

# Open Dutch WordNet

*Marten Postma, Emiel van Miltenburg, Roxane Segers, Anneleen Schoen, and Piek Vossen*

RELEASE

1.3

DATE

January 30, 2016

LICENSE

CC BY-SA 4.0

WEBSITE

[wordpress.let.vupr.nl/odwn/](http://wordpress.let.vupr.nl/odwn/)

CO-FUNDED BY

Nederlandse TaalUnie

# Outline

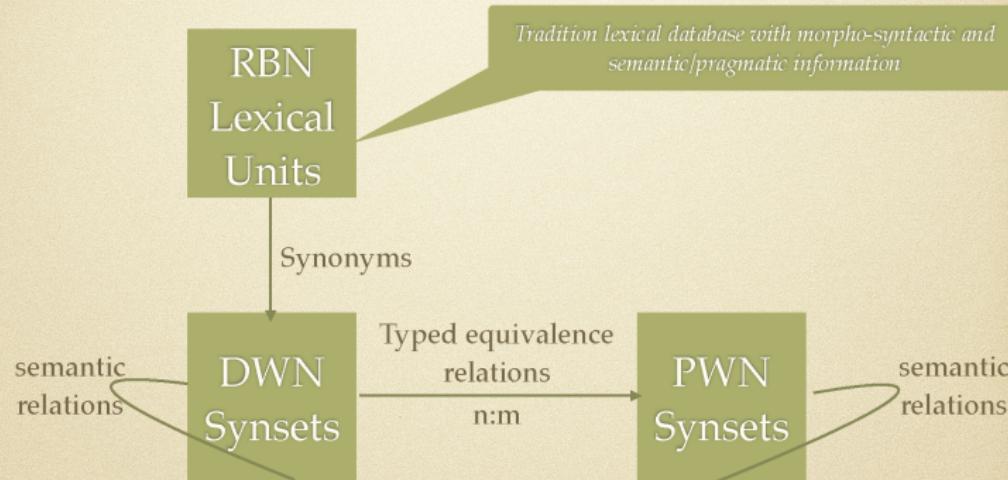
1 Why

2 How

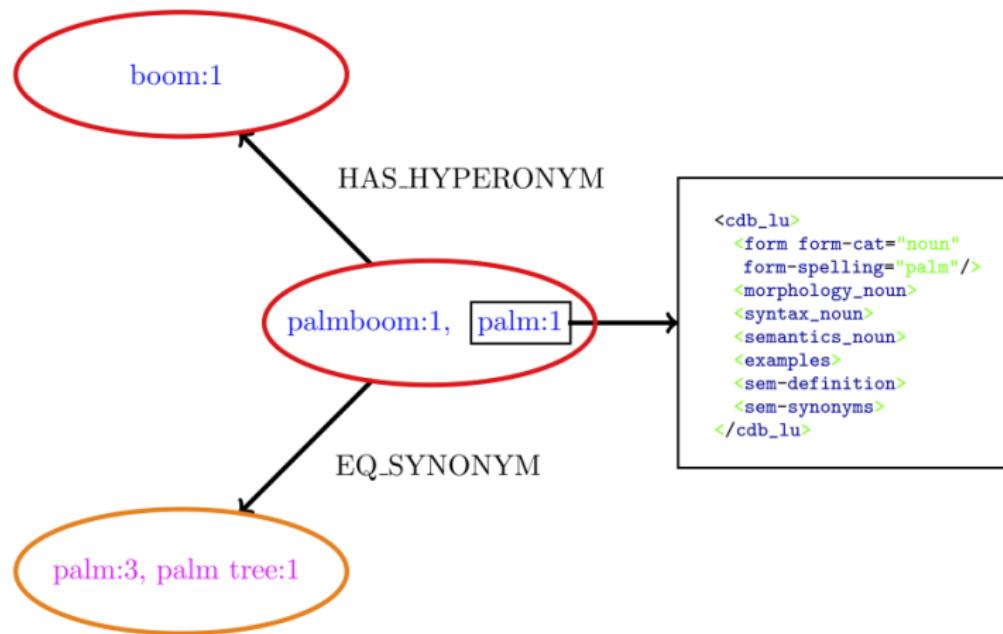
3 What

# Cornetto 2.0

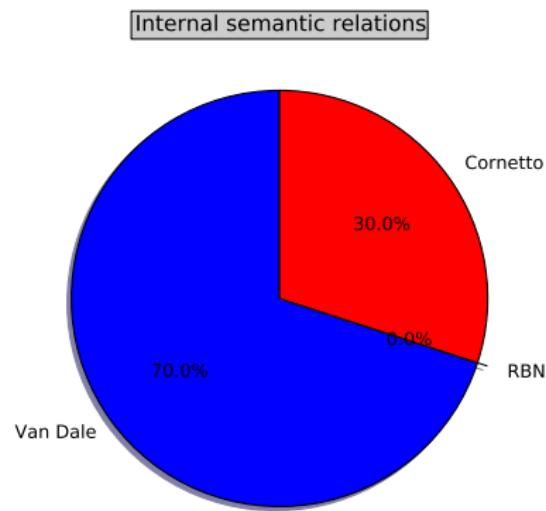
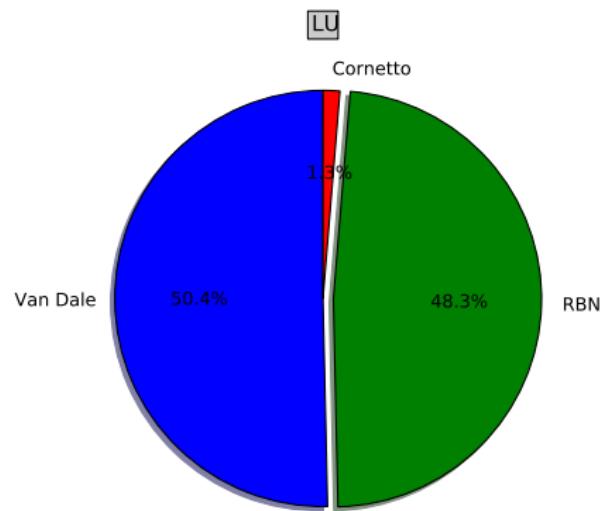
## Cornetto2.0



# Structure Cornetto 2.0



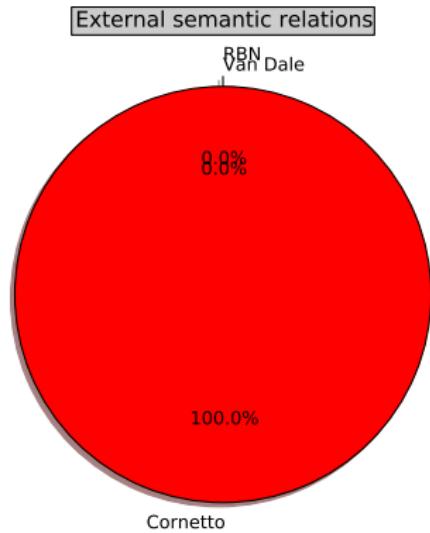
# Provenance components Cornetto 2.0



# Hiking trip by Piek Vossen



# Distribution external semantic relations: only Cornetto



# Summary of Why

- ① Cornetto 2.0 is not open source, due to the Van Dale content inside.
- ② In order to create Open Source Dutch Wordnet, we must remove the Van Dale content from Cornetto 2.0 and replace it with open source content.
- ③ the open source external semantic relations are a good starting point.

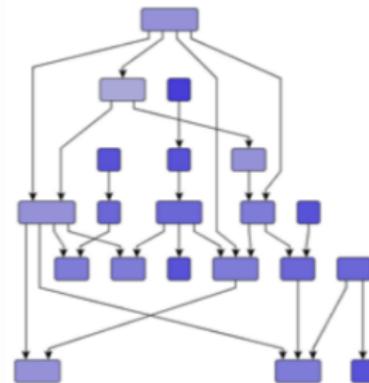
# Creating Open Source Dutch WordNet



Wiktionary  
The free dictionary

Google translate

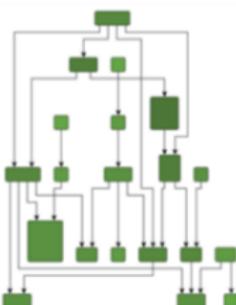
**WordNet**  
A lexical database for English



filtering  
equivalence  
relations



palm:3



# Manual Annotation

Cornetto synset  $\longleftrightarrow$  Wordnet synset:

- ① synset depth
- ② number of children
- ③ number of ESRs

During a period of four months, working 8 hours a week, four students manually checked 12,966 ESRs, of which 6,575 were removed.

A decision tree algorithm (Weka, Hall et al., 2009) was used to filter all relations that had not been checked manually. Features included:

- ① synset depth
- ② number of ESRs
- ③ translation overlap
- ④ average semantic similarity score (Leacock & Chodorow)

The manual annotation was split in a training part and a development part. 32,258 external semantic relations were removed.

Open Source Dutch WordNet is then created in the following way:

- ① Take all synsets and relations from English WordNet.
- ② If *eq\_synonym* or *eq\_near\_synonym* relation to WordNet synset then insert LU's not from Van Dale into WordNet synset.
- ③ Else if *eq\_has\_hyperonym* then add as new synset to English Wordnet linked as a *hyponym*
- ④ Else if Cornetto *hypernym* has equivalence relation to WordNet, then add as new synset to English Wordnet linked as a *hyponym*.
- ⑤ Using open source dictionaries, the English synonyms in WordNet are translated into Dutch.

# Open source dictionaries

Oliver (2014) used Wiktionary, OmegaWiki, and Google to translate synonyms from English WordNet to other languages, including Dutch:

- ① **monosemous**: the synonym is simply translated
- ② **polysemous**: gloss overlap and semantic relations are used to find the best synset.

A similar strategy with OPUS, Bing, and Google Translate was used to enrich the adjectival part of the resource.

## Manual editing round 2

- ➊ **excessive synsets:** manual inspection of synsets containing 5 or more synonyms
- ➋ **synonym filtering:** it could happen that the same synonym was placed in multiple synsets, we manually selected the correct one.

8.257 synonyms were checked in this phase.

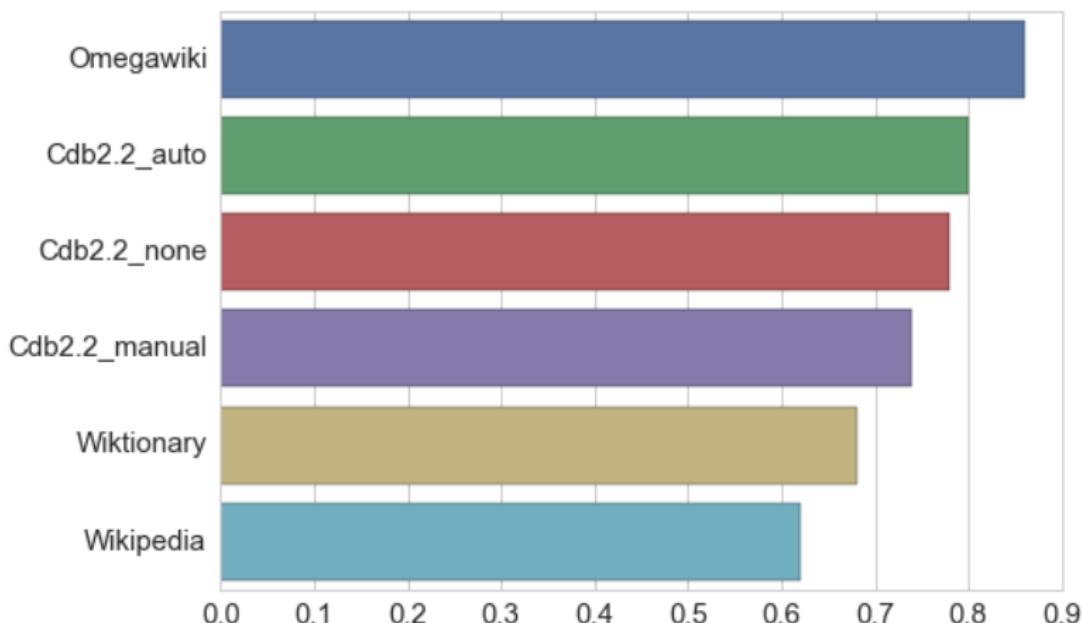
# Main statistics

Main statistics about the resource:

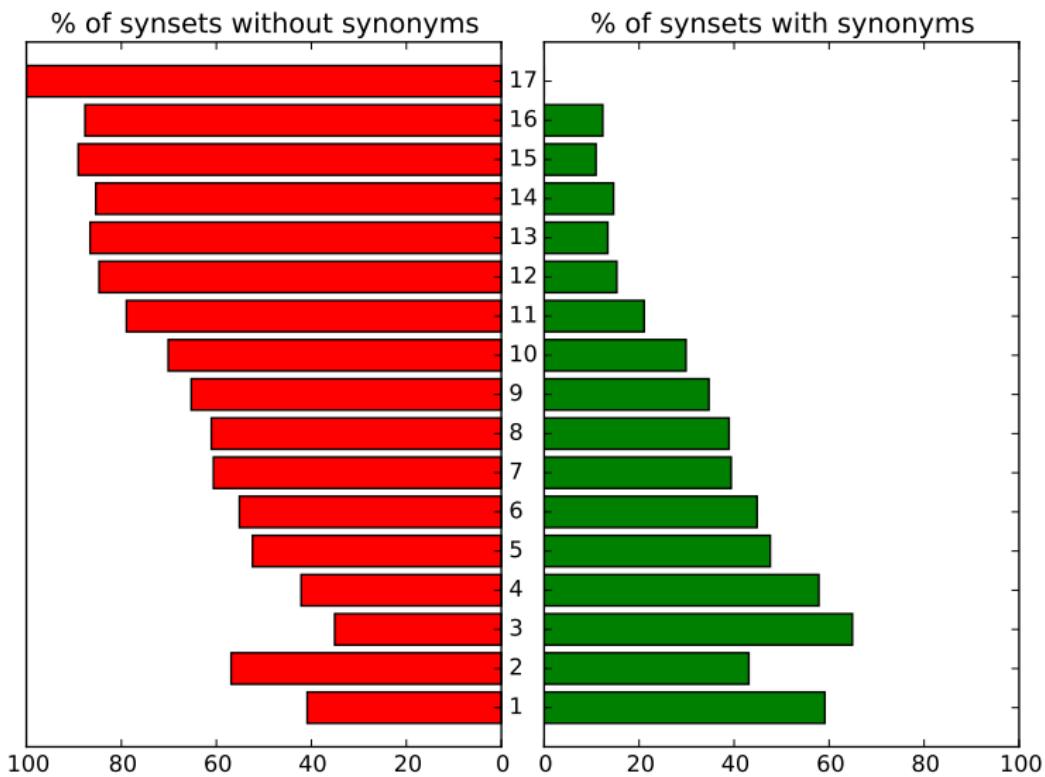
- ① synsets: 117.914
- ② 51.588 synsets contain a Dutch synonym
- ③ Number of synonyms per resource:

Provenance	instances	% of all LE
cdb2.2_Auto	32806	35.5
cdb2.2_None	19073	20.7
wiktionary	17968	19.5
cdb2.2_Manual	13075	14.2
omegawiki	12589	13.6
google	8374	9.1
opus	612	0.7
bing	506	0.5
wikipedia	375	0.4

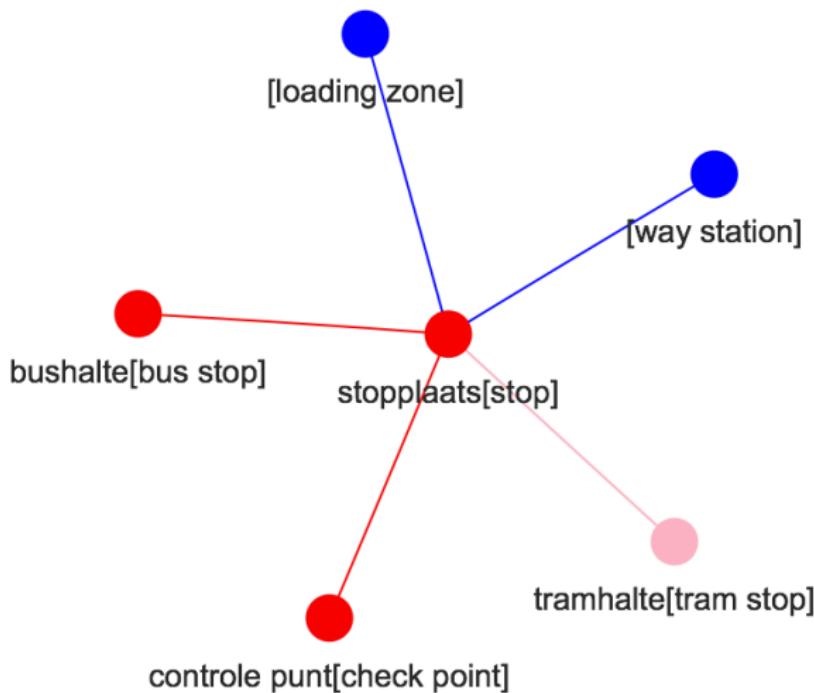
# Confidence of polysemous synonyms



# Depth distribution



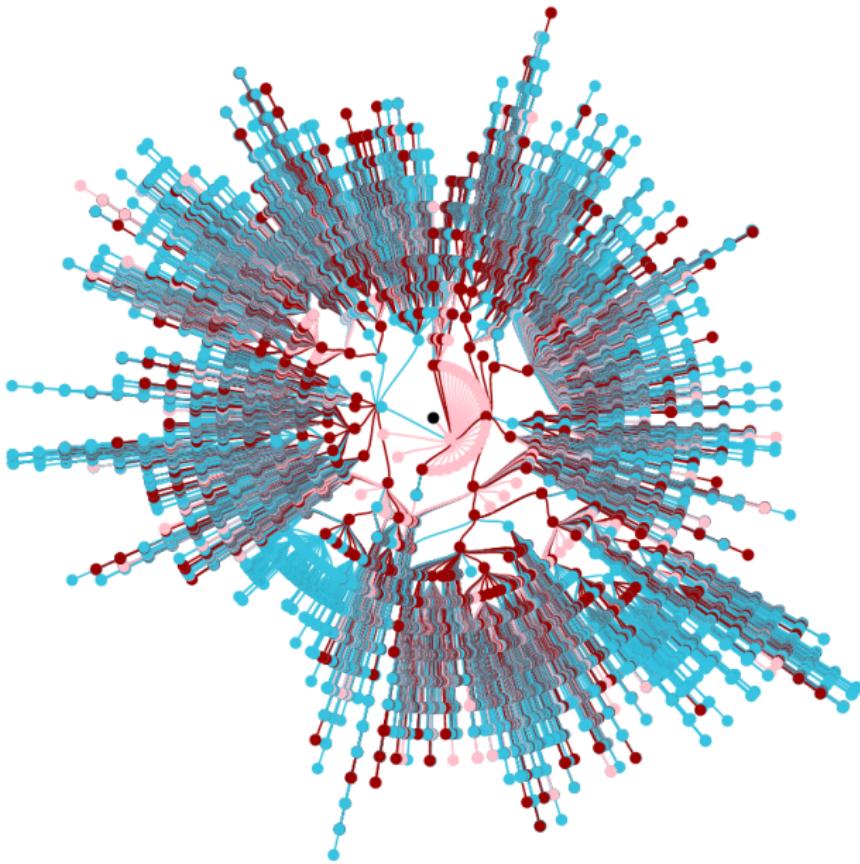
# New Concepts



Why  
oooooo

How

What



# Formats and Usage

## Global WordNet Grid LMF

<https://github.com/MartenPostma/OpenDutchWordnet>

## Open MultiLingual WordNet

Thank you! Questions?