

Cornetto



September, 21st, 2007
STEVIN programmaday

ru | STEVIN

Goal

- 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
- <http://www.let.vu.nl/onderzoek/projectsites/cornetto/start.htm>

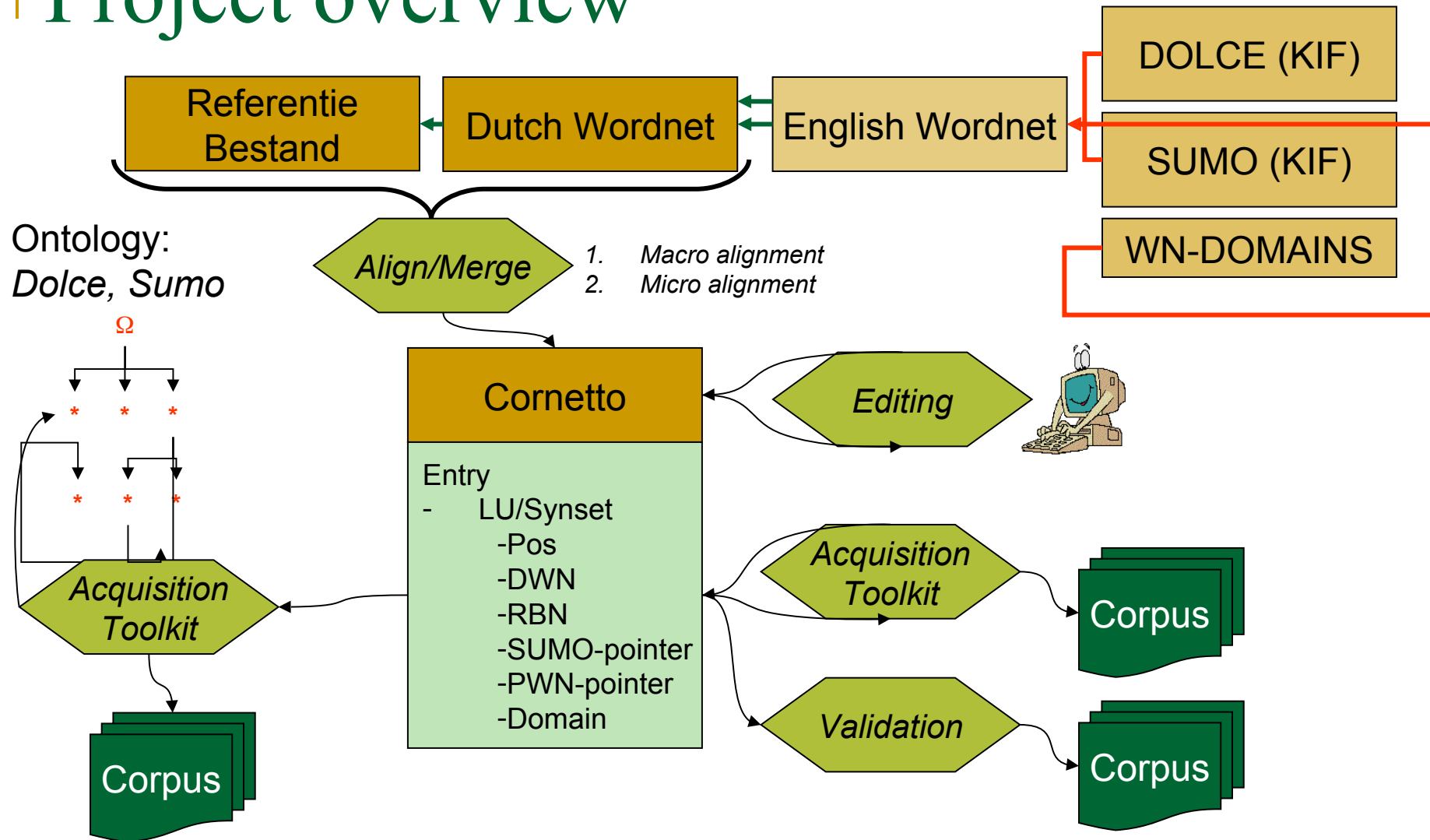
Consortium

- 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

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

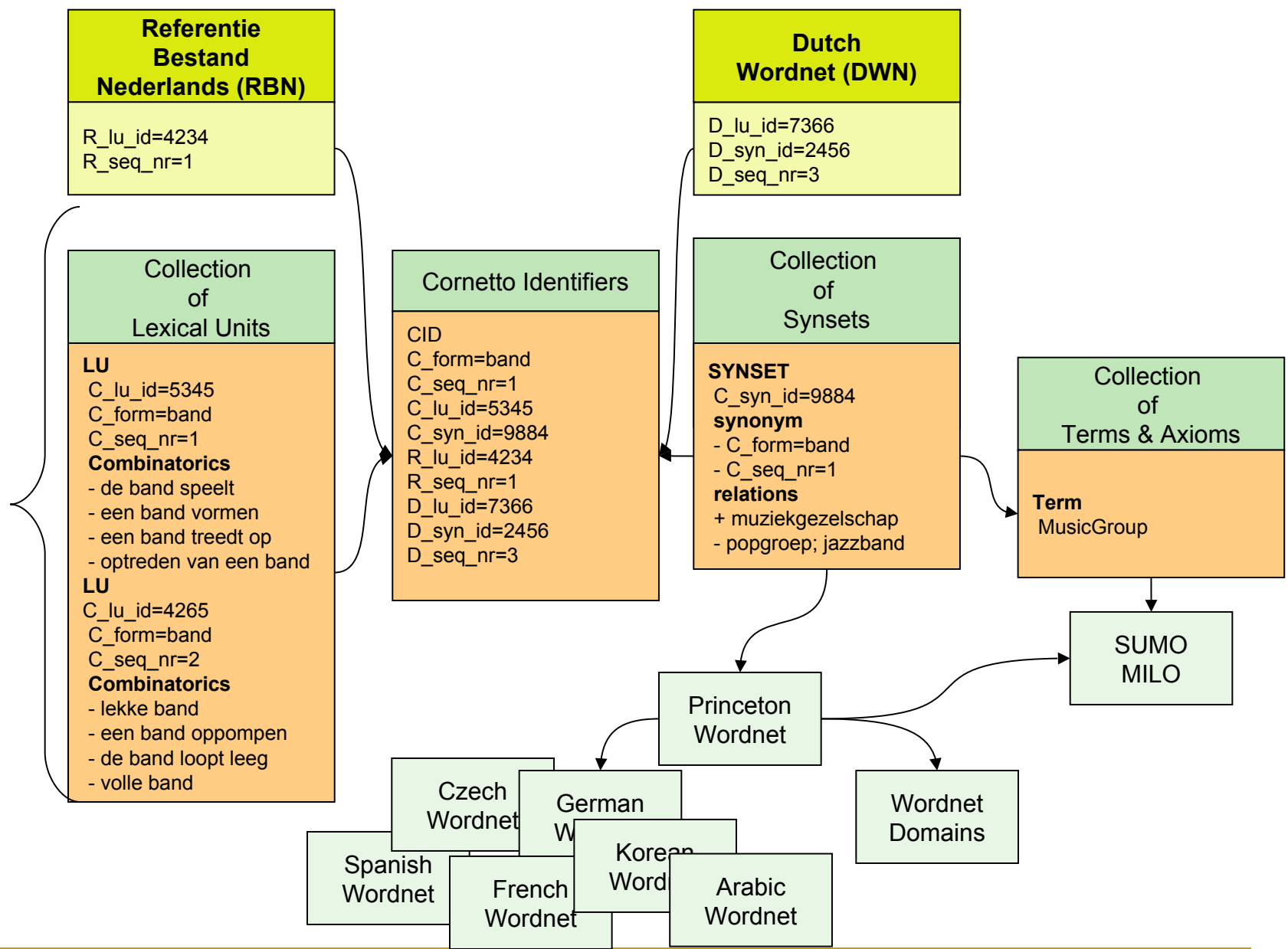
Project overview



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

Cornetto Database (CDB)



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

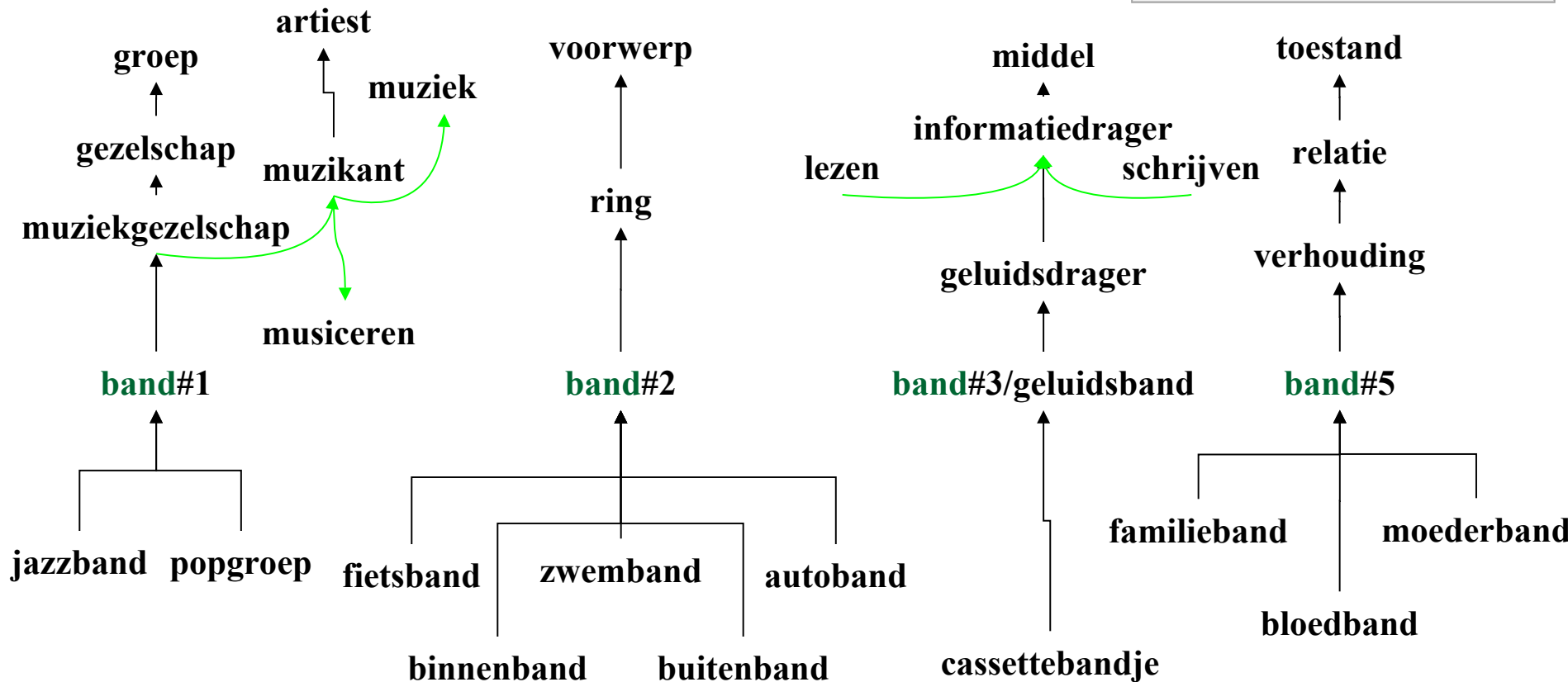
de band afspelen

Combinatoriek

een goede/sterke band

de banden verbreken

een band hebben met iemand



Editing strategies

Macro alignment: LU's vs. synsets
Ontology assignment

- ❑ Frequent verbs
- ❑ Nouns with a semantic shift in RBN
- ❑ Adjectives
- ❑ High-polysemous words
- ❑ Words with low scores

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 a 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> |

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.