

Cornetto



Vrije Universiteit Amsterdam

ru | STEVIN

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

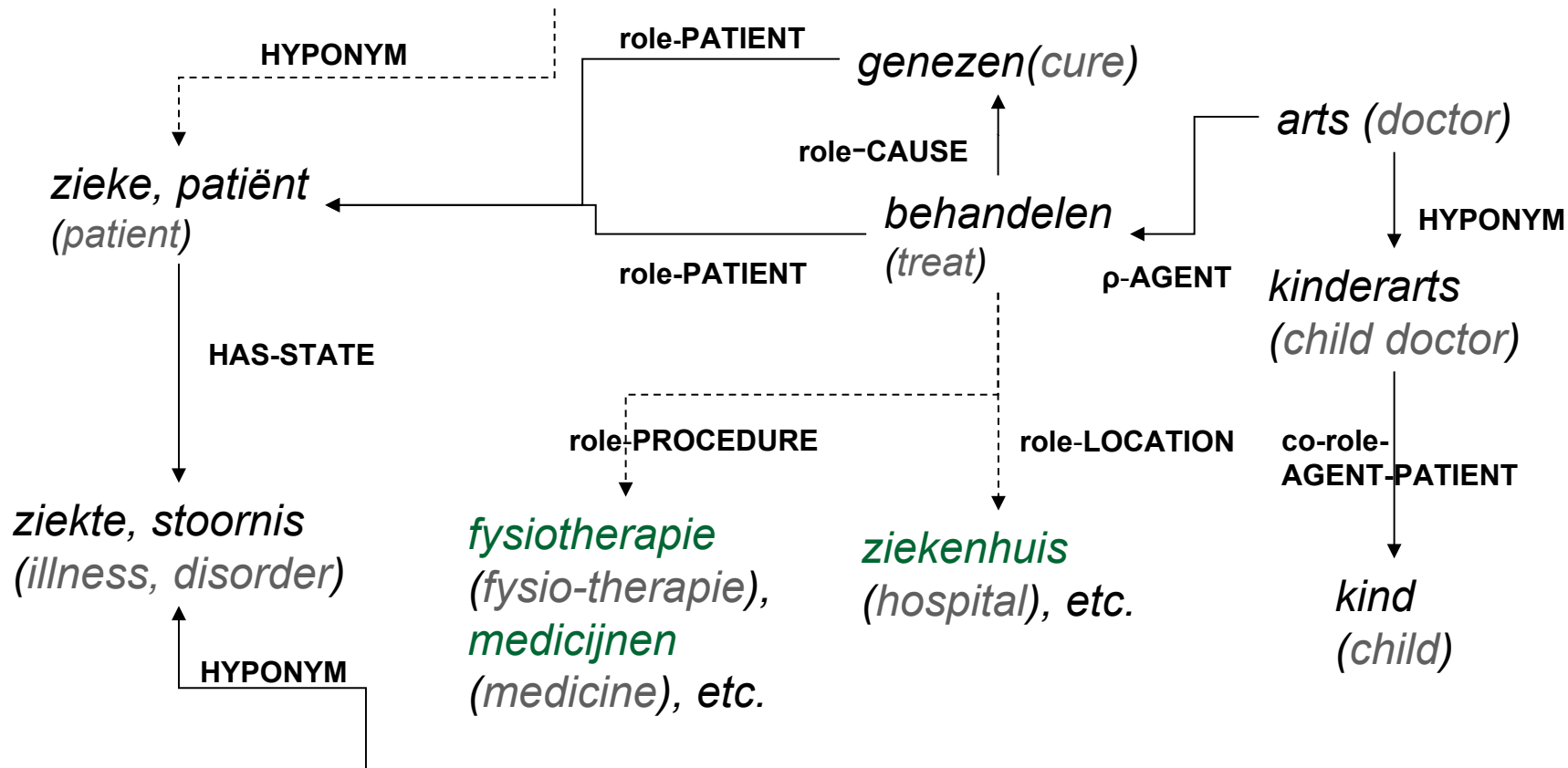
Consortium

- Vrije Universiteit Amsterdam, Faculteit der Letteren, Language, Cognition, and Communication:
 - 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



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).

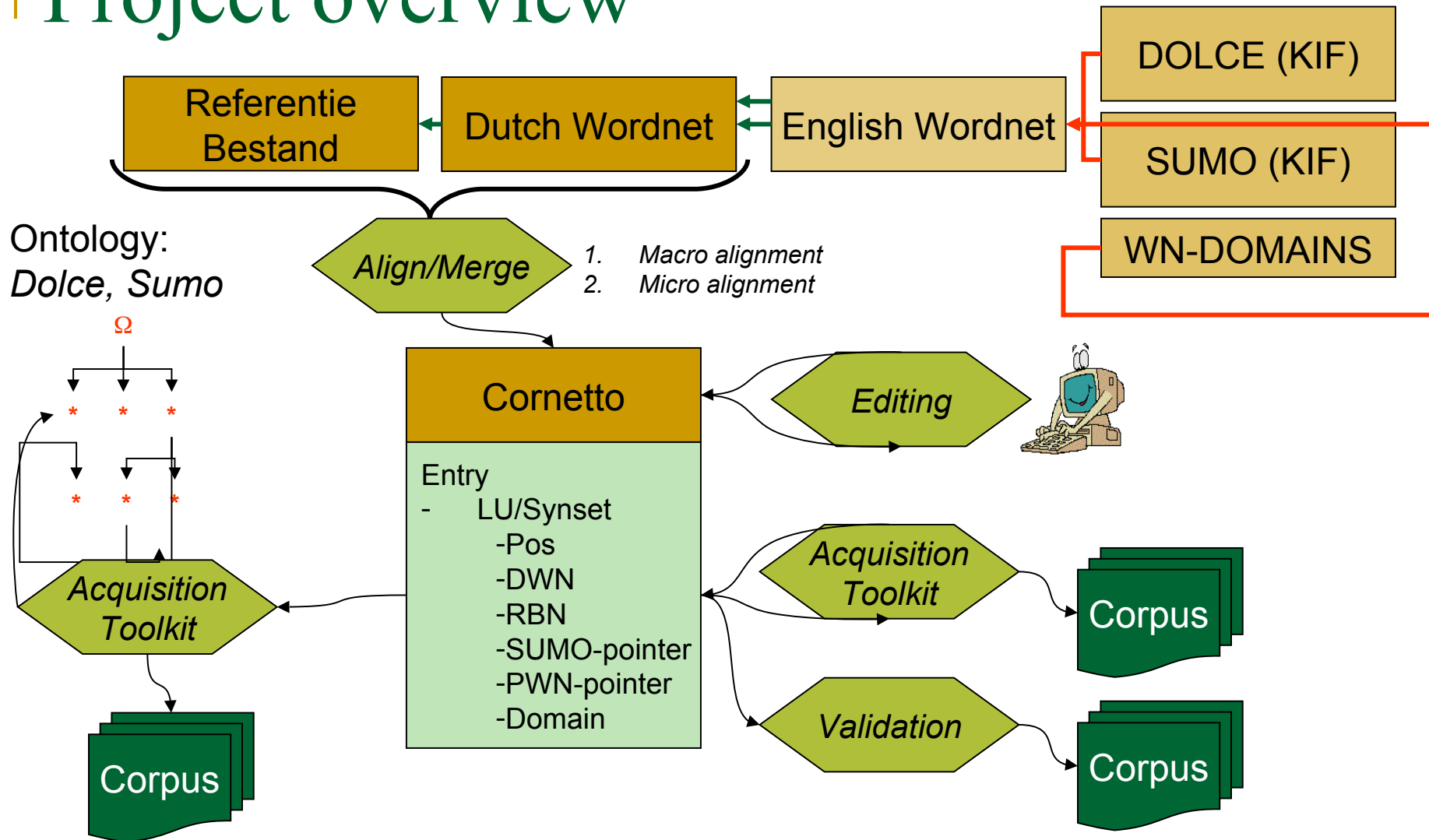
Combinatorics

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>

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



Two types of semantic resources

- **RBN:**

- ❑ Lexical units
- ❑ Form-meaning relation
- ❑ Linguistic knowledge to use forms in a meaning
- ❑ Semantic-syntactic combinatoric frames

■ DWN:

- ❑ Based on WordNet (Fellbaum 1998): relational conceptual model of the lexicon
- ❑ Synsets: set of synonyms that represent a concept
- ❑ Semantic relations between synsets, i.e. abstracts from the specific form of particular synonyms



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

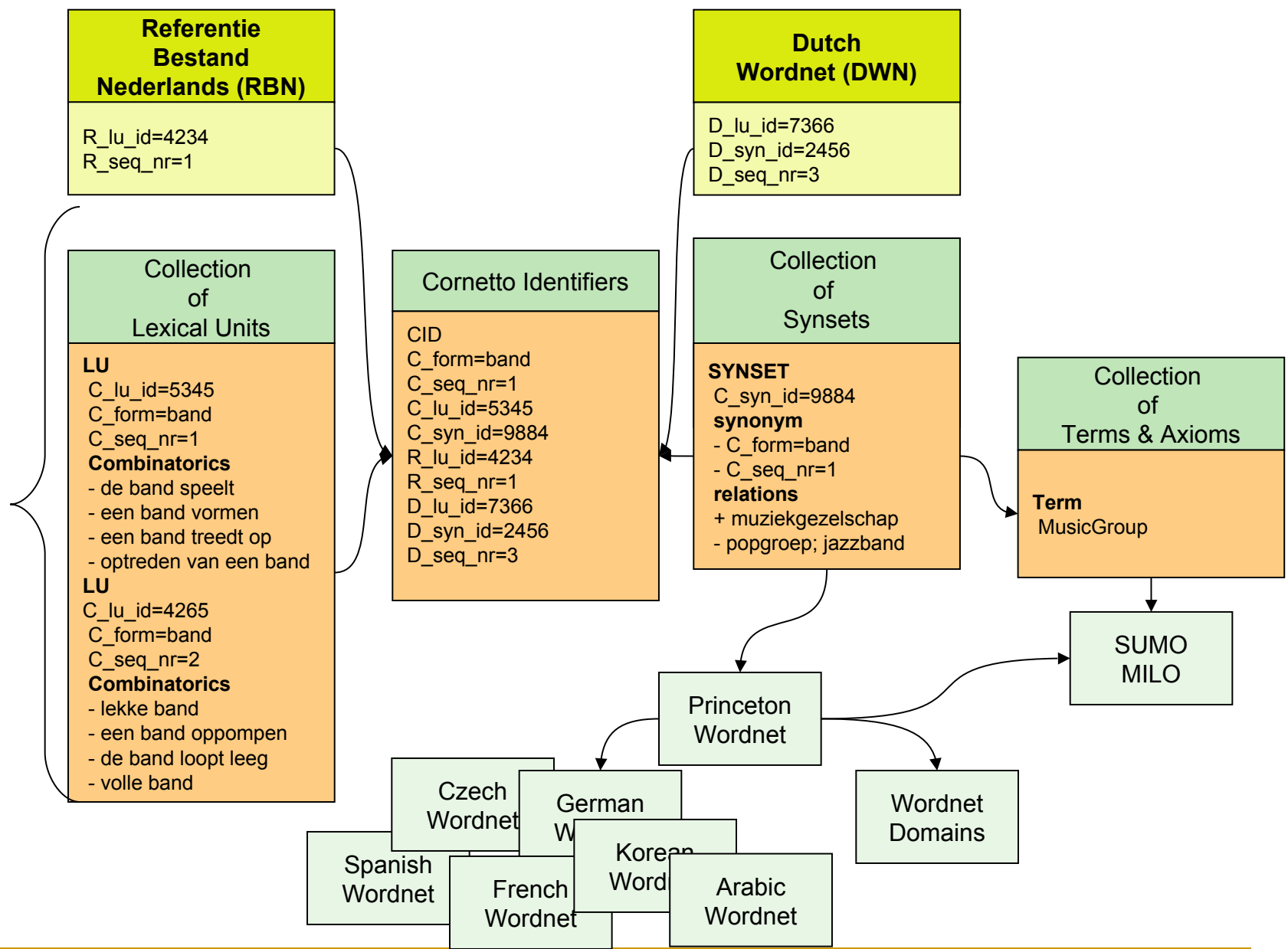
■ Collections:

- ❑ 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)

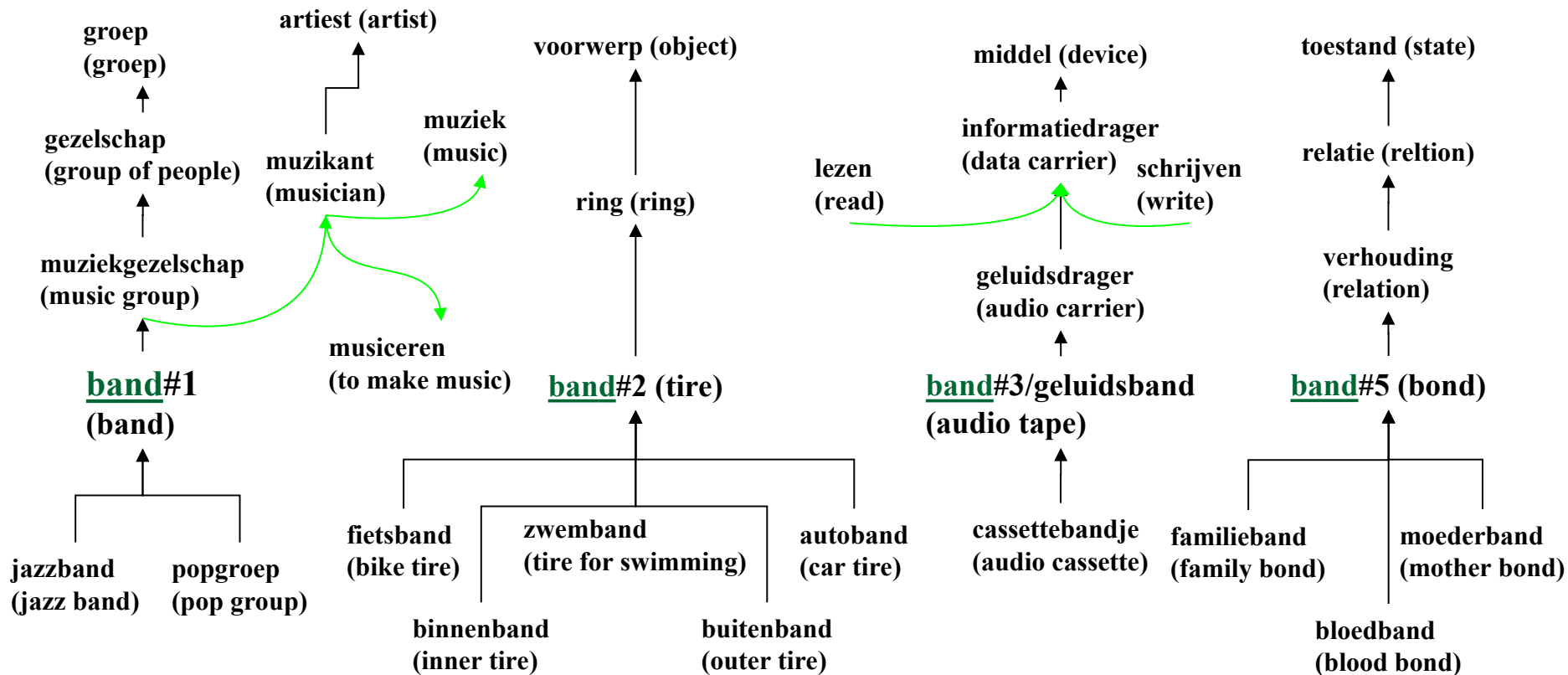


Combinatorics
in een band spelen (to play in a band)
een band oprichten (to start a band)
de band speelt (the band plays)

Combinatorics
de band oppompen (to pump air in a tire)
een band plakken (to fix a whole in a tire)
een lekke band (flat tire)
de band springt (the tire explodes)

Combinatorics
de band starten (to start a tape)
op de band opnemen (to record on a tape)
de band afspelen (to play from a tape)

Combinatorics
een goede/sterke band (a good strong bond)
de banden verbreken (to break all bonds)
een band hebben met iemand (to have a bond with s.o.)



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...

Cornetto Synset

vertrouwen

[*] d_v-8319 [verb] vertrouwen:1, betrouwen:2, leunen:2, rekenen:8, steunen:5, verlaten:6, bouwen:2 **zich verlaten,**

[*] d_n-26003 [noun] vertrouwen:1, betrouwen:3 KARAK geloof in iemands betrouwbaarheid

[*] d_v-8320 [verb] vertrouwen:2 **betrouwbaar achten**

Concept Relations Tree Revtree Edit Xml

C_SY_ID: **d_n-26003**

Synonyms: **vertrouwen:1/r_n-40551, betrouwen:3/d_n-401510**

PoS specific: NOUN_HET

Definition: geloof in iemands betrouwbaarheid

VLIS domains: KARAK

SUMO: (+, , **believes**)

-->> [HAS_HYPERONYM] gedachte:6/d_n-22494

-->> [HAS_HYPERONYM] gevoel:5, aandoening:3, affect:1, emotie:1, gemoedsaandoening:1, gemoedsbeweging:1, gevoelen:2, sentiment:1/d_n-39425

-->> [HAS_HYPERONYM] houding:3, attitude:1, gedraging:1, gedragslijn:1, habitus:3, opstelling:2, positie:6/d_n-34156

-->> [NEAR_ANTONYM] wantrouwen:1, achterdocht:1, argwaan:1, suspicie:1/d_n-14991

-->> [XPOS_NEAR_SYNONYM] **vertrouwen:2/d_v-8320**

Equivalence relations:

[EQ_SYNONYM] **faith:2, trust:5/ENG20-05597303-n**

WordNet domains: **factotum**

Querying a dictionary is complete. Item(s): 3

Cornetto LU

vertrouwen

[*] r_n-40551 noun, vertrouwen:1, geloof in betrouwbaarheid

[*] r_v-9410 verb, vertrouwen:1, rekenen op

[*] r_v-9411 verb, vertrouwen:2, als betrouwbaar beschouwen

Form Examples Edit Xml

type: **fixed** subtype: **gracol**

Example: 79166

canonical form: **iemands vertrouwen genieten**

type: **fixed** subtype: **lexcol**

Example: 79167

canonical form: **iemand zijn vertrouwen schenken**

type: **free**

Example: 79168

canonical form: **in vertrouwen**

type: **fixed** subtype: **gracol**

Example: 79169

canonical form: **vertrouwen stellen in [iemand/iets]**

type: **free**

Example: 79170

canonical form: **iemand in vertrouwen nemen**

type: **fixed** subtype: **idiom**

Example: 79552

canonical form: **het volste vertrouwen**

type: **fixed** subtype: **lexcol**

Example: 79553

canonical form: **erg goed van vertrouwen zijn**

type: **fixed** subtype: **idiom**

Querying a dictionary is complete. Item(s): 3

rekenen:4, uitbrengen:1 COMM (een boodschap) mondeling overbrengen

tree Edit Xml

ONYM] meedelen:1, not_found:0/d_v-4150
ONYM] spreken:7, kletsen:3, praten:2, reppen:3/d_v-6668
YM] toezeggen:1, beloven:1/d_v-1142
YM] benadrukken:1, aanzetten:12, accentueren:3, beklemtonen:2,
:3, onderstrepen:3, tamboereren:2/d_v-1158
YM] uitlaten:5, beoordelen:1/d_v-1167
YM] pretenderen:1, beweren:1/d_v-1369
YM] dicteren:1/d_v-2001
YM] distantiëren:1/d_v-2028
YM] goedemorgenzeggen:1/d_v-2763
YM] goedenachtzeggen:1/d_v-2765
YM] goedendagzeggen:1/d_v-2766
YM] herhalen:3/d_v-2937
YM] herkauwen:3/d_v-2947
YM] hummen:1/d_v-3059
YM] inbrengen:5, aandragen:2/d_v-3125
YM] meesmuilen:1/d_v-4189
YM] nazeggen:1, echoën:3, herhalen:4/d_v-4412
YM] noemen:3/d_v-4473
YM] ontkennen:1, miskennen:1, ontveinzen:1/d_v-4793
YM] opmerken:2, aantekenen:4/d_v-5161
YM] opzeggen:2/d_v-5352
YM] stellen:5, poneren:1/d_v-6741
YM] toegeven:6, erkennen:2/d_v-7153
YM] toesnauwen:1, aanblaffen:2, sissen:2, toebijten:1, toeblassen:1,
n:4/d_v-7184

zeggen

r_v-10227 verb, zeggen:1, in gesproken woorden uiten
r_v-10228 verb, zeggen:2, betekenen
d_v-343372 verb, zeggen:3
d_v-343373 verb, zeggen:4

Form Examples Edit Xml

Syntax:
trans: tran separ: onsch class: main peraux: h valency: tri reflexiv: nrefl subje
complementation: np, np + dat, np + of, np + wh, np + utterance, np np

Semantics:
type: action
resume: in gesproken woorden uiten
definition: XVD: in gesproken woorden uiten
hypernyms: uiten

Examples:
Example: 6679
canonical form: zo gezegd, zo gedaan
type: fixed subtype: idiom

Example: 6680
canonical form: nee/ja zeggen
type: free

Example: 6681
canonical form: iets over iets/iemand zeggen
type: fixed subtype: gracol

Example: 6682
canonical form: iemand zeggen dat...
type: free

Example: 6683
canonical form: hij zei: 'geef de appelmoes eens door'
type: free

en:1, betrouwen:2, leunen:2, rekenen:8, steunen:5, verlaten:6, bouwen:2 zich verlaten,
wen:1, betrouwen:3 KARAK geloof in iemands betrouwbaarheid
en:2 betrouwbaar achten

tree Edit Xml

2/r_v-9411

TRANSITIVE

r achten

wbaar

, (+, , IntentionalPsychologicalProcess)

[SYNONYM] aanmerken:3, houden:27, oordelen:3, schatten:2, zien:7, achten:1,
en:1/d_v-1225

[SYNONYM] wantrouwen:1/d_v-8766

[SYNONYM] vertrouwen:1, betrouwen:3/d_n-26003

[SYNONYM] trust:1, swear:5, rely:1, bank:8/ENG20-00665312-v

[SYNONYM] believe:3, trust:3/ENG20-00697189-v

psychology

vertrouwen

[*] r_n-40551 noun, vertrouwen:1, geloof in betrouwbaarheid
[*] r_v-9410 verb, vertrouwen:1, rekenen op
[*] r_v-9411 verb, vertrouwen:2, als betrouwbaar beschouwen

Form Examples Edit Xml

C_LU_ID: r_v-9411 C_SEQ_NR: 2
[verb] vertrouwen

Morphology:
type: **simp**morph flex conjugation: type: **regular**,
flex mode: **inf** flex tense: **ntense** flex number: **nnumber** flex person: **nperson**

Syntax:
trans: **tran** separ: **onsch** class: **main** peraux: **h** valency: **di** reflexiv: **nrefl** subje
complementation: **np**

Semantics:
type: **action**
resume: **als betrouwbaar beschouwen**
definition: **XVD: geloven dat de of het genoemde betrouwbaar is**
hypernyms: **geloven**

Examples:

Example: 7386
canonical form: **zoals de waard is vertrouwt hij zijn gasten**
type: **fixed**

Example: 7389
canonical form: **het hele zaakje niet vertrouwen**
type: **free**

Example: 7390
canonical form: **iemand kunnen vertrouwen**
type: **free**