GrETEL 2.0

Liesbeth Augustinus
Ineke Schuurman
Vincent Vandeghinste
Frank Van Eynde

CLARIN ERIC 2013 Prague
GrETEL

How to combine the **data-oriented approach** of treebank mining with the **knowledge-oriented** method of theoretical and descriptive linguistics? And in a user-friendly way?
For Dutch (LASSY/CGN)

XPath
- Not user-friendly
- Knowledge of Alpino grammar necessary

= problematic for non-technical linguists
• Verify theory through data with corpus or treebank examples
• Time consuming, requires some effort

How to make interaction between computational linguistics and theoretical linguistics possible?
Why GrETEL?

GrETEL meets a demand:

• Many theoretical/descriptive linguists want tools that are easy to use, if not:
  – they don’t start using them from the beginning (the tools scare them)
  – they stop using them

• Even when improved at a later stage, it may already be a dead loss

GrETEL turns out to be easy to use
  – You don’t need to know a query language and is used at several universities in both Flanders and the Netherlands (teaching)
Treebanks

2 treebanks, each +/- 1M tokens:
• Lassy small (written Dutch)
• CGN (spoken Dutch)

Treebanks to be added in the near future
• Lassy large (written Dutch, <500M tokens)
• AfriBooms treebank (written Afrikaans, 55K tokens)
Using GrETEL

An example:

“look for all NP nodes in which the head noun is modified by the adjective ‘politiek’, e.g. politieke discussies”

• XPath:

```
//node[@cat="np" and node[@rel="mod" and @pos="adj" and @root="politiek"] and node[@rel="hd" and @pos="noun"]
```

• GrETEL:

```
“ze voeren politieke discussies” (415 hits in 406 sentences)
```

[they had political discussions]
Thanks for your attention!

Questions?

gretel@ccl.kuleuven.be

http://nederbooms.ccl.kuleuven.be