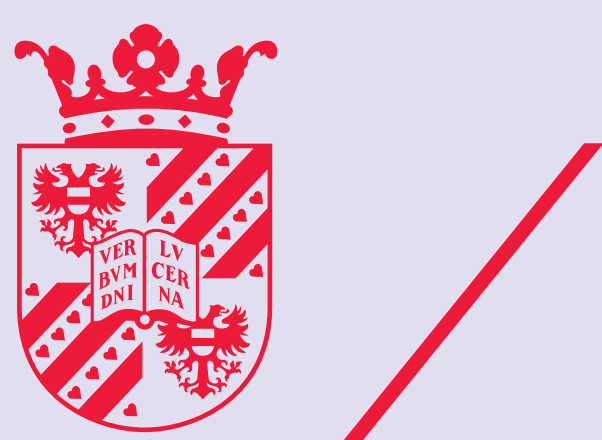
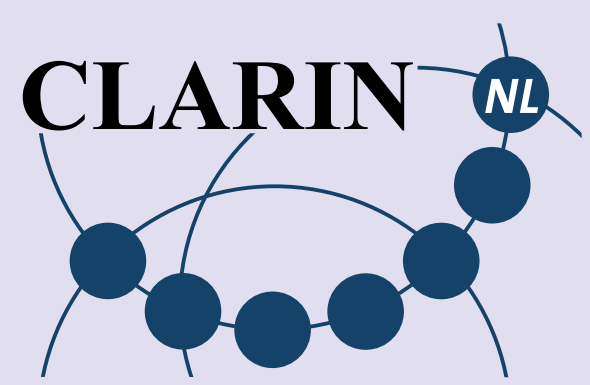




PaQu: Parse & Query

Gertjan van Noord



In PaQu kun je zoeken in *syntactisch* geannoteerde corpora. De syntactische annotatie geeft aan:

- wat de eigenschappen van de woorden zijn
- welke woorden bij elkaar horen
- wat de relatie tussen woorden en woordgroepen is

PaQu ondersteunt twee zoekmethoden:

- Eenvoudig zoeken naar woorden en woord-paren
- Zoeken naar specifieke constructies met XPath

In de eenvoudige zoekinterface kun je zoeken naar woord-paren waarbij je kunt specificeren:

• Woord en/of Lemma

• Woordsoort

adj	bijvoeglijk naamwoord	bw	bijwoord
vnw	voornaamwoord	vz	voorzetsel
n	zelfstandig naamwoord	ww	werkwoord
lid	lidwoord	tw	telwoord
vg	voegwoord	mwu	deel van uitdrukking

• Grammaticale relatie

hd	hoofd	su	onderwerp
obj1	lijdend voorwerp	mod	bepaling
obj2	meewerkend voorwerp	pc	voorzetselvoorwerp
svp	scheidbaar deel van werkwoord	ld	locatief complement
sup	voorlopig onderwerp	pobj1	voorlopig lijdend voorwerp

lemma: 259× bier, 11× drank, 10× sake, 10× toverdrank, 6× pils, 5× abdijbier, 3× aan, 3× wisseldrank, 3× witbier, 2×

woord

obj1

hoofdwoord

postag

help

Zoeken

Wissen

Reset

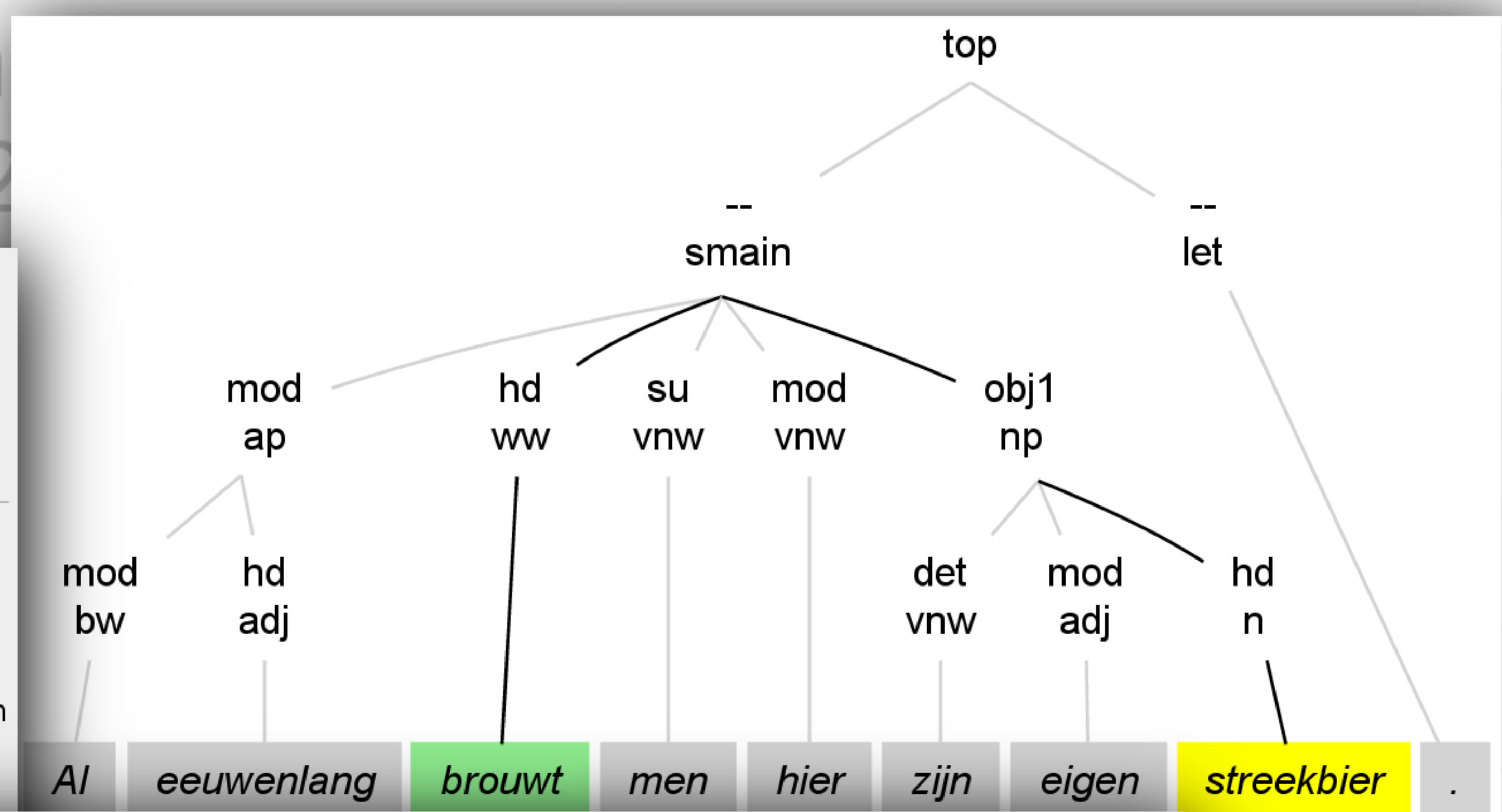
`hlemma` IN ("Brouwen", "brouwen") AND `postag` = "n" AND `rel` = "obj1"

1. Al eeuwenlang **brouwt** men hier zijn eigen **streekbier**.
◦ **streekbier:n** — **obj1** — **brouwt:ww**

2. Bierbrouwerij De Drie Ringen is gevestigd in een voormalige stadsboerderij en **brouwt** op ambachtelijke wijze **bier**.
◦ **bier:n** — **obj1** — **brouwt:ww**

3. Een groot **aantal** bieren wordt **gebrouwen** in of is gerelateerd aan E Bourgogne des Flandres Blond , Bourgogne des Flandres Bruin , Bi , Brugs (Tarwebier) , Brugse Babbelaar , Brugse Bok , Brugse Zo , Zot Dubbel , De Garre Tripel , Steenbrugge Blond , Steenbrugge D , Steenbrugge Tripel , Steenbrugge Wit , Straffe Hendrik en Straffe Quadrupel .
◦ **aantal:n** — **obj1** — **gebrouwen:ww**

4. Zo wordt **bier** **gebrouwen** doordat gisten de aanwezige suikers om tot het smaakgevende en conserverende alcohol .
◦ **bier:n** — **obj1** — **gebrouwen:ww**



Mijn corpora

Status	Titel	Regels	Toegang	Opmerkingen
gereed	Het geloof der kameraden	2.561	openbaar	
bezig 46%	In de oude brouwerij	72		Bron: doorlopende tekst
gereed	Louis Couperus — Nippon	3.553	openbaar	
gereed	Pierre Kemp — Zuster Beatrijs	846	openbaar	
gereed	R.J. Spitz: Beatrijs — Het Middelnederlandsche gedicht in proza	285	privé	

Je kunt in PaQu op verschillende manieren de gevonden resultaten bekijken:

- Bekijk de zinnen
- Bekijk de syntactische annotatieboom
- Bekijk frequentieoverzicht

In PaQu zijn een aantal corpora standaard beschikbaar, zoals de Lassy corpora. Daarnaast heeft de gebruiker de mogelijkheid om ook eigen teksten syntactisch te doorzoeken.

- Lever je eigen teksten aan
- De teksten worden automatisch syntactisch geannoteerd
- Zoek vervolgens in je eigen teksten

PaQu is ontwikkeld door Peter Kleiweg en Gertjan van Noord, met financiële steun van Clarin. De woordparen-zoekinterface is gebaseerd op een eerdere versie die door Erik Tjong Kim Sang is gemaakt. De XPath-zoekinterface is gebaseerd op de Dact software die door Daniël de Kok en collega's is gemaakt. De automatische syntactische annotatie gebruikt de Alpino parser. De Lassy corpora zijn ontwikkeld in het Stevin Lassy project.

Referenties

Gertjan van Noord, Ineke Schuurman, Gosse Bouma: Lassy syntactische annotatie. http://www.let.rug.nl/vannoord/Lassy/sa-man_lassy.pdf (2011)

Frank Van Eynde: Part Of Speech Tagging En Lemmatisering Van Het D-Coi Corpus http://www.let.rug.nl/vannoord/Lassy/POS_manual.pdf (2005)

Gertjan van Noord: At Last Parsing Is Now Operational. Piet Mertens, Cedrick Fairon, Anne Dister, Patrick Watrin (editors): TALN06. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles. Page 20–42 (2006)

Gertjan van Noord, Gosse Bouma, Frank van Eynde, Daniël de Kok, Jelmer van der Linde, Ineke Schuurman, Erik Tjong Kim Sang, Vincent Vandeghinste: Large Scale Syntactic Annotation of Written Dutch: Lassy. In: Essential Speech and Language Technology for Dutch. Springer (2013)