_2317|dotteren:1, verb C=0/0 D=false

+ d_3201|injecteren:1, injiciëren:1, inspuiten:1, prikken:4, verb C...

+ d_4124|masseren:1, verb C=0/0 D=false

+ d_4903|ontwormen:1, verb C=0/0 D=false

+ d_5044|opereren:2, verb C=0/0 D=false

+ d_6000|reponeren:1, verb C=0/0 D=false

+ d_6754|steriliseren:1, verb C=2/2 D=false

+ d_6815|stoppen:7, verb C=0/0 D=false

+ d_7824|verbinden:3, verb C=2/2 D=false

Querying a dictionary is complete. Found 16 item(s).

text category: vb

2007, Leuven

ru | STEVIN

19

behandelen

Search

Search in all

verb, behandelen:1
 verb, behandelen:1, omgaan met iets of iemand
 verb, behandelen:2
 verb, behandelen:2, bespreken
 verb, behandelen:3
 verb, behandelen:3, trachten te genezen
 verb, behandelen:4
 verb, behandelen:5
 verb, behandelen:6

Form Syn-Sem Examples Edit Xml

New

Delete

Save

Clear form

Lu id

r_v-1172

Seq. nr.

3

Form

Spelling behandelen

Length

Spelvar

Morphology

Type derivation

Structure

Flex mode inf

Tense

ntense

Number nnumber

Person

nperson

Flex conjugation

Add

regular

Remove

Syntax

Trans Tran

Separ Onsch

Class Main

Peraux

H

Valency Tri

Reflexiv

Nrefl

Subject Pers

Complementation

Add

nil

Remove



behandelen



Search

Search in all

verb, behandelen:1
verb, behandelen:1, omgaan met iets of iemand
verb, behandelen:2
verb, behandelen:2, bespreken
verb, behandelen:3
verb, behandelen:3, trachten te genezen
verb, behandelen:4
verb, behandelen:5
verb, behandelen:6

Form Syn-Sem Examples Edit Xml

```
- <cdb_lu c_seq_nr="3" type="swu" 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>
      <sy-comp>np</sy-comp>
      <sy-comp>pp [...] np</sy-comp>
      <sy-compl-text>iemand voor /met iets behandelen</sy-compl-text>
    </sy-complementation>
  </syntax_verb>
  - <semantic_verb>
```

behandelen

Search

Search in all

[verb] behandelen:2 bespreken of beschrijven

[verb] behandelen:3 op de vereiste wijze bewerken

[verb] behandelen:4, bejegenen:1 op een bepaalde wijze omgaan met (iem.)

[verb] behandelen:5 MED_T (een patiënt) als arts trachten te genezen van een ziekte

[verb] hanteren:1, behandelen:1, omgaan:7

[verb] verdedigen:2, behandelen:6 pleiten voor

Concept

Relations

Tree

Revtree

Edit

Xml

New

Delete

Save

Clear form

Synset Id

d_1062

☐ Base concept

Synonyms: Cid Id, Lu Id, Status

Add

168519

verb, behandelen:5

Remove

Domain Relations

Add

Semantic Relations: Target, Relation

Add

[noun] arts:1, dokter:1, esculaap:2, geneesheer:1,

INVOLVED_AGENT

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[verb] beademen:1 | MED_T | (een patiënt) kunstmat

HAS_HYPONYM

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[noun] behandeling:2 | | het omgaan met iets

XPOS_NEAR_SYNONYM

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[verb] besnijden:2 | MED_T | de voorhuid wegnemen

HAS_HYPONYM

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[verb] bestralen:2 | CUL, IND | produkten met behulp

HAS_HYPONYM

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[verb] castreren:1, ontmannen:1, snijden:7 | MED_T

HAS_HYPONYM

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

[verb] curetteren:1 | MED_T | (de baarmoeder) scho

HAS_HYPONYM

Remove

☐ Coordinative ☐ Disjunctive ☐ reversed ☐ Factive ☐ Negative

Querying a dictionary is complete.

behandelen		Search	Search in all
[verb] behandelen:2	bespreken of beschrijven		
[verb] behandelen:3	op de vereiste wijze bewerken		
[verb] behandelen:4, bejegenen:1	op een bepaalde wijze omgaan met (iem.)		
[verb] behandelen:5	MED.T (een patiënt) als arts trachten te genezen van een ziekte		
[verb] hanteren:1, behandelen:1, omgaan:7			
[verb] verdedigen:2, behandelen:6	pleiten voor		

Concept Relations Tree Revtree Edit Xml

```
- <cdb_synset c_sy_id="d_1062" count_childs="22" d_synset_id="d_1062">
- <synonyms>
  <synonym status="" c_cid_id="168519" form="behandelen" c_lu_id="d_24306"/>
</synonyms>
- <wn_internal_relations>
- <relation factive="" reversed="false" relation_name="INVOLVED_AGENT" negative="false" coordinative="false" disjunctive="false" target="d_10091">
  <author name="Piek" score="0.0" status="YES" date="19961217" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_980">
  <author name="Paul" score="0.0" status="YES" date="19961206" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="XPOS_NEAR_SYNONYM" negative="false" coordinative="false" disjunctive="false" target="d_15850">
  <author name="Piek" score="0.0" status="YES" date="19961129" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_1243">
  <author name="Paul" score="0.0" status="" date="19961206" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_1282">
  <author name="Paul" score="0.0" status="" date="19961206" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_1728">
  <author name="Paul" score="0.0" status="" date="19961206" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_1856">
  <author name="Paul" score="0.0" status="" date="19961206" source_id="d_1062"/>
</relation>
- <relation factive="" reversed="false" relation_name="HAS_HYPONYM" negative="false" coordinative="false" disjunctive="false" target="d_1899">
  <author name="Paul" score="0.0" status="" date="19961206" source_id="d_1062"/>
</relation>
```

Next steps

- Done:
 - Macro alignment & database
- In progress:
 - Editing
 - Revising critical alignments
 - Defining ontology constraints
 - Revising word meanings based on ontology distinctions
 - Revising ontology assignment
 - Micro-level alignment
 - Automatic acquisition
 - Task-based evaluation



The end....