

# Open Source

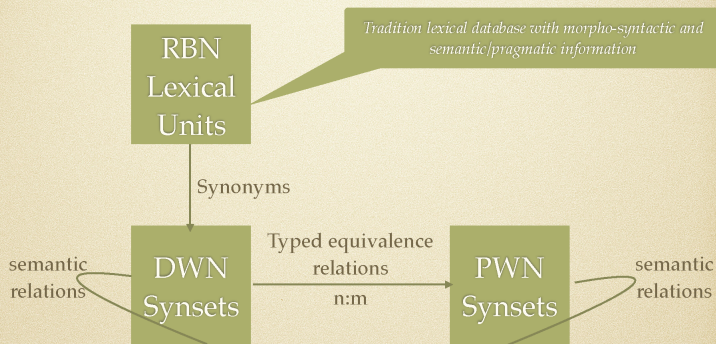
# Dutch WordNet

MARTEN POSTMA	m.c.postma@vu.nl
PIEK VOSSEN	p.t.j.m.vossen@vu.nl
RELEASE	1.0
DATE	December 1, 2014
LICENSE	CC BY-SA 4.0
WEBSITE	<a href="http://wordpress.let.vu.nl/odwn/">wordpress.let.vu.nl/odwn/</a>
CO-FUNDED BY	Nederlandse TaalUnie

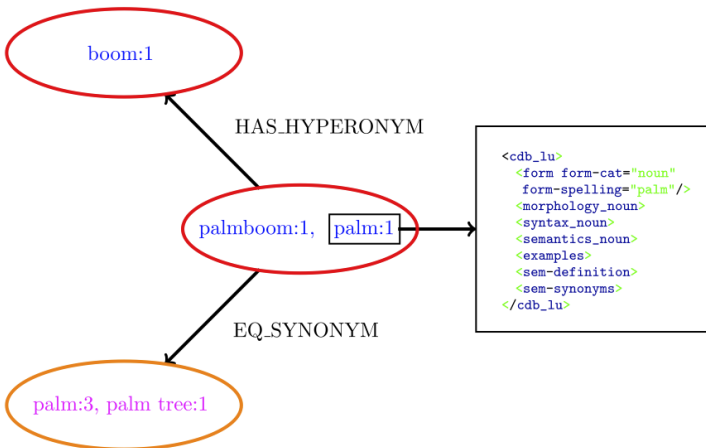
# Outline

- 1 Why
- 2 How
- 3 What

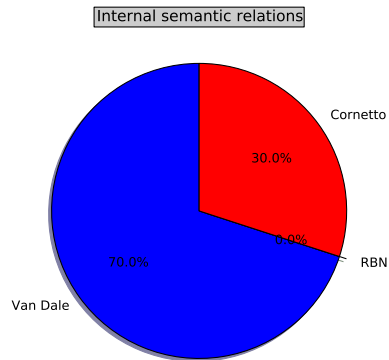
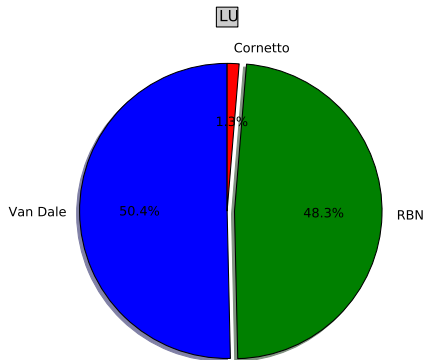
# Cornetto 2.0



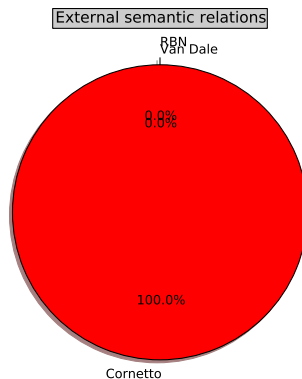
# Structure Cornetto 2.0



# Provenance components Cornetto 2.0



# Distribution external semantic relations: only Cornetto



# Summary of Why

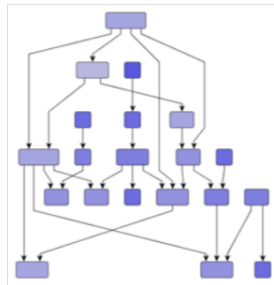
- 1 Cornetto 2.0 is not open source, due to the Van Dale content inside.
- 2 In order to create Open Source Dutch Wordnet, we must remove the Van Dale content from Cornetto 2.0 and replace it with open source content.
- 3 the open source external semantic relations are a good starting point.

# Creating Open Source Dutch WordNet



## WordNet

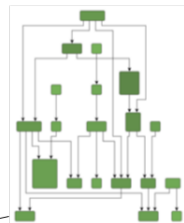
A lexical database for English



filtering  
equivalence  
relations



palm:3





# Manual Annotation

The manual annotation consisted of checking whether an equivalence relation between a Cornetto synset and a Wordnet synset was valid or not.

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 semantic similarity measures, position in the hierarchy, translations and many more. By doing this, 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 resources (Wikipedia (Wikipedia, 2014;Foundation,2014a), Wiktionary (Foundation, 2014b), Google Translate (Google,2014), BabelNet (Navigli et al., 2010)), the English synonyms in WordNet are translated into Dutch and added if at least two resources propose the same synonym for the same synset.

# Example

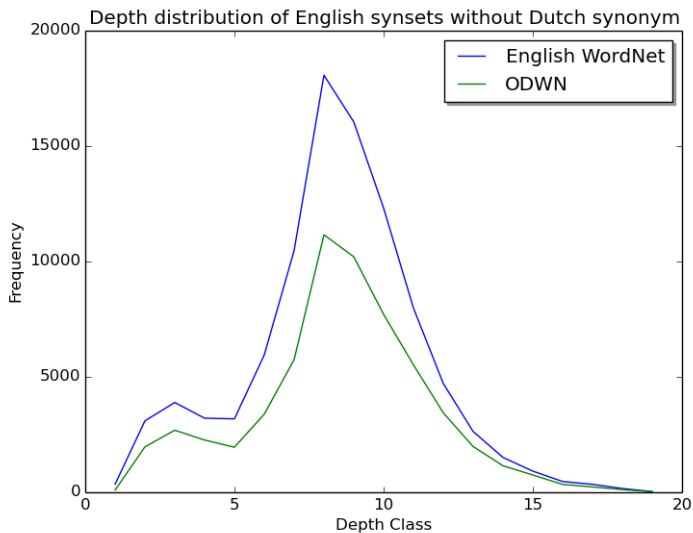
```
<cdb_synset c_sy_id="eng-30-06722453-n" posSpecific="NOUN" comment="EQ_SYNONYM">
  <synonyms>
    <synonym c_lu_id="r_n-39115" c_lu_id-previewtext="uitspraak:2" status="cdb2.2_Manual"/>
    <synonym c_lu_id="o_n-102764766" c_lu_id-previewtext="dictum:1" status="cdb2.2_Manual"/>
    <synonym c_lu_id="t_n-679214835" c_lu_id-previewtext="verklaring:4" status="google_api+wiktionary"/>
  </synonyms>
  <definition>a message that is stated or declared;
  a communication (oral or written) setting forth particulars or facts etc</definition>
  <wn_internal_relations>
    <relation relation_name="HAS_HYPERONYM" target="eng-06598915-n" source="pwn"/>
    <relation relation_name="HAS_HYPERONYM" target="eng-06284225-n" source="odwn"/>
  </wn_internal_relations>
</cdb_synset>
```

# Main statistics

Main statistics about the resource:

- ① synsets: 116992 (of which 21636 are new)
- ② 34613 synsets contain a Dutch synonym
- ③ Number of synonyms per resource:
  - ① **wiktionary**: 20707
  - ② **google**: 22708
  - ③ **wikipedia**: 422
  - ④ **babelnet**: 20440

# Distribution depth Dutch synonyms



# Editor

`https://debvisdic.let.vu.nl:9002/editor/`

# Future work

- ① Convert resource to RDF (currently in XML and LMF)
- ② add adjectives to resource
- ③ translate the Dutch resumes
- ④ add resource to Global WordNet Grid
- ⑤ ....



Thank you! Questions?