

The Cornetto Database

Content design and progress



RU | STEVIN

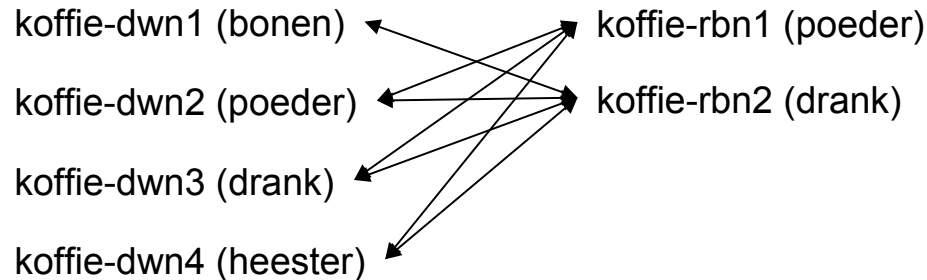
Overview

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



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)



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;

Cornetto Database (CDB)

Referentie Bestand Nederlands (RBN)

R_lu_id=4234
R_seq_nr=1

Dutch Wordnet (DWN)

D_lu_id=7366
D_syn_id=2456
D_seq_nr=3

Collection of Lexical Units

LU
C_lu_id=5345
C_form=band
C_seq_nr=1
Combinatorics
- de band speelt
- een band vormen
- een band treedt op
- optreden van een band

LU
C_lu_id=4265
C_form=band
C_seq_nr=2
Combinatorics
- leuke band
- een band oppompen
- de band loopt leeg
- volle band

Cornetto Identifiers

CID
C_form=band
C_seq_nr=1
C_lu_id=5345
C_syn_id=9884
R_lu_id=4234
R_seq_nr=1
D_lu_id=7366
D_syn_id=2456
D_seq_nr=3

Collection of Synsets

SYNSET
C_syn_id=9884
synonym
- C_form=band
- C_seq_nr=1
relations
+ muziekgezelschap
- popgroep; jazzband

Collection of Terms & Axioms

Term
MusicGroup

SUMO MILO

Princeton Wordnet

Wordnet Domains

Czech Wordnet
German Wordnet
Korean Wordnet
Arabic Wordnet
French Wordnet
Spanish Wordnet



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 kluiif*)
- **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.

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;

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
 [n] bas length: shortening

Morphology:

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

Cornetto Synset

bas

d_n-10187 [noun] bas1, bas1toen1
d_n-15805 [noun] bas1, bas1toen1
d_n-10194 [noun] bas1, bas1toen1
d_n-33176 [noun] bas1, bas1toen1

Concept Relations Tree

C_SY_ID: d_n-10187
Synonyms: basgitaar

PoS specific: NOU
Definition: als bas g
Differentiae: laag, v
VLIS domains: MU
SUMO: (+, , Guita

-->> [HAS_HYP
-->> [HAS_HYP
-->> [CO_INSTR
-->> [HAS_MER

WordNet

Querying a d

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.

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 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

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

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

ru | STEVIN

19

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 foreground window is titled 'Cornetto Synset' and shows the search results for the term 'drinkwater' (d_n-16556). The search bar contains 'drinkwater' and the results list shows 'd_n-16556 [noun] drinkwater:1 CUL water dat geschikt is om gedronken te worden'. The main content area displays the following information:

- 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)
- Relations:** -->> [HAS_HYPERONYM] water:1, majem:1/d_n-20623
- WordNet domains:** chemistry

The status bar at the bottom of the foreground window indicates 'Querying a dictionary is complete.' and 'Item(s): 1'. The background windows show other synsets, such as 'n_n-503673 [noun] bergwater:1 AARDR' and 'bluswater' (d_n-16013 [noun]).

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..

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

Editing

Three editing strategies

Macro alignment: LU's vs. synsets

- ❑ Frequent verbs
- ❑ Nouns with a semantic shift in RBN
- ❑ Adjectives

Editing

LU's vs. synsets:

1. frequent verbs

Editing

LU's vs. synsets:

2. Nominal LU's with semantic shifts

Semantic Shifts

A (peripheral) aspect of a meaning that is close connected with the central meaning.

□ 'Article' (a text in e.g. a magazine) central meaning *He wrote an article*

□ 'Article' (the pages on which it is printed) aspect
He burned the article

A shift can be seen as an extension of a meaning (RBN) or, with another approach, as a different meaning (DWN).

A semantic shift in a LU predicts a shift from one semantic type to another semantic type.

Types of shifts

- There are 29 different types of shifts in the RBN.

Some examples for nouns:

- | | | |
|----------------------------------|------------------|-----------------------|
| □ Human → Place | <i>bakker</i> | <i>(baker/bakery)</i> |
| □ Dynamic → Nondynamic | <i>adaptatie</i> | <i>(adaptation)</i> |
| □ Nonhuman → Concrother | <i>vijg</i> | <i>(fig)</i> |
| □ Institution → Place → Artefact | <i>school</i> | <i>(school)</i> |

Cornetto Synset

school Search Search in all

d_n-36313 [noun] school:3 ONDERW inrichting waar onderwijs gegeven v
 d_n-41924 [noun] school:4, schoolgebouw:1 ONDERW, GEBOUW gebouw
 d_n-13847 [noun] school:5 KUNST, WET groep personen, in de kunst of v
 d_n-19465 [noun] school:6 DIERK aaneengesloten menigte vissen

Concept Relations Tree Revtree Edit Xml

C_SY_ID: d_n-36313
 Synonyms: school:3/r_n-33122

PoS specific: NOUN_DE
 Definition: inrichting waar onderwijs gegeven wordt
 Differentiae: algemeen
 VLIS domains: ONDERW
 SUMO: (=, , School)

Cornetto LU

school Search Search in all

r_n-33120 noun, school:1, onderwijsinstelling
 r_n-33121 noun, school:2, groep vissen
 r_n-33122 noun, school:3, richting in kunst of wetenschap
 d_n-254617 noun, school:4
 d_n-254621 noun, school:5
 d_n-254624 noun, school:6

Form Examples Edit Xml

C_LU_ID: r_n-33120 C_SEQ_NR: 1
 [noun] school

Morphology:
 type: simpmorph plurforms: scholen

Syntax:
 gender: mf article: de

Semantics:
 reference: common countability: count type: institut
 shift: 4:inst>pl>art subclass: institutie
 resume: onderwijsinstelling

Guidelines

- ❑ Split LU's with a semantic shift if there is (ontological) evidence to do so.
- ❑ Editing a LU with a semantic shift means editing the whole entry.
- ❑ Map the new, splitted LU to a corresponding synset, or create a new synset.
- ❑ Split peripheral shifts if there is a corresponding synset.
- ❑ Peripheral shifts with no corresponding synset are evaluated on frequency and their degree of independency to the central meaning.

Benefits

- ❑ Enrichment of both synsets and LU's
- ❑ Shift can be used as pointer to create new synsets
- ❑ Shifts can predict the semantic relation between synsets
- ❑ Consistency in both synsets and LU's



The end....