

GE HAD DIEN EEN KEER MOETEN EN ZIEN!

Het “Gesproken Corpus van de (Zuidelijk-)Nederlandse Dialecten”

Anne Breitbarth & Anne-Sophie Ghyselen

Dag van de Nederlandse Zinsabouw, 21/12/2018

ONTKENNING IN ZUIDELIJK NEDERLANDS

- (1) Hij heeft **niemand nie** gezien.
Hij gaat **nooit nie** naar school.

(Aelbrecht 2008)

- (2) k=een al overal gezocht in us en k'**en** vinden ze **nievers**.

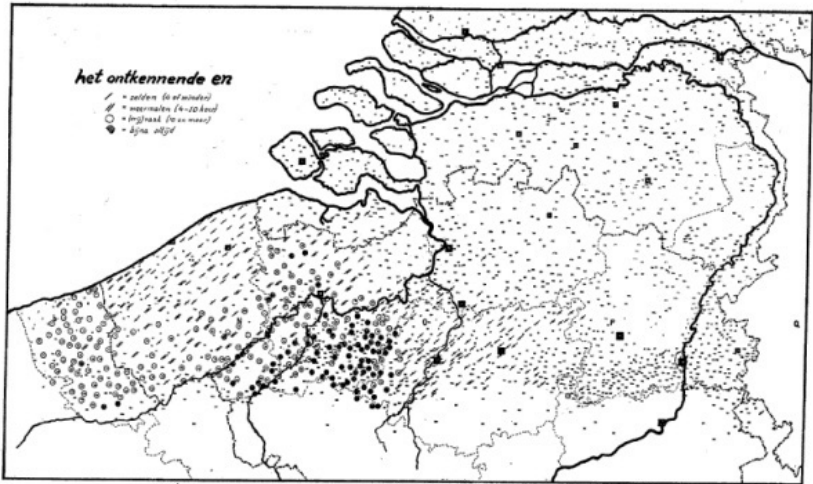
(1124p Lapscheure; Breitbarth & Haegeman 2014:70)

ACHTERGROND

- retentie van *en* als overblijfsel van de Jespersencyclus (vb. Beheydt 1998)
- goed beschreven voor zuidelijk-Nederlandse (vooral Vlaamse) dialecten (vb. Koelmans 1967, Haegeman 1995, van der Auwera & Neuckermans 2004, Neuckermans 2008)
- syntactische positie historisch direct vóór het verbogen werkwoord (vb. Haegeman 1995; 2002):

NEGATIEF EN

- Koelmans (1967): frequentie van ontkenkend *en* in RND-zinnen (max. 27): vooral in zuidelijk OV en zuidoostelijk WV



MINDER NEGATIEF EN

Breitbarth & Haegeman (2014; 2015): *en* is...

- niet helemaal optioneel (3)
- niet meer puur ontkennend, maar discoursmarkeerder met een procedurale betekenis, nl. dat de propositie **onverwacht** is in de context van een gegeven discoursachtergrond (4)

- (3) Waarom (**en**) ee=j da nu were nie gedaan? (Breitbarth & Haegeman 2014: 73)
- a. 'Waarom heb je dat niet nog een keer gedaan?' (*en* OK)
- b. 'Herinner mij even: Waarom heb je het niet gedaan?' (**en*)
- (4) a. k'een al overal gezocht in us en k'**en** vinden ze nievers.
(I124p Lapscheure; Breitbarth & Haegeman 2014:70)
- b. Ge woont al vijftien jaar in Gent, in g'**en** ken nog d'Universiteit nie?
(I241 Gent, Overdiep 1937:456)
- c. Ik meenden da(t) (i)k den duvel bij de ... bij de benen (h)oo, maar (i)k **en** (h)od en nog bij de staart niede.
(I256a Wichelen; Van Keymeulen 1975)

- Neuckermans (2008): aanwijzingen voor een verdere ontwikkeling (Neuckermans: grammaticalisatieproces): *en* in niet-negatieve zinnen, ook voor een niet-finiet werkwoord

- (5) a. Ze pakte eu portefeuille waar dase eu sleutelke in **en** doet
(0248 Halle; Neuckermans 2008: 176)
- b. Ij zal nog wel **en** komen. (0022, Merelbeke; Neuckermans 2008: 175)

- Neuckermans (2008: 175–6): “affirmatief” gebruik (vooral in hoofdzinnen) zeer zeldzaam, maar voetnoot p. 176:

*Dit betekent niet dat deze constructie onmogelijk is in de rest van Vlaanderen.
Volgens prof. Magda Devos komt deze constructie bijvoorbeeld ook voor in Gent.*

WAT WE NOG GRAAG ZOULDEN WETEN

- Geografische distributie in *spontaan gesproken* taal?
- Welke factoren beïnvloeden het optreden van *en*?
 - discourscontext
 - syntactische omgeving
 - lexicaal materiaal in de omgeving
 - eigenschappen van niet-negatief gebruik
- Missing link voor historische taalkunde

Nood aan nieuwe data-bronnen!

TIJD VOOR NIEUWE DATA?

- meest gebruikte data tot nu toe:
 - 1 geëliciteerd materiaal, bvb. SAND / DynaSAND (<http://www.meertens.knaw.nl/sand/>), RND (www.dialectzinnetjes.ugent.be)
 - 2 waardeoordelen van moedertaalsprekers, (toevallige) observaties
- probleem: bepaalde structuren zijn niet of zelden geattesteerd (vooral in de SAND) door opvragingen a.d.h.v. vragenlijsten, vb. $V > 2$

H116p	Torhout	veldwerker	[v=359] Met zo n weer je kun nie veel doen. [/v]	context	▶
		informant3	[a=n] Met zukke weer kun je nie veel doen. [/a] de drie informanten keuren deze zin af; nochtans komen er nogal wat inversieloze zinnen voor in de spontane spraak.	context	

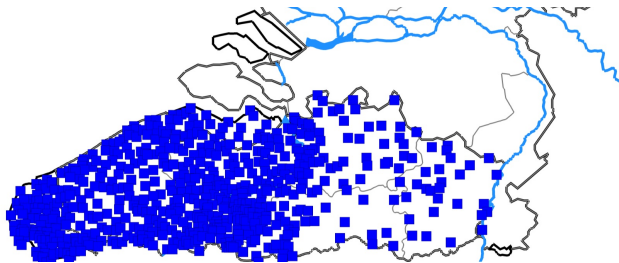
- bepaalde discourscontexten moeilijk opvraagbaar

(OUDE) NIEUWE DATA:

Het Gesproken Corpus van de (zuidelijk-)Nederlandse Dialecten
(GCND)

STEMMEN UIT HET VERLEDEN

- UGent: unieke verzameling van 783 bandopnames (ca. 700 uur) uit 550 plaatsen (Vanacker & De Schutter 1967), licentiaatsverhandelingen



STEMMEN UIT HET VERLEDEN

- profiel van de respondenten:
 - honkvaste sprekers (m/v) met geringe scholing, vaak analfabeten
 - geboren rond 1900 (oudste spreker 1871!)
 - bijzondere, verdwenen/verdwijnende beroepen
- inhoud:
 - levensverhalen, “oral history”
 - opkomst van technologische vernieuwingen: elektriciteit, auto's, fietsen...
 - ambachten, gewoontes, taalsituatie (vooral in Frans-Vlaanderen)

- beschikbaarheid en bruikbaarheid voor taalkundig onderzoek:
 - Sinds 2014 zijn de geluidsbestanden digitaal beschikbaar op www.dialectloket.be
 - probleem: gedigitaliseerd \neq doorzoekbaar



Dialectloket
een website over taalvariatie van de Universiteit Gent



STEMMEN UIT HET VERLEDEN

■ 318 transcripties uit de jaren 60 en 70; geen uniform protocol, kwaliteitsverschillen

van kamere. Waarda'k goeng, ze lag in een andere kamere, maar 't was in den doo' kamere (doodkamer). Ze lag ip sterven. 'k Kom erbi, 'k zie't, ze vraag no mi, 'eur kleed voor an te doen en een laken voor in te draaie. 'k Zegge morgenuchtend is ze dood, 'k goenk me' lakens, 't was ollemolle (allemaal) gedaan. 'k Zegge morgenuchtend is ze dood. Oo'k do 's nuchtens (Als ik daar 's ochtends) bi komme, ze leef' nog/ ja/ ze stonden erbi voor of te

L. D'r was zeker veel armoe dan in dien tijd (h)ier ?

S. Hojojo, d'r was (h)ier gee(n) werk (h)aast, newar, 't waren gelukkigen die bij nen boer mochten gas(n) werken ... voor nen boter(h)am (h)é, maar voor gee(n) geld, zolle. Da(t) wa(s) ne gelukkige mens die da(t) mocht doen. De mensen kwamen uit Frankrijk newar, dan ...

1) *Elbaldygen.* 29. dec. 1964

T: Jaa, en Gabriel doe je ons ne keer iets kunnen vertellen, over
over de vroegere oeders, tijd ^{aan} en over Elbaldygen van vroeger.

S: Hevel jong, ik zegge ik ik men echtes, Elbaldygen maaden.

De tijd dringt!

Doelstellingen:

- 1 Banden transcriberen – transcriptieprotocol
- 2 Transcripties annoteren
 - POS-tagging
 - Parsing
 - topic tags / trefwoorden
- 3 Online corpus via CLARIN infrastructuur + open source search software

Momenteel...

- 2018-19 financiering pilootproject FWO (€26.000; FW018/KAN/009)
- 2018-2021 postdoctoraal project FWO M. Farasyn over Frans-Vlaams (FWO 12P7919N)
- 2019-2020 subsidies door de provincies Zeeland (NL) en Oost-Vlaanderen
- Pijplijn ontwikkelen + begroten van tijd/budget nodig om hele collectie te transcriberen en annoteren

Hoe dialect transcriberen?

Twee lagen:

(i) dichter bij het dialect

- regelmatige klankveranderingen gestandaardiseerd
- functiewoorden en morfologie zoals in het dialect
- clitische elementen met # met elkaar verbonden

(ii) dichter bij de standaardtaal

- geen vertaling van dialectlexemen
- functiewoorden en morfologie gestandaardiseerd
- clitische elementen opgesplitst in afzonderlijke lexemen

- (6) maar a#k#ik mijn tanten horen klappen...
maar als ik ik mijn tanten hoor klappen...

TRANSCRIPTIEPROTOCOL

The screenshot displays the ELAN 5.0 software interface for a file named 'N72_leper.eaf'. The top menu bar includes 'File', 'Edit', 'Annotation', 'Tier', 'Type', 'Search', 'View', 'Options', 'Window', and 'Help'. Below the menu is a toolbar with tabs for 'Grid', 'Text', 'Subtitles', 'Lexicon', 'Comments', 'Recognizers', 'Metadata', and 'Controls' (which is currently selected).

The 'Controls' panel features three main sliders: 'Volume' (set to 100), 'N72_leper.wav' (set to 100), and 'Rate' (set to 100). There are also 'Mute' and 'Solo' checkboxes for the audio track.

Below the sliders, the current time is shown as '00:24:32.161' and the selected range as 'Selection: 00:24:29.483 - 00:24:31.907 2424'. A set of playback controls (play, stop, previous, next, etc.) is visible.

The main workspace shows a timeline with an audio waveform and two transcription tiers. The top tier, 'N72_leper....', shows a selected segment from 00:24:29.500 to 00:24:32.0. The bottom tier shows two transcription entries:

- N72_sp1 [702]**: maar a#k#ik mijn tanten horen klappen... wel \
- N72_sp1_ndl [702]**: maar als ik ik mijn tanten hoor klappen... wel \

TRANSCRIPTIEPROTOCOL



- nauwkeurige controle van 1000 tags in de transcripties van 9 opnames (één per dialect of overgangsbegied)

	TreeTagger	LeTS	FROG
Oudenburg (H24)	94.2	92.9	95.2
Maldegem (I154)	94.7	94.4	97.3
Westdorpe (I166)	90.9	93.0	98.8
Sint-Niklaas (I175)	96.3	91.4	94.6
Gent (I241)	91.8	88.4	92.9
Ieper (N72)	90.4	82.5	92.6
Hardifort (N94)	90.1	88.5	93.8
Sint-Joris-Weert (p130)	91.8	90.5	89.3
Uikhoven (Q013)	95.9	92.8	95.9
Total	92.9	90.5	94.5

- FROG herkent zelfs dialectwoorden (*gewrocht, ne < een, bé, ...*); doet al chunking / (dependentie-)parsing (grammaticale functies)

- 45 opnames getranscribeerd
- Transcripties voorzien van POS-tags, chunking en parsing informatie met FROG (Van den Bosch et al. 2007)
- Netwerk van vrijwilligers voor manuele correctie

WORK IN PROGRESS

- manuele correctie van de transcripties (vrijwilligers) en de POS-tags in de 45 opnames van het pilootproject
- output van FROG moet nog van Folia-XML (<https://proycon.github.io/fofia/>) naar ELAN-XML overgezet worden om POS-tags aan tijdscores te koppelen

ONS TO-DO-LIJSTJE

- Afsluiting pilootproject
- Infrastructuur-aanvraag 2019
- Eventueel onderzoeksprojecten met infrastructuur-component

WE HAVE A DREAM

- 700 uur cultureel erfgoed ontsloten
 - voor syntactisch onderzoek
 - voor volgende generaties
- Eventueel uitbreiding naar het Noorden?

CASESTUDY: (NIET-)NEGATIEF EN

CASE: (NIET-)NEGATIEF EN

Neuckermans (2008)¹:

- partikel **en** in ontkennende zinnen:
 - in bijzinnen frequenter dan in hoofdzinnen
 - normaal gezien direct voor finiet werkwoord
- 23 plaatsen: *en* voor **niet-finiet** werkwoord i.p.v. *Vfin*
- daarnaast: niet (echt) negatief gebruik:
 - 1 restrictief (Neuckermans: 19 plaatsen, vooral WV/OV, 7 FV) en expletief (9 plaatsen, vooral WV/OV)
 - 2 in niet-negatieve bijzinnen (7 plaatsen in Antwerpen & Brabant)
 - 3 in niet-negatieve hoofdzinnen (slechts in 0022 Merelbeke)



CASE: (NIET-)NEGATIEF EN

GCND (dus ver):

- partikel **en** in (niet alleen) ontkennende zinnen:
in hoofdzinnen frequenter dan in bijzinnen!

	hoofdzin	bijzin
ontkennend	178	46
restrictief	15	1
niet-ontkennend	18	2

- aard van de data? (gesproken \rightsquigarrow meer paratactisch?)
 - tekenen van fossilisering? 27 x *ik en weet/wist...*; 24 x *we en hebben / hadden ...*
- (7) allez waar dat hij dan dan verdwenen is. waarschijnlijk terechtgekomen in de andere armee ewaar ...
ik en weet het niet (1227 Vinkt)

CASE: (NIET-)NEGATIEF EN

- *en* in het GCND (tot dusver): geografische distributie (frequentie per 10K woorden) bevestigt Koelmans' bevindingen (op basis van RND-data, dus niet spontaan gesproken taal) – vooral geconcentreerd in overgangsgedebied WV-OV (en FV)



CASE: (NIET-)NEGATIEF EN

- Neuckermans (2008: 168–70): ‘expletief’ (=negatief-polair) gebruik van *en* “zeldzaam”, “[komt] sporadisch voor[...] na een comparatief, na het voegwoord *voordat* en in conditionele zinnen”
- GCND: indicatie dat er meer en andere negatief-polaire contexten mogelijk zijn; probleem met geëliciteerde data bij Neuckermans (RND en SAND)?

- (8) a. *ja/neen-vraag*
ja maar **en** bestaat dat nog? dat weet ik niet. (080 Waregem)
- b. *ja/neen-vraag / (niet?) meer:*
van aan de hoek van het Oostakkerstraatje is dat misschien overgenomen die apothekerij? Dat ze zij dat meer **en** doet? (1241 Gent)
- c. restrictie van \forall -kwantor:
maar jij bent een grote dommerik en al die die dat **en** heeft zijn dommeriken (1131 Waterland-Oudeman)

CASE: (NIET-)NEGATIEF EN

Meer nog niet in de literatuur besproken gebruik van *en*:

1 **indirecte ontkenning** (directe ontkenning in de hoofdzin licenseert *en* in niet-negatieve bijzin)

- (9) als het niet schoon is ... de mensen gaan niet binnen [...] als het zo maar naar de winkel of ... of naar de patisserie **en** is of gelijk wat ... je gaat **niet** binnen (080 Waregem)

2 "indirect-restrictief"?

- (10) en ze danste maar het waren **maar juist** haar voeten die van de grond **en** gingen (N42 Pittem)

3 ***en* in hoofdzin + ingebedde vraag** – Middelnederlands!

- (11) a. ik **en** weet **of** dat nu nog veel meer gedaan werd (0265 Ronse)
b. Ic **en** weet **of** si overeen die name mingen van desen tween (Postma 2002: 50)

CASE: (NIET-)NEGATIEF EN

- niet-negatief *en* in hoofdzinnen (incl. voor niet-finiete werkwoorden! – (12c))
waarschijnlijk verder verspreid dan in Neuckermans' data:

- (12) a. maar wil jij jonge gasten we **en** hebben al een keer jaren gehad dat we een paar jonge gasten hebben hé dus studenten van de universiteit hé (I158 Eeklo)
- b. maar ja het bij het leger. Maar we **en** hebben eerst eerst in dingen gelegen in uh ... in Casteau (0152 Ninove)
- c. met zijn beste kleren aan ... je had DIEN een keer moeten **en** zien (N42 Pittem)

- opmerkelijk: *maar* of contrastieve intonatie in (12a-12c)

CASE: (NIET-)NEGATIEF EN

- bevestiging van Breitbarth & Haegeman (2014; 2015): *en* als discoursartikel die propositie uit de voorgaande context als onverwacht markeert – (13)?

- (13) a. ja ik **en** was al gelijk nog op de wereld niet als de eerste auto hier gekomen is want hij is hier gepasseerd de eerste september achtienhonderdriënnegentig (I154 Maldegem)
- b. ewaar en die mensen trapten dat natuurlijk te voet af want uh fietsen **en** bestonden er nog niet (I227 Vinkt)

- meer onderzoek nodig; sporadische voorbeelden waar zo'n contrast moeilijker te zien is:

- (14) Nou nen boerenzoon of ne rijke mens die nie(t) geren soldaat **en** wierd en nen arme mens die verkocht (h)em dan. (I256a Wichelen; Van Keymeulen 1975)

- andere losse eindjes: niet-negatief *en* in context *we (en) hebben*:

- (15) a. zo we **en** hadden dan een keer een zeug en uh ik zeg tegen onze Romain ik zeg die zeug laat haar zoog niet komen (I156 Sint-Laureins)
- b. allez we **en** hebben we dat dan veel gehad hé (I156 Sint-Laureins)

BESLUIT

- nog nodig: kwantitatief en kwalitatief detailonderzoek naar syntagmatische en paradigmatische correlaten van het gebruik van *en* in ontkennende en niet-ontkennende contexten
- voordeel GCND: spontaan gesproken taal – geen beïnvloeding door een vragenlijst, natuurlijke discoursomgeving
- nadeel: geen experimentele condities aftoetsbaar, toeval
- nog veel meer data nodig (en subsidies)
- heel diverse onderzoeksvragen mogelijk

Bedankt!



Veronique De Tier
Melissa Farasyn
Lien Hellebaut
Jacques Van Keymeulen
alle transcribenten
... en de vrijwilligers!

- Aelbrecht, Lobke. 2008. *Niemand nie* in Vlaamse spreektaal. *TABU* 36: 89–102.
- Behedyt, G. 1998. Het gebruik en de vorm van de negatie in het zuidelijke Nederlands in een diachronisch perspectief (15e–20e eeuw). Licentiaatsverhandeling Universiteit Leuven.
- Bosch, A. van den, Busser, G.J., Daelemans, W., and Canisius, S. (2007). An efficient memory-based morphosyntactic tagger and parser for Dutch, In F. van Eynde, P. Dirix, I. Schuurman, and V. Vandeghinste (Eds.), *Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting*, Leuven, Belgium, pp. 99–114
- Breitbarth, A. & L. Haegeman. 2014. The distribution of preverbal *en* in (West) Flemish: syntactic and interpretive properties. *Lingua* 147: 69–86.
- Breitbarth, A. & L. Haegeman. 2015. *'En' en is niet wat we dachten*: a Flemish discourse particle. *MIT Working Papers in Linguistics* 75: 85–102.
- Breitbarth, A., M. Farasyn & L. Haegeman. 2018. "t En is niet spijtig": De distributie van (niet echt) ontkennend en in het Wichels. In T. Colleman, J. De Caluwe, V. De Tier, A.-S. Ghyselen, L. Triest, R. Vandenberghe & U. Vogl (eds.), *Woorden om te bewaren – Huldeboek voor Jacques Van Keymeulen*, 11–24. Gent.
- DynaSAND (2006) = S. Barbiers, H.J. Bennis, G. de Vogelaer, M. Devos & M.H. van der Ham. *Dynamische Syntactische Atlas van de Nederlandse Dialecten (DynaSAND)*. Amsterdam: Meertens Instituut. <http://www.meertens.knaw.nl/sand/>.
- Ghyselen, A.S. & J. Van Keymeulen 2014. Dialectcompetentie en functionaliteit van het dialect in Vlaanderen anno 2013. *Tijdschrift voor Nederlandse Taal- en Letterkunde* 130, 117–139.
- Ghyselen, A.S. & J. Van Keymeulen. 2016. Implicational scales in colloquial Belgian Dutch. *Dialectologia et Geolinguistica* 24(1): 62–82.
- Haegeman, L. 1992. *Theory and description in generative syntax. A case study in West Flemish*. Cambridge University Press, Cambridge.
- Koelmans, L. 1967. Over de verbreiding van het ontkennende en. *De Nieuwe Taalgids* 60: 12–18.
- Lybaert, C., B. De Clerck, Saelens, J. & L. De Cuyper. 2017. A corpus-based analysis of V2 variation in West Flemish and French Flemish dialects. *Journal of Germanic Linguistics*.
- Neuckermans, A. 2008. Negatie in de Vlaamse dialecten volgens de gegevens van de Syntactische Atlas van de Nederlandse Dialecten (SAND). PhD thesis, Ghent University.
- Postma, G.J. 2002. De enkelvoudige clitische negatie in het Middelnederlands en de Jespersen cyclus. *Nederlandse Taalkunde* 7: 44–82.
- Vanacker, V.F. (1967). Syntaktische Daten aus Französisch-Flämischen Tonbandaufnahmen. In: L.E. Schmitt (red.), *Verhandlungen des 2. Internationalen Dialektologenkongresses*. Sonderband *Zeitschrift für Mundartforschung* 3–4, 844–855.
- Vanacker, V.F. & G. De Schutter. 1967. Zuidnederlandse dialecten op de band. *Taal en Tongval* 19: 35–51.
- Van Keymeulen, J. 1975. Enkele syntactische kenmerken van het Wichels dialect. Licentiaatsverhandeling Universiteit Gent.