

# Investigating *heel, erg, zeer* with PaQu

Jan Odijk, Tin-dag, Utrecht, 2015-02-07

## (1) Background

- a. AutoSearch: (enrich,) upload & search (expected March 2015, INL)
  - PoS-tags, Corpus Modern Dutch interface
- b. PaQu: upload (, enrich) and search (July 2015, V1 available, RUG)  
syntactic structures, Groningen Word Relations Search Application, User tests on-going

## (2) Basic Facts

- a. *Heel erg zeer* are (near-)synonyms meaning 'very'
- b. *Heel* can modify adjectival (A) predicates only
- c. *Erg en zeer* can modify A, verbal (V) and prepositional (P) predicates

## (3) Illustration

- a. Hij is daar heel /erg / zeer blij over
- b. Hij is daar \*heel / erg /zeer in zijn sas mee
- c. Dat verbaast mij \*heel / erg / zeer  
(*very* in English is like Dutch *heel* (v. *very much*)  
See [Odijk 2011, 2014] for more data and qualifications

## (4) Assessment of the facts

- a. Distinction is purely syntactic
- b. Cannot be derived from semantic differences
- c. No correlation found with other known facts (see (5))
- d. Cannot be derived from general (universal) principles
- e. → must be acquired by L1 learners of Dutch

## (5) Correlation with other differences? Unlikely! Then we must find {*heel*} v. {*erg, zeer*} .

But we only find other oppositions:

- a. Meaning? {*heel, zeer*} v. {*erg*}
- b. Adverbial Inflection? {*heel, erg*} v. {*zeer*}
  - Hele / erge / \*zere grote handen
- c. Comparative / Superlative? {*heel, zeer*} v. {*erg*}
  - \*heler / erger / \*zeerder ziek (dan Piet)
  - \*Het heelst / het ergst / \*het zeerst ziek
- d. Modification by *heel erg zeer*? {*heel, zeer*} v. {*erg*}
  - \*heel heel<sup>1</sup> / \*erg heel / \* zeer heel ziek
  - Heel erg / ?erg erg<sup>1</sup> / zeer erg ziek
  - \*Heel zeer / \*erg zeer / \*zeer zeer<sup>1</sup> ziek
- e. Early use by children? {*heel*} v. {*erg*} v. {*zeer*} [Odijk 2014]
- f. Formality? {*zeer*} (formal) v. {*heel, erg*} (neutral)

---

<sup>1</sup> Well-formed as repeated modifiers

**(6) Research Questions**

- a. How can children acquire the fact that *erg* and *zeer* can modify A, V and P predicates (in L1 acquisition)?
- b. How can children acquire the fact that *heel* can modify A but canNOT modify V and P predicates (in L1 acquisition)?
- c. What kind of evidence do children have access to for acquiring such properties?
- d. Is there a relation with the time of acquisition?
- e. Is there a role for *indirect negative evidence* (absence of evidence interpreted as evidence for absence)?
- f. ...

**(7) CHILDES corpora**

- a. Use Dutch CHILDES corpora to investigate this
- b. Problem: ambiguity of the relevant words
- c. Dutch CHILDES corpora do NOT have (reliable) pos-tags and no syntactic parses at all
- d. Done manually for Van Kampen Corpus [Odijk 2014:91]
- e. PaQu (Parse and Query) automates this

**(8) Ambiguity of *heel*, *erg*, *zeer***

word	Morphosyntax	Syntax	Meaning
<i>heel</i>	A	Mod N	(1) 'whole' (2) 'large'
		Mod A	'very'
	Vf		(1) 'heal' (2) 'receive'
<i>erg</i>	N utrum		'erg'
	N neutrum		'evil'
	A	Mod N, predc	'bad', 'awful'
		Mod A V P	'Very'
<i>zeer</i>	N		'pain'
	A	Mod N, predc	'painful'
		Mod A V P	'very'

**(9) PaQu**

- a. Search for morpho-syntactic information and syntactic dependency relations

- b. Distinction relevant ones v. irrelevant ones can now be made mostly automatically  
 c. <http://zardoz.service.rug.nl:8067/>
- (10) **Small Experiment** (was intended as a user test)
- Take all Dutch CHILDES corpora
  - Select all adult utterances containing *heel, erg* or *zeer*
  - Clean the utterances, e.g.
    - *ja, maar <we be> [//] we bewaren (he)t ook* →
    - *ja, maar we bewaren het ook*
  - Gather statistics and draw conclusions
- (11) **Accuracy**
- Manual annotation of Van Kampen corpus used as gold standard (Acc)
  - Alpino makes finer distinctions: I mapped these
  - Annotation errors in the gold standard: revised gold standard (Rev Acc)

word	Acc	Rev Acc
heel	0.94	0.95
erg	0.88	0.91
zeer	0.21	0.21

- (12) **Caveats**
- It concerns (cleaned) adult speech
  - It concerns relatively short sentences, explicitly separated
  - It mostly concerns a very local grammatical relation
  - Most problematic for *zeer*: *zeer doen*

(13) **Results**

Results	mod A	mod N	Mod V	mod P	predc	other	unclear	Total
<b>heel</b>	886	46	2	2	14	0	2	<b>952</b>
<b>erg</b>	347	27	109	0	187	5	0	<b>675</b>
<b>zeer</b>	7	1	83	0	19	21	7	<b>138</b>

(14) **Interpretation**

- Overwhelming # examples for mod A for *heel*
- Large # examples for mod A and mod V for *erg*
- Very few examples for *zeer* (mod V mostly wrong parses)
- No examples of mod P / mod V for *heel* at all (the 4 are wrong parses)
- PP predicates with *zeer, erg*: *op prijs stellen, in de smaak vallen* only (analyzed as mod V) – 3 occurrences

(15) **Conclusions**

- a. Linguistics
  - No examples for mod P: how to explain *heel v. erg, zeer*?  
Overwhelmingness of mod A for *heel*?
  - Are the current Dutch CHILDES corpora representative enough to draw reliable conclusions?
- b. PaQu
  - PaQu is very useful for doing better and more efficient manual verification of hypotheses
  - In some cases its fully automatically generated parses and their statistics can reliably be used directly (though care is required!)

(16) **Future Work**

- a. Similar experiments for the children's speech (cf. [Odijk 2014:34])
- b. Similar experiments for *te v. overmatig; worden v. raken* and others
- c. Extend PaQu to include all relevant `metadata`
- d. Extend PaQu to natively support common formats such as CHAT, Folia, TEI, ...
- e. Make similar system for GrETEL (<http://nederbooms.ccl.kuleuven.be/eng/gretel>)
- f. Manually verify (parts of) parses for CHILDES corpora (UU AnnCor project)

## References

- [Odijk 2011] Odijk, J. , "User Scenario Search", internal CLARIN-NL document, April 13, 2011. [[docx](#)]
- [Odijk 2014] Odijk, J. , 'CLARIN: What's in it for Linguists?', Uilendag lecture, Utrecht, Mar 27, 2014. [[pptx](#)]