



# NEDERBOOMS

## Treebank Mining for Data-based Linguistics

Liesbeth Augustinus  
Vincent Vandeghinste  
Ineke Schuurman  
Frank Van Eynde

LOT Summer School - June, 2014



# NEDERBOOMS

- **Exploitation of Dutch treebanks for research in linguistics**
- CLARIN-VL project
- October, 2010 – February, 2012



# NEDERBOOMS

- **Exploitation of Dutch treebanks for research in linguistics**
- CLARIN-VL project
- October, 2010 – February, 2012
- **Goals:**
  - User-friendly tools
  - Access to large data files
  - Fast and accurate



# GrETEL

- **G**reedy **E**xtraction of **T**rees for **E**mpirical **L**inguistics
- Query engine for treebanks



# GrETEL

- **Greedy Extraction of Trees for Empirical Linguistics**
- Query engine for treebanks
- **Treebank** = syntactically annotated corpus  
e.g. Penn Treebank (English), TüBa (German),  
LASSY, CGN (Dutch)



# TREEBANKS

CGN treebank	LASSY small
Spoken Dutch	Written Dutch
Stylistic & regional differences conversations vs read texts NL vs VL	Stylistic differences Wikipedia vs legal texts
± 1M tokens	± 1M tokens
130k sentences	65k sentences
Manually corrected	Manually corrected

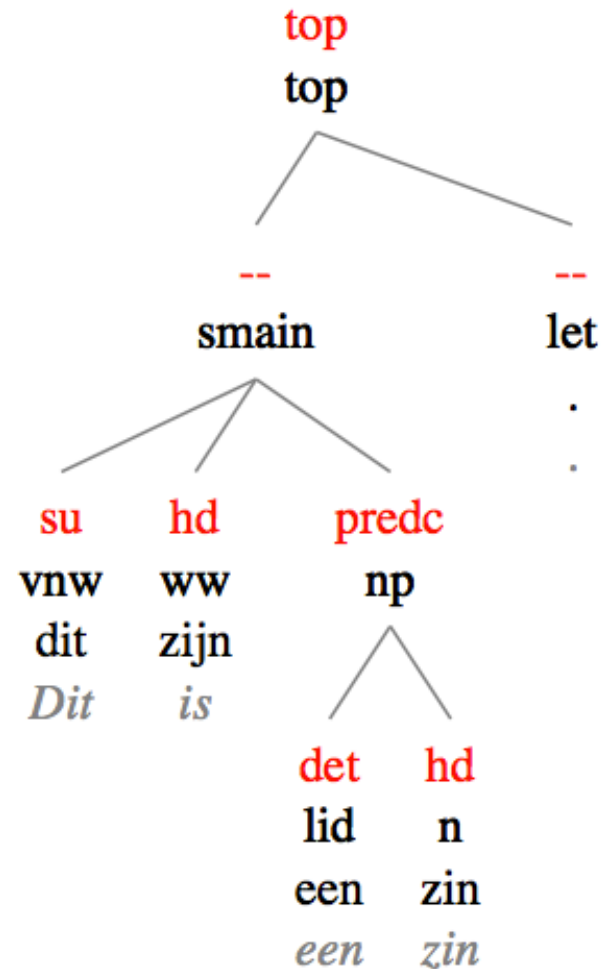
# GrETEL

- **Greedy Extraction of Trees for Empirical Linguistics**
- Query engine for treebanks
- **Treebank** = syntactically annotated corpus  
e.g. Penn Treebank (English), TüBa (German),  
LASSY, CGN (Dutch)
- **Parser**  
e.g. Alpino (Van Noord 2006)



# ALPINO PARSER

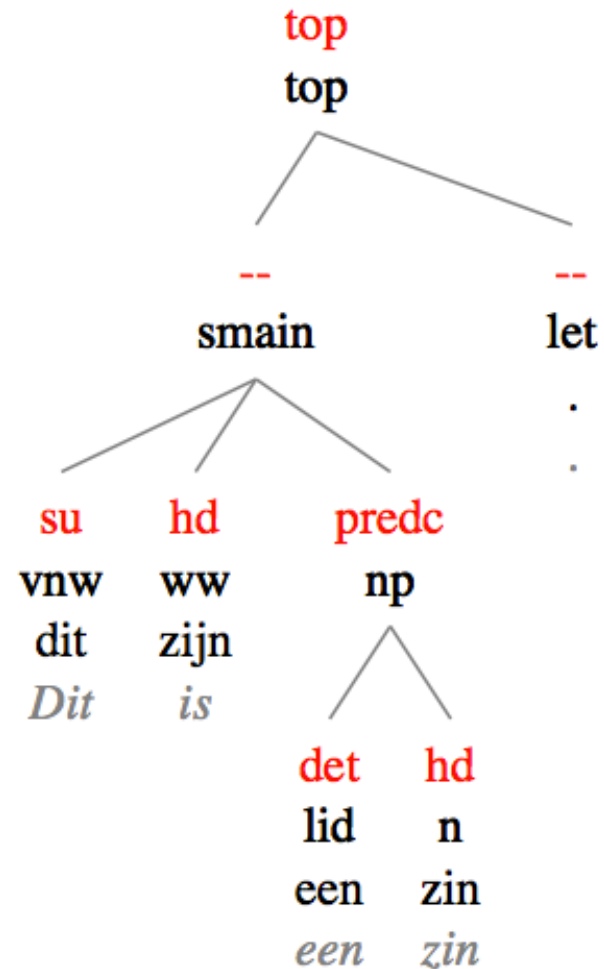
*Dit is een zin.* >> ALPINO parser >>  
'This is a sentence.'





# ALPINO PARSER

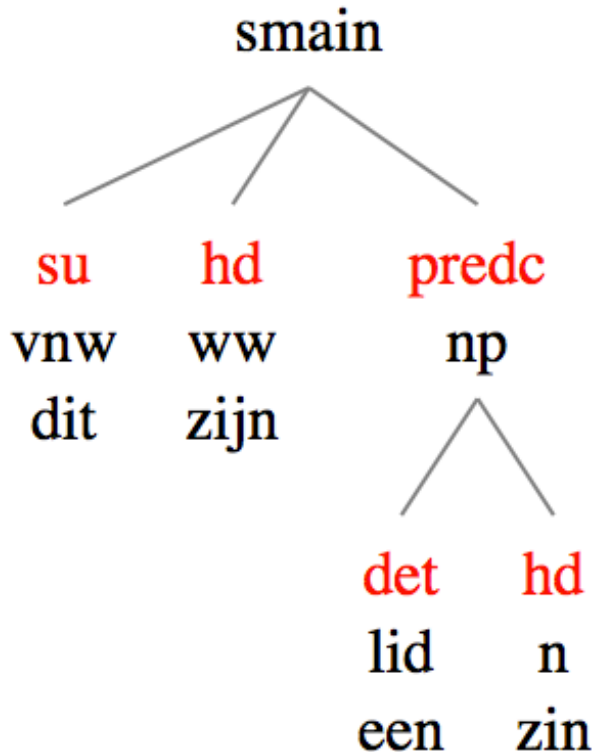
*Dit is een zin.* >> ALPINO parser >>  
'This is a sentence.'



XML trees

Query language: **XPath**

# XPATH



```
//node[@cat="smain" and  
node[@rel="su" and  
@pt="vnw" and @lemma="dit"]  
and node[@rel="hd" and  
@pt="ww" and @lemma="zijn"]  
and node[@rel="predc" and  
@cat="np" and  
node[@rel="det" and  
@pt="lid" and @lemma="een"]  
and node[@rel="hd" and  
@pt="n" and @lemma="zin"]]]
```

# XPATH



```
//node[@cat="smain" and  
node[@rel="su" and  
@pt="vnw" and @lemma="dit"]  
and node[@rel="hd" and  
@pt="ww" and @lemma="zijn"]  
and node[@rel="predc" and  
@cat="np" and  
node[@rel="det" and  
@pt="lid" and @lemma="een"]  
and node[@rel="hd" and  
@pt="n" and @lemma="zin"]]]
```

# XPATH



```
//node[@cat="smain" and  
node[@cat="su" and  
@pt="adit" and  
@pt="w" and @lemma="zijn"]  
and node[@cat="dc" and  
@cat="n" and  
@pt="n" and @lemma="zin"]]]
```

# XPATH



# GrETEL

- **Greedy Extraction of Trees for Empirical Linguistics**
- **Query treebanks by example**



# GrETEL

- **Greedy Extraction of Trees for Empirical Linguistics**
- **Query treebanks by example**
  - ➔ No or limited knowledge of data structures and/or formal query languages needed





# the user

1. Example sentence

2. Indicate relevant items  
of the sentence

3. (Adapt XPath)  
Select treebank

4. Inspect results



# GrETEL

• Parser (Alpino)

• Automatically generate  
XPath expression

• Present results



# OUTLINE

- GrETEL in a nutshell
- **GrETEL demo**
  - **Case study**
  - Search options
- Conclusions and future work

# CASE STUDY

## *Infinitivus Pro Participio (IPP) constructions in Dutch*

*Hij heeft Marie horen zingen.*

‘He has heard Mary sing.’

*... dat Jan niet is kunnen komen.*

‘... that Jan was not able to come.’

# CASE STUDY

## *Infinitivus Pro Participio (IPP) constructions in Dutch*

*Hij heeft Marie horen/\*gehoord zingen.*

‘He has heard Mary sing.’

*... dat Jan niet is kunnen/\*gekund komen.*

‘... that Jan was not able to come.’

# GrETEL ONLINE

Contact



Nederbooms

- About
- Projects
- ▼ Tools
  - GrETEL
  - GrETEL for LASSY
  - GrETEL for CGN
  - Manual and docs
  - History

Home > Tools > GrETEL

## GrETEL

### What is *GrETEL*?

Greedy Extraction of Trees for Empirical Linguistics

GrETEL is a query engine in which linguists can use a natural language example as a starting point for searching a treebank with limited knowledge about tree representations and formal query languages. By allowing users to search for constructions which are similar to the example they provide, we hope to bridge the gap between traditional and computational linguistics.

# INPUT



## Nederbooms

Home > Tools > GrETEL > GrETEL for CGN

### GrETEL for CGN (v1.2)

Please provide an **input example**

- About
- Projects
- Tools
  - GrETEL
    - GrETEL for LASSY
    - GrETEL for CGN**
    - Manual and docs
    - History

# SELECTION MATRIX

## GrETEL for CGN (v1.2)

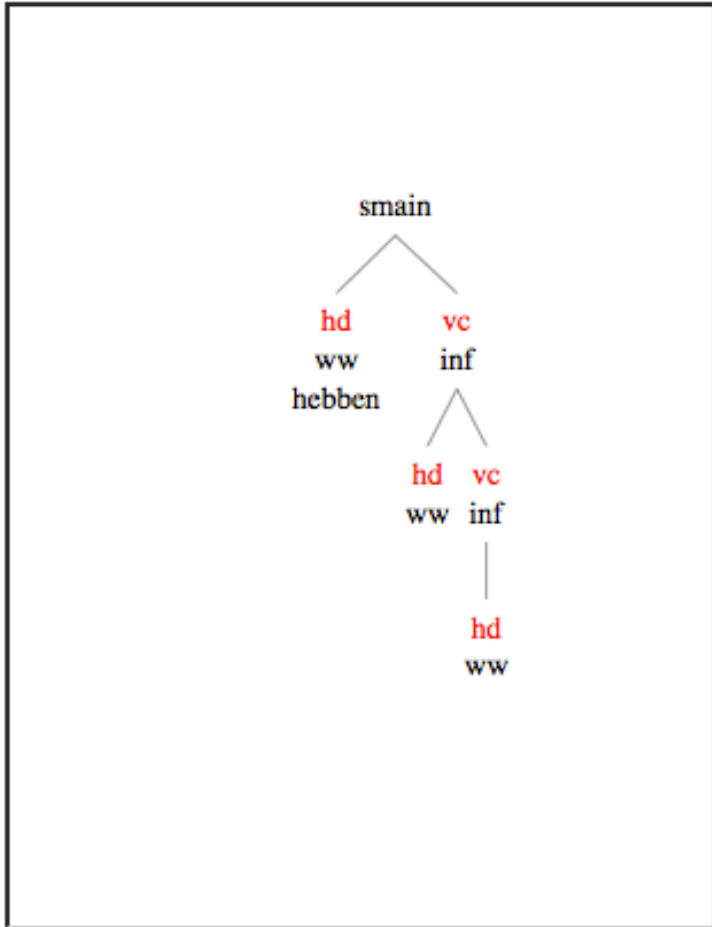
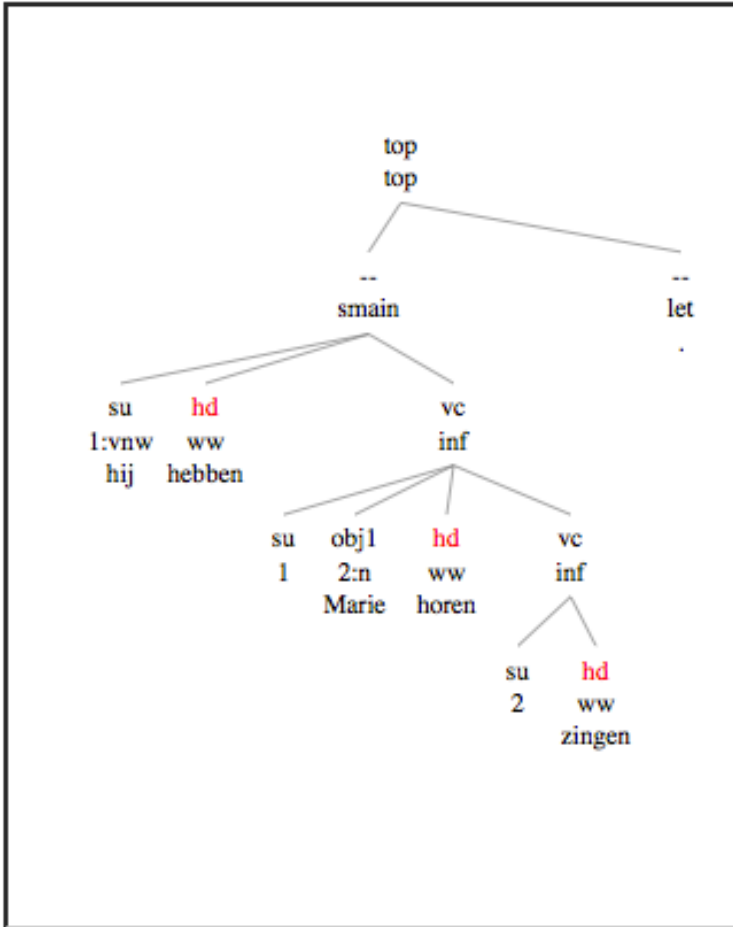
Please indicate the relevant parts of the sentence

<b>sentence</b>		Hij	heeft	Marie	horen	zingen	.
<b>relevant nodes</b>	<b>pos</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
	<b>extended pos</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<b>lemma</b>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	<b>token</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>optional nodes</b>		<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

# SELECTION GUIDELINES

## GUIDELINES

- **pos**: Short part-of-speech tag (e.g. noun, verb, prep) [[list of all pos tags](#)]
- **extended pos**: Long part-of-speech tag. Besides the part-of-speech, the tags contain more detailed information on person, gender, number, tense, agreement etc. For example: `N(soort,mv,basis)`, `WW(pv,tgw,ev)`, `VNW(pers,pron,nomin,vol,2v,ev)`. [[list of all pos tags](#)]
- **lemma**: Word form that generalizes over inflected forms. For example: *zin* is the lemma of *zin*, *zinnen*, and *zinnetje*; *gaan* is the lemma of *ga*, *gaat*, *gaan*, *ging*, *gingen*, and *gegaan*. Lemma is case insensitive (except for proper names). Note that the short part-of-speech tag will be taken into account as well.
- **token**: The exact word form. This is a case sensitive feature. Note that the short part-of-speech tag and the lemma will be taken into account as well.



# XPATH GENERATOR

**XPath query** generated from the input example. You can adapt it if necessary. If you are dealing with a long query, the [XPath beautifier](#) might come in handy.

[[download original XPath](#)]

```
//node[@cat="smain" and node[@rel="hd" and @pt="ww" and @lemma="hebben"] and
node[@rel="vc" and @cat="inf" and node[@rel="hd" and @pt="ww"] and node[@rel="vc" and
@cat="inf" and node[@rel="hd" and @pt="ww"]]]]
```



# TREEBANK SELECTION

## CGN core corpus

Treebank		Contents	# Sentences	# Words	# Sentences	# Words	# Sentences	# Words
<input checked="" type="checkbox"/> NL	<input checked="" type="checkbox"/> VL		NL		VL		TOTAL	
<input checked="" type="checkbox"/> NA	<input checked="" type="checkbox"/> VA	Spontaneous conversations ('face-to-face')	50,239	302,828	22,881	147,418	73,120	450,246
<input checked="" type="checkbox"/> NB	<input checked="" type="checkbox"/> VB	Interviews with teachers of Dutch	2,484	25,724	4,289	34,158	6,773	59,882
<input checked="" type="checkbox"/> NC	<input checked="" type="checkbox"/> VC	Telephone conversations (recorded via a switchboard)	11,649	70,084	3,142	19,984	14,791	90,068
<input type="checkbox"/> ND	<input checked="" type="checkbox"/> VD	Telephone conversations (recorded on MD)	0	0	929	6,309	929	6,309
<input checked="" type="checkbox"/> NE	<input type="checkbox"/> VE	Simulated business negotiations	3,123	25,524	0	0	3,123	25,524
<input checked="" type="checkbox"/> NF	<input checked="" type="checkbox"/> VF	Interviews/discussions/debates (broadcast)	6,290	75,167	2,617	25,122	8,907	100,289
<input checked="" type="checkbox"/> NG	<input checked="" type="checkbox"/> VG	(Political) discussions/debates/meetings (non-broadcast)	1,166	25,125	543	9,009	1,709	34,134
<input checked="" type="checkbox"/> NH	<input checked="" type="checkbox"/> VH	Lessons recorded in the classroom	3,064	26,004	1,395	10,116	4,459	36,120
<input checked="" type="checkbox"/> NI	<input checked="" type="checkbox"/> VI	Live (sports) commentaries (broadcast)	2,251	25,002	1,026	10,147	3,277	35,149
<input checked="" type="checkbox"/> NJ	<input checked="" type="checkbox"/> VJ	Newsreports (broadcast)	2,259	25,084	536	7,686	2,795	32,770
<input checked="" type="checkbox"/> NK	<input checked="" type="checkbox"/> VK	News (broadcast)	1,923	25,353	558	7,306	2,481	32,659
<input checked="" type="checkbox"/> NL	<input checked="" type="checkbox"/> VL	Commentaries/columns/reviews (broadcast)	1,857	25,082	601	7,431	2,458	32,513
<input checked="" type="checkbox"/> NM	<input checked="" type="checkbox"/> VM	Ceremonious speeches/sermons	444	5,190	107	1,894	551	7,084
<input checked="" type="checkbox"/> NN	<input checked="" type="checkbox"/> VN	Lectures/seminars	593	14,921	701	8,159	1,294	23,080
<input type="checkbox"/> NO	<input checked="" type="checkbox"/> VO	Read speech	0	0	3,256	44,144	3,256	44,144
<b>CGN core</b>		Complete treebank	87,342	671,088	42,581	338,883	129,923	1,009,971

# RESULTS

## IPP constructions in CGN

*Hij heeft Marie horen zingen.*

‘He has heard Mary sing.’

→ **344 hits**

# RESULTS: table

**RESULTS:** 344 matches in 339 sentences (out of 129,923 sentences)

[Download results\[`csv`\]](#)

[Show/hide detailed results](#)

TREEBANK	HITS	MATCHING SENTENCE	SENTENCES IN TREEBANK
NA	94	93	50,239
VA	64	63	22,881
NB	3	3	2,484
VB	9	9	4,289
NC	13	13	11,649
VC	10	10	3,142
VD	4	4	929

# RESULTS: data

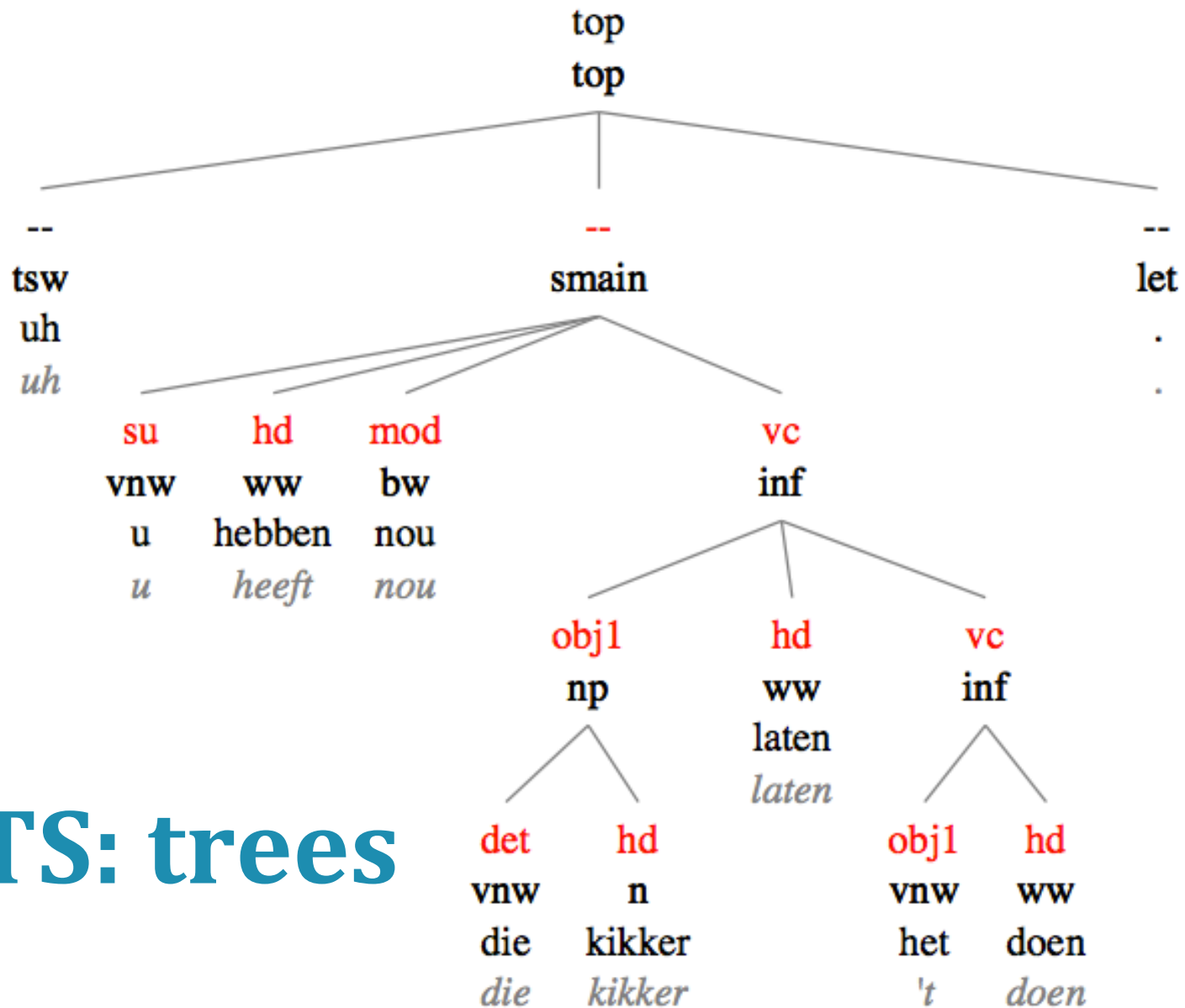
SENTENCE ID	MATCHING SENTENCE	HITS	DISPLAY OPTIONS
fna000724__28	'k heb alleen geen lolly's kunnen krijgen want die kon je alleen weer in combinatie krijgen met ja zo'n zak met zo'n pretpakket met allerlei andere dingen en ik denk ja wat moet 'k met al die zoetigheid .	1	[full screen] [XML]
fvg600014__64	'k heb blijkbaar uh mij toch laten vertellen dat in de Kempen men een groot een grote mest*a mestverwerkingsfabrieken wil opstarten .	1	[full screen] [XML]
fva400258__41	'k heb d'r heel veel zien staan xxx .	1	[full screen] [XML]
fna000940__32	nou meid ik 'k heb daar nou inwendig zitten janken op die bank daar hoor .	1	[full screen] [XML]
fva400083__238	ja 'k heb daarstraks ook één achter de bomen zien lopen daarachter een rooie ...	1	[full screen] [XML]
fva400454__114	en 'k heb dan ne keer 's avonds beginnen oefeningen maken en dan doe ik dat doe 'k gans de nacht verder hé .	1	[full screen] [XML]
fva400370__34	'k heb dat ergens zien liggen en 'k heb dat meegebracht .	1	[full screen] [XML]

# RESULTS: data

SENTENCE ID	MATCHING SENTENCE	HITS	DISPLAY OPTIONS
fna000724__28	'k heb alleen geen lolly's kunnen krijgen want die kon je alleen weer in combinatie krijgen met ja zo'n zak met zo'n pretpakket met allerlei andere dingen en ik denk ja wat moet 'k met al die zoetigheid .	1	[full screen] [XML]
fvg600014__64	'k heb blijkbaar uh mij toch laten vertellen dat in de Kempen men een groot een grote mest*a mestverwerkingsfabrieken wil opstarten .	1	[full screen] [XML]
fva400258__41	'k heb d'r heel veel zien staan xxx .	1	[full screen] [XML]
fna000940__32	nou meid ik 'k heb daar nou inwendig zitten janken op die bank daar hoor .	1	[full screen] [XML]
fva400083__238	ja 'k heb daarstraks ook één achter de bomen zien lopen daarachter een rooie ...	1	[full screen] [XML]
fva400454__114	en 'k heb dan ne keer 's avonds beginnen oefeningen maken en dan doe ik dat doe 'k gans de nacht verder hé .	1	[full screen] [XML]
fva400370__34	'k heb dat ergens zien liggen en 'k heb dat meegebracht .	1	[full screen] [XML]

“greedy” search

u heeft nou die kikker uh 't laten doen .  
fnj007296\_\_51



# RESULTS: trees

# RESULTS

## IPP constructions in CGN

*Hij heeft Marie horen zingen.*

‘He has heard Mary sing.’

→ **344 hits**

*... dat Jan niet is kunnen komen.*

‘... that Jan was not able to come.’

→ **24 hits**

# MORE RESULTS

## Option 1: Use different queries

*Hij heeft Marie horen zingen.*

‘He has heard Mary sing.’

→ 344 hits

*... dat hij Marie heeft horen zingen.*

‘... that he has heard Mary sing.’

→ 79 hits

*... dat Jan niet is kunnen komen.*

‘... that Jan was not able to come.’

→ 24 hits

*Jan is niet kunnen komen.*

‘Jan was not able to come.’

→ 120 hits



# MORE RESULTS

## Option 2: Adapt the query

```
//node[@cat="smain" and node[@rel="hd" and @pt="ww" and  
@lemma="hebben"] and node[@rel="vc" and @cat="inf" and  
node[@rel="hd" and @pt="ww"] and node[@rel="vc" and  
@cat="inf" and node[@rel="hd" and @pt="ww"]]]]
```

```
//node[(@cat="smain" or @cat="ssub") and node[@rel="hd"  
and (@lemma="hebben" or @lemma="zijn")] and  
node[@rel="vc" and @cat="inf" and node[@rel="hd" and  
@pt="ww"] and node[@rel="vc" and @cat="inf" and  
node[@rel="hd" and @pt="ww"]]]]
```

→ 566 hits (one sentence matches twice: [fva400364](#) 10)

# OUTLINE

- GrETEL in a nutshell
- **GrETEL demo**
  - Case study
  - **Search options**
- Conclusions and future work

# SEARCH OPTIONS

→ Below annotation matrix

## OPTIONS

- Respect word order
- Ignore properties of the dominating node
- Split extended pos tags

# SEARCH OPTIONS

## PP-over-V

- V + PP
- ... *dat hij opstond met een kater.*  
'... that he woke up with a hangover.'
  
- PP + V
- ... *dat hij met een kater opstond.*  
... that he with a hangover woke-up  
'... that he woke up with a hangover.'

# SEARCH OPTIONS

## PP-over-V in LASSY small

- V + PP
- ... *dat hij opstond met een kater.*  
'... that he woke up with a hangover.'

WR-P-E-I-0000039352.p.3.s.39	De verscheidenheid qua bordjes is groot , zo loopt bijvoorbeeld Schutterij 't Zandakker Gilde Sint Jan Venray met een prachtig roodkoperen geslagen draagbord , terwijl <b>andere verenigingen met een houten draagbord lopen</b> .	1	[full screen] [XML]
dpc-vla-001171-nl-sen.p.29.s.6	Het was in het Kaaitheater dat <b>Anne Teresa De Keersmaecker een podium vond in het begin van haar carrière</b> .	1	[full screen] [XML]
dpc-eli-000943-nl-sen.p.24.s.1	Het Comité voor geneesmiddelen voor menselijk gebruik ( CHMP ) heeft geconcludeerd dat <b>Ariclaim een bescheiden effect vertoonde in studies</b> maar dat dit effect een voordeel kan zijn voor vrouwen met matige tot ernstige stress-urine-incontinentie , daar er momenteel geen alternatieve farmacologische behandeling voor deze aandoening bestaat .	1	[full screen] [XML]
dpc-vla-001161-nl-sen.p.54.s.1	Een goed jaar na de start van het nieuwe financieringsinstrument kunnen we besluiten dat <b>ARKimedes na een zwakke start nu toch wel goed op gang komt</b> .	1	[full screen] [XML]

2,895 matches in 2,769 sentences

But: results include PP + V as well!

# SEARCH OPTIONS

## PP-over-V in LASSY small

- V + PP + word order option
- ... *dat hij opstond met een kater.*  
'... that he woke up with a hangover.'

WS-U-E-A-0000000218.p.21.s.3	Het ging destijds eigenlijk al meteen mis : het voorprogramma met ondermeer André van Duin en de Fouryo's was veel te rustig voor een publiek <b>dat kwam voor rok-en-rol</b> .	1	[full screen] [XML]
dpc-med-000686-nl-sen.p.26.s.3	Het belangrijkste voordeel van het product is de associatie van een lage dosis ethyniloestradiol met een progestativum van de tweede generatie ( cfr de polemiek van het einde van vorige eeuw over het thrombo-embolisch risico <b>dat lager is met tweede generatie progestativum in vergelijking met deze van de derde generatie</b> . )	1	[full screen] [XML]
WR-P-E-I-0000024561.p.4.s.68.3	Brels eerste album op het Barclay-label bestaat voor een groot deel uit studio-versies van materiaal <b>dat luisteraars al kenden van het Olympia-concert</b> .	1	[full screen] [XML]
dpc-rou-000983-nl-sen.p.8.s.11	Het verhaal van Rodenbach , <b>dat net als wijn rijpt op eikenhouten vaten</b> , past daar perfect bij , maar was nog niet verteld . "	1	[full screen] [XML]

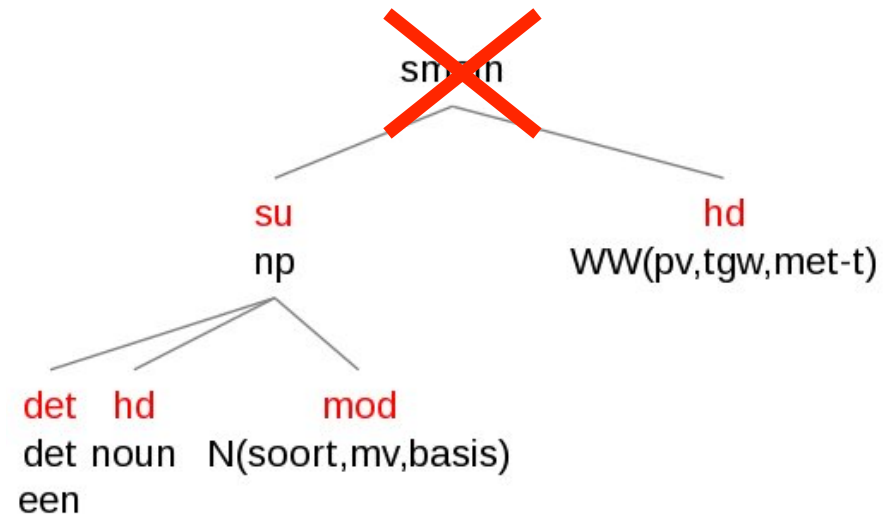
789 matches in 777 sentences

Results only include V + PP

# SEARCH OPTIONS

## OPTIONS

- Respect word order
- Ignore properties of the dominating node
- Split extended pos tags



XPath query generated from the input example. You can adapt it if necessary. [[download original XPath](#)]

```
//node[@cat="sm n"] and node[@rel="su" and @cat="np" and node[@rel="det" and @root="een" and @pos="det"] and node[@rel="hd" and @pos="noun"] and node[@rel="mod" and @postag="N(soort,mv,basis)" and @pos="noun"]] and node[@rel="hd" and @postag="WW(pv,tgw,met-t)" and @pos="verb"]
```

# SEARCH OPTIONS

## OPTIONS

- Respect word order
- Ignore properties of the dominating node
- Split extended pos tags

XPath query generated from the input example. You can adapt it if necessary. [[download original XPath](#)]

```
//node[@cat="smain" and node[@rel="su" and @cat="np" and node[@rel="det" and @rest="con" and @pos="det"] and node[@rel="hd" and @pos="noun"] and node[@rel="mod" and @postag="N(soort,mv,basis)" and @pos="noun"]] and node[@rel="hd" and @postag="WW(pv,tgw,met-t)" and @pos="verb"]]
```



# SEARCH OPTIONS

## OPTIONS

- Respect word order
- Ignore properties of the dominating node
- Split extended pos tags

**XPath query** generated from the input example. You can adapt it if necessary. [[download original XPath](#)]

```
//node[@cat="smain" and node[@rel="su" and @cat="np" and node[@rel="det" and @root="een" and @pos="det"] and node[@rel="hd" and @pos="noun"] and node[@rel="mod" and @postag="N(soort,mv,basis)" and @pos="noun"]] and node[@rel="hd" and @postag="WW(pv,tgw,met-t)" and @pos="verb"]]
```

**XPath query** generated from the input example. You can adapt it if necessary. [[download original XPath](#)]

```
//node[@cat="smain" and node[@rel="su" and @cat="np" and node[@rel="det" and @root="een" and @pos="det"] and node[@rel="hd" and @pos="noun"] and node[@rel="mod" and @getal="mv" and @graad="basis" and @pos="noun" and @ntype="soort"]] and node[@rel="hd" and @pvtijd="tgw" and @wvform="pv" and @pos="verb" and @pvagr="met-t"]]
```

# SEARCH OPTIONS

## OPTIONS

- Respect word order
- Ignore properties of the dominating node
- Split extended pos tags

**XPath query** generated from the input example. You can adapt it if necessary. [[download original XPath](#)]

```
//node[@cat="smain" and node[@rel="su" and @cat="np" and node[@rel="det" and @root="een" and @pos="det"] and node[@rel="hd" and @pos="noun"] and node[@rel="mod" and @postag="N(soort,mv,basis)" and @pos="noun"]] and node[@rel="hd" and @postag="WW(pv,tgw,met-t)" and @pos="verb"]]
```

**XPath query** generated from the input example. You can adapt it if necessary. [[download original XPath](#)]

```
//node[@cat="smain" and node[@rel="su" and @cat="np" and node[@rel="det" and @root="een" and @pos="det"] and node[@rel="hd" and @pos="noun"]] and node[@rel="mod" and @getal="mv" and @graad="basis" and @pos="noun" and @ntype="soort"]] and node[@rel="hd" and @pvtijd="tgw" and @wvform="pv" and @pos="verb" and @pvagr="met-t"]]
```

# SEARCH OPTIONS

## LASSY Small

<input checked="" type="checkbox"/>	Treebank	Contents	# Sentences	# Words
<input checked="" type="checkbox"/>	DPC	Dutch Parallel Corpus	11,716	193,029
<input checked="" type="checkbox"/>	Wikipedia	Dutch Wikipedia pages	7,341	83,360
<input checked="" type="checkbox"/>	WR-P-E	E-magazines, newsletters, teletext pages, web sites, Wikipedia	14,420	232,631
<input checked="" type="checkbox"/>	WR-P-P	Books, brochures, guides and manuals, legal texts, newspapers, periodicals and magazines, policy documents, proceedings, reports, surveys	17,691	281,424
<input checked="" type="checkbox"/>	WS-U	Auto cues, news scripts, text for the visually impaired	14,032	184,611
	<b>LASSY Small</b>	Complete treebank	65,200	975,055

## OPTION

Include context (one sentence before and after the matching sentence)

# SEARCH OPTIONS

SENTENCE ID	MATCHING SENTENCE	HITS	DISPLAY OPTIONS
WR-P-E- I-0000020972.p.4.s.59.2	Levenscyclus <b>Een kolonie hommels sterft elk najaar</b> , alleen de bevruchte jonge koninginnen blijven in leven en overwinteren alleen .	1	[full screen] [XML]
WR-P-P- H-0000000076.p.1.s.7	Tien meter lang . ' <b>Een swingende groep kleuter-groupies probeert ondertussen uitgelaten het podium te beklimmen</b> . Het gaat hier niet om lieve luistermuziek maar om stevige kleuterpop .	1	[full screen] [XML]
WR-P-P- I-0000000111.p.2.s.1	<b>Een tiental commissies in het Amerikaanse Congres begint eveneens aan zijn onderzoek naar het failliet</b> . Een van de commissies wil informatie over het energiebeleid krijgen van de Amerikaanse vice-president Dick Cheney .	1	[full screen] [XML]
WS-U-E- A-0000000206.p.18.s.7	Een sprinkhaan kan per dag zijn eigen gewicht aan voedsel opeten . <b>Eenmaal geland verslindt een wolk sprinkhanen in korte tijd de oogst van een heel seizoen</b> . Alleen in Mali werden de afgelopen dagen al 42 springhanenzwermen waargenomen .	1	[full screen] [XML]
WS-U-E- A-0000000011.p.24.s.1	<b>Er dient zich een nieuwe generatie schaatsters aan</b> , dat beweest Mark Tuitert afgelopen weekend in Heerenveen : want hij was na vier afstanden de sterkste : en is de nieuwe Europese Kampioen allround .	1	[full screen] [XML]
WR-P-E- I-0000039352.p.3.s.259	De eerste ronde valt het zwaarste , getuige het feit dat meestal meer dan de helft van de deelnemers dan reeds ' sneuvelt ' . <b>Gelukkig voor de talloze toeschouwers krijgt een aantal schutters weer snel een ritme te pakken</b> .	1	[full screen] [XML]

# OUTLINE

- GrETEL in a nutshell
- GrETEL demo
  - Case study
  - Search options
- **Conclusions and future work**

# CONCLUSIONS

- **GrETEL**: search engine for Dutch treebanks
- Input = natural language example
- Output = sample of similar sentences
- Syntactic concordancer
- Available online (via *Mozilla Firefox*)
- No installation required



# FUTURE WORK

- **GrETEL 2.0**
  - CLARIN-NTU
  - Query very large treebanks
  - Improve user interface
  - Include SoNaR corpus (ca 500M tokens, 42M sentences)

**\* coming soon \***

- **AfriBooms**
  - GrETEL for Afrikaans
  - Include other treebank formats



# Try it yourself at

<http://nederbooms.ccl.kuleuven.be/eng/gretel>

## Thanks for your attention!



**KU LEUVEN**