

# Focus Particles inside PPs: A comparison between Dutch, English, and German

Gosse Bouma, Petra Hendriks, and Jack Hoeksema

April 2005

## Focus Particles

- Kim **only** listens to JAZZ
- Kim **only** LISTENS to jazz
- Kim **even** PLAYS soccer
- Kim was **also** warned by her DOCTOR

## Overview

- Literature Overview,
- Using Corpus Data
  - ★ Tagged
  - ★ Syntactically Analyzed
  - ★ Unanalyzed
- Discussion
  - ★ Using Corpora
  - ★ Using WWW

## Previous Work

- A **Focus Particle** selects some constituent as **Focus**,
- F Particle and Focus often, but not always, **adjacent**
- F Particle attaches to a maximal projection

## Jacobs 83

- German, GPSG
- F Particle attaches to some projection of V

daß Luise [**nur** [ihrem Arzt ein Auto vermachte]]

\*Peter träumt von **nur** seiner Frau

\*Den Sohn **nur** des Grafen liebt sie

[**nur** [ihrem Arzt vermachte Luise ein Auto]]

## Bayer 86

- English, German, QR at LF
- QR is movement, subject to island constraints
  - ★ No extraction from PP → no F Part in PP
  - ★ No F Part inside PPs in German
  - ★ F Part inside PPs ok in English
- *Many speakers find ? and \* examples of Rooth ok*
  - ★ They are sending eggs to **even** Marie

## Rooth 85

- English, MG
- *only* adjoins to VP, DP,
- *even* adjoins to VP, DP, S
- F Particles can occur anywhere where Q NPs can occur..

\*The library is closed on **only** Sunday

\*The entrance **only** to the Santa Monica freeway was blocked

## Büring and Hartmann 2001

- Variant of Jacobs 83 for German
- In English, F Particles attach to projection of V or DP
- No motivation given for the difference

## Literature Summary

- F Particles ok inside PP and DP in English
  - ★ But Rooth notes these are marginal
- F Particles impossible inside PP and DP in German
  - ★ Bayer: P-stranding → F Part inside PP
- Hoeksema & Zwarts (91), Hoeksema (99)
  - ★ **Dutch**: no P-stranding, but F Part inside PP

## Using Corpus Data

- Authors do not agree on the data
- Reasons why F Particles inside PPs are relatively rare, are not well understood
- Maybe Corpus Data can help

## Dutch, 1st Attempt

- Algemeen Dagblad 1994, 17 million words,
  - ★ Annotated automatically with **Part of Speech**
- Search for Prep + Focus Particle bigrams

## Dutch allows F Part in PP

Voor de ouderen **met alleen een AOW-uitkering** stijgt dan het netto-inkomen.

Zonder toestemming gebruikmaken **van alleen de namen van voetballers en topsporters** mag evenmin

Bovendien hebben we de komende weken een druk programma, **met ook doordeweekse wedstrijden**.

Dat gaat soms tot de grenzen **van ook ons incasservermogen**

Tot in de vezels **van zelfs de grootste burgerplurk en**

**VVD'er-van-het-oude-type** bleek paniek te zijn doorgedrongen

De gevolgen kunnen variëren van depressies en echtscheidingen **tot zelfs suïcide**

## False Hits

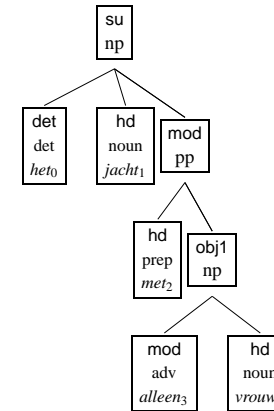
De ploeg won vier wedstrijden **zonder ook maar** een **tegendoeelpunt**

Maar ze kunnen niet worden gevonden **zonder [ook over andere kwesties te praten]**

In de notitie stelt de bond **voor ook arbeiders van andere bedrijven** in te huren

Schröder **uit ook kritiek** op D66 en PvdA

## Dependency Trees



## Dutch, 2nd Attempt

- AD, NRC, 1994, 1995, 78 mln words,
  - ★ Parsed Automatically using Alpino
- Search for PP containing an NP containing a focus Particle

## Results

P + alleen	P-Part-NP		Part-P-NP	
	#	%	#	%
met	412	32	889	68
van	86	21	321	79
voor	30	2	1389	98
tot	12	12	92	88
door	11	2	642	98
in	11	0.4	2442	99.6
op	11	1	1091	99
in plaats van	10	91	1	9

## Results

P + ook	P-Part-NP		Part-P-NP	
	#	%	#	%
met	49	2	2962	98
van	39	1.4	2638	98.6
in	6	0.04	13641	99.96
tot	4	0.5	797	99.5

P + zelfs	P-Part-NP		Part-P-NP	
	#	%	#	%
van	51	10	455	90
tot	33	9	315	91
met	26	3	788	97
voor	9	1.4	644	98.6

## Trigram Variation

- Google Counts for fixed P-Part-Det/N

P + alleen + det/N	P-Part-det/N		Part-P-det/N	
	#	%	#	%
de	6,260	22	21,700	78
die	442	32	934	68
deze	173	17	830	83
mijn	1560	36	2800	64
mannen	101	45	124	55
vrouwen	157	51	151	49
geld	24	8	281	92

## Lexical Variation

- among F Particles and Prepositions
- **Met** seems to allow P Part order more easily than other Ps
- **Van** usually modifies nouns → Part-P order marked as well

Dus niet [ een verwijzing [ van alleen de huisarts ] ]

? Dus niet een verwijzing alleen van de huisarts

Dus niet alleen een verwijzing van de huisarts

- Many examples involve Phrasal Prepositions (*op grond van, in plaats van, op rekening van*).

## German

- Cosmas corpus (web access), 2 billion words!

mit nur	P-Part-det		Part-P-det	
	#	%	#	%
dem	3	0.07	4425	99.93
der	0	0	3639	0
diesem	3	1.6	184	98.4
dieser	0	0	155	100

Man kann sich die Maschine **mit nur dem Piloten** oder 1-3 Passagieren auswählen

Der Stadtrat hat bei einer Sitzung **mit nur diesem**

**Tagesordnungspunkt** ein Ansuchen...abgelehnt

## English

- British National Corpus, <http://view.byu.edu>, 100 mln words

	P-Part-det		Part-P-det	
	#	%	#	%
with only the	186	37	319	63
for only the	44	8.5	476	91.5
in only the	19	1.5	1213	98.5
by only the	8	1.4	545	98.6
of only the	24	9	251	91
to only the	10	1.5	655	98.5

- English patterns like Dutch

## Discussion

- Dutch was never addressed by Jacobs, Rooth, Bayer, B&H
  - ★ But Bayer makes a prediction about Dutch...
- Heterogenous Data
  - ★ Newspaper vs Google
  - ★ Anyone can place stuff in the internet..
  - ★ How to quantify Google data?

## Conclusions from Corpus Data

- F Part inside PP rare in all three languages,
- But not even excluded in German,
- Lexical, Syntactic, Semantic & Pragmatic factors seem to play a role in distribution
- Suggests purely syntactic accounts cannot be correct

## On Google Counts

- Veronis, Beaver, Liberman (Language Log)
- Illogical behaviour of OR
  - ★ Chirac: 3.2 mln, Chirac or Sarkozy 1.7 mln, Chirac and Sarkozy 1.6 mln, Chirac and Chirac: 1.7 mln, Chirac Chirac: 1.7 mln
  - ★ Bouma : 457.000, Bouma OR Bouma: 503.000
- Number of Hits fluctuates strongly
  - ★ the (Feb) : 8 bln , the (Mar) : 3.2 bln
- Other Oddities
  - ★ Economist in English/General pages: 60%, the: 1%