

-
-
-

Wat hebben de lexicograaf en de foneticus elkaar te vertellen?

Truus Kruyt

**aangepaste versie t.b.v. werkcollege Corpusonderzoek
Universiteit Münster 2007/2008**

-
-
-

Overzicht van de presentatie

Relaties tussen fonetiek en lexicografie : vroeger

fonetiek in de woordenboeken

Relaties tussen fonetiek en lexicografie : nu

Corpus Gesproken Nederlands (CGN)

Relaties tussen fonetiek en lexicografie : straks

het CGN en de Geïntegreerde Taalbank

-
-
-

Fonetiek in woordenboeken: uitspraak (1)

19 typen woordenboeken (1) *(naar Geeraerts & Janssens 1982)*

Verklarende handwdb	klemtoon
Spreekwoorden-wdb	-
Synoniemen-wdb	-
Vreemde woorden wdb	-
Spelling-wdb	-
Uitspraak-wdb	~ IPA
Afkortingen-wdb	-
Etymologische en <i>naamkundige</i> wdb	-
Historisch-taalkundige wdb	-
Wdb op auteurs of <u>werken</u>	lijst van rijmwoorden

-
-
-

Fonetiek in woordenboeken: uitspraak (2)

19 typen woordenboeken (2) *(naar Geeraerts & Janssens 1982)*

Nieuwe woorden **wdb**

Wdb Zuid-Ned./Surinaams Ned.

Dialect-wdb

Terminologische wdb

Wdb van slang en geheimtaal

Kinder-**wdb**

Rijm-wdb

Puzzel-**wdb**

Retrograde en frequentie-**wdb**

-

zelden klemtoon/uitspraak
vroeger vaak geen uitspraak

-

-

uitspraak (in spellingvorm)
soms klemtoon-geordend

-

-

-
-
-

Fonetiek in woordenboeken: gesproken-taalgebruik

Woordenboeken o.b.v. gesproken taal (min of meer) :

Historisch-taalkundige wdb (WNT), Nieuwe woorden *wdb*,
Wdb Zuid-Ned. /Surinaams Ned., Dialect-*wdb*,
Wdb van slang en geheimtaal, Retrograde en frequentie-*wdb*

Dus niet:

Verklarende hand*wdb*, Spreekwoorden-*wdb*, Synoniemen-*wdb*,
 Vreemde woorden *wdb*, Spelling-*wdb*, Uitspraak-*wdb*,
 Afkortingen- *wdb*, Rijm-*wdb*, Puzzel-*wdb*,
 Terminologische *wdb*, *Wdb* op auteurs of werken,
 Kinder-*wdb*, Etymologische en naamkundige *wdb*

-
-
-

Van vroeger naar nu

Technologische ontwikkelingen : er kan meer

Relatief groot corpus van gesproken taal haalbaar:

British National Corpus

10 miljoen woorden

Corpus Gesproken Nederlands

10 miljoen woorden

Multimedia-applicaties mogelijk : wdb-tekst met geluid

Internationale ontwikkelingen : meer raakvlakken

EU: ontwikkeling 'Languages Resources' (**ELREC-conferenties**)

vormen van standaardisatie, bep. tools

lexiconbouwers: spraak-/taaltechnologen + lexicografen

-
-
-

Wat doet een lexicograaf ?

een woordenboek maken

Beschrijving van eigenschappen van de woordenschat

welke eigenschappen?

afhankelijk van type wdb

ingang:

trefwoord + 'vaste' verbindingen

door analyse van het 'woord'gebruik

o.a. in elektronische corpora

(m.u.v. Engels, in commerciële lexicografie relatief recent)

'Vanouds': corpora van geschreven taal (10-tallen miljoenen woorden)

Sinds 'kort': corpora van gesproken taal ('slechts' 10 miljoen woorden)

-
-
-

Corpus Gesproken Nederlands CGN: wensen v.e. wdb-maker

Wat denkt de lexicograaf nodig te hebben?

Getranscribeerd CGN

- * niet genormaliseerd naar geschreven taal
- * informatie over klemtoon
- * informatie over accent (*prominentie*)

Flexibele retrieval op CGN

- * woordvormen en trefwoorden (*met POS*) in context
- * subcorpusselectie
- * meta-informatie
- * syntactische patronen (**maar: in hoeverre haalbaar?**)
- ? link met gesproken taaluitingen
- * link met andere corpora: *geschreven, dialect, internet ...*

-
-
-

Corpus Gesproken Nederlands CGN: een 1^{ste} verkenning

Vraag: wat leert het CGN ?

- * geschreven taal *versus* gesproken taal
- * Vlaams *versus* Nederlands

Analyse van transcriptiefiles op cd-rom CGN:

10 bestanden Nederlands + 10 bestanden Vlaams
willekeurige keuze *maar* :

grotere bestanden (ca. 30 Kb) (N.B. >1 spreker per bestand)
streven naar diversiteit (veel leraren: klas / over lezen)

Hulp Vlaamse collega's + wdb

-
-
-

Corpus Gesproken Nederlands CGN: een 1^{ste} verkenning

Welke gespreksonderwerpen?

Nederlandse sprekers:

gesprek (*van alles*)

diner-gesprek (*van alles*)

telefoongesprek (*van alles*)

telefoongesprek (*van alles*)

lezing waterhuishouding

toespraak Gelders Landschap

lezing ruimtelijke ordening

gesprek ziekten, behandeling

interview leraar Nederlands

lezing woningbouw

Vlaamse sprekers:

toespraak hoogleraar

leraar/klas: buitenl. leerl.

leraar over lezen

leraar/klas: normen

interview politici

interv. boerinnen: dioxinecrisis

interview over 20^{ste} eeuw

interview kok over koken

interview zanger over cd

interview over Chinees boek

-
-
-

Corpus Gesproken Nederlands CGN: 1^{ste} indrukken (1)

Gesproken taal vs. Geschreven taal (1)

Algemene aspecten:

*gesproken-taalkenmerken (*zie hierna*) → *andere frequenties*

*afgebroken woorden → *wat kun je daarmee?*

*weinig verband pauzes / zinsgrenzen → *transcriptie slecht leesbaar*

**toespraak*: interactie tekst / beeld → *interpretatie tekst soms lastig*

Planning van het spreken:

*zeer veel woordherhalingen (*ook >1 woord*)

bijv. nou als iemand net dood is en ze en ze en en uh

*veel substantieven met 2 of meer (verschillende) lidwoorden

*expressies als: *uh, uhm, zeg maar, ik bedoel, weet ik wat, hoe moet ik het zeggen, ik weet niet precies, zal ik maar zeggen*

-
-
-
-
-
-
-
-

-
-
-

Corpus Gesproken Nederlands CGN: 1^{ste} indrukken (2)

Gesproken taal vs. Geschreven taal (2)

Interactie tussen sprekers / spreker en publiek:

spreker: toch?, vind je niet?, sorry, iets anders, stel u voor, nu zou je kunnen zeggen, je kunt het niet goed zien maar, da's leuk man zo'n..

luisteraar: ja, nee, mm, hum, goh, jawel, ja joh, nee oké, natuurlijk, oh, ah, he, aha, misschien, nou nou, inderdaad, uiteraard, sjonge jonge, zou best kunnen ja, volgens mij niet, dat niet, dat zal wel ja, 't zal toch wel?, ja hoezo?, ja precies, ja dat klopt, dacht 't wel, vind ik ook wel, nou ik weet niet, meen je dat nou?, leg eens uit, dat bedoel ik ja, begrijp dat niet, is prima, kan ook

-
-
-

Corpus Gesproken Nederlands CGN: 1^{ste} indrukken (2)

Gesproken taal vs. Geschreven taal (3)

Woordenschat: *wsch. eerder in gesproken taal dan in schrijftaal*

rondbanjeren, hommeles, tof, geinig, uitkienen, klussen, nep, snappen (*begrijpen*), plenty, hartstikke, heisa, platweg

tjee, voilà, verdorie, god, hup, nou, wel, eerlijk gezegd, denk ik, maar ja, zou ik zeggen, moet ik zeggen, gewoon, best wel, is wel makkelijk, waarom ... dan

tuurlijk, mekaar, effe, effetjes, ie, d'r (*er*), d'rop *etc.*, hun (*hen*), wou je, kan je, je kan, 'r zus, d'r tegenaan

-
-
-

Corpus Gesproken Nederlands CGN: 1^{ste} indrukken (3)

Gesproken taal vs. Geschreven taal (4)

Syntaxis:

- * niet lopende zinnen, onafgemaakte zinnen
- * weglating eerste woord van de uiting
bijv. is gratis, hebben ze mooi gedaan, heeft Enst gelijk in
- * overbodig pronomen
bijv. de markt die heeft zich.. , Fenicië dat was gelegen..
- * tangconstructies
*bijv. het enige is ik kan op die vraag kan ik geen antwoord
geven op dit moment*
- * "zeggen van" + directe rede
bijv. als jij zegt van 'k wil hier de Gelderlander

-
-
-

Corpus Gesproken Nederlands CGN: 1^{ste} indrukken (4)

Vlaams vs. Nederlands (1) *deze kleur = 'zou je niet opschrijven'*

Lexicaal:

allee, awel, bah (ja), ocharmen, amai, (chapeau) zonne,
oelala, vermits, nochtans, plots, gans, plezant, aanpalende
(dorpen), aanleunende (provincies), debiet, miserie, valies,
omzendbrief, omhaling, kepie, geraken, bekomen, buitengaan

Met andere betekenis:

zetel (bank), aanduiden (voordragen), terug (opnieuw),
juist (correct), gelijk welke (om het even), nog straffer gaan (kras,
sterk), titularis (klasseleraar), licentie (~doctoraalopleiding)

-
-
-

Corpus Gesproken Nederlands CGN: 1^{ste} indrukken (5)

Vlaams vs. Nederlands (2) *deze kleur* = 'zou je niet opschrijven'

Uitdrukkingen:

katje mie katje were, op de sukkel zijn, aan hetzelfde zeel trekken,
uit de brand slepen, met de vlag gaan zwaaien, schrik hebben,
rond de pot draaien, met iemands voeten spelen, in gang schieten,
(familie) op de buiten, zich uit de slag trekken

Morfologisch:

verkleinwoord: ietske, spelleke, zetelke, vingerkes, afbeeldingske

meervoud: om negen uren ; artikels?

ne keer (een keer), *den* indruk van, *uwen* boterham

zo een (zo'n), *vandemorgen*

-
-
-
-
-
-
-
-

-
-
-

Corpus Gesproken Nederlands CGN: 1^{ste} indrukken (6)

Vlaams vs. Nederlands (3) *deze kleur = 'zou je niet opschrijven'*

Varianten op het Nederlands:

duizendvijfhonderd, een uur en half (*anderhalf*), met vier (*z'n vieren*), bij manier van spreken, laat ons (werken) (*laten we ...*)

onder de vakantie (*in*), werken aan de radio (*bij*)

van god en van alleman verlaten, *een man en een paardenkop*

het gedoe daarrond, weeral (*alweer*), maar alleen (*alleen maar*),
maar altijd (*altijd maar*)

dat is meer als een geluk ja, evenveel dan wij

daarjuist (*zojuist*), als + *verl. tijd* (*toen*)

-
-
-

Corpus Gesproken Nederlands CGN: 1^{ste} indrukken (7)

Vlaams vs. Nederlands (4) *deze kleur = 'zou je niet opschrijven'*

Syntaxis: constructie met van + infinitief-constructie

bijv. een tweede punt is van nieuwsgierig te blijven
mogelijkheden die wel vereisen van risico's te nemen
mijn vader die zwoer van dat nooit te doen

Syntaxis: constructie met dat

bijv. de manier waarop dat die mensen denken
waardoor dat je de tijd krijgt
materies waarvan dat je zelf ook niks kent
ze snapt niet waarom dat wij lachen
terwijl dat er eigenlijk helemaal niks aan was

-
-
-

Corpus Gesproken Nederlands CGN: 1^{ste} indrukken (8)

Vlaams vs. Nederlands (5) *deze kleur = 'zou je niet opschrijven'*

Syntaxis: hulpww ~ gaan voor zullen

bijv. ik ga niet ingaan op
ik ga nog verder gaan
je gaat een boete moeten betalen

Syntaxis: hulpww ~ positie van het voltooid deelwoord

bijv. als hun agent uit het systeem moet genomen worden
wij dachten dat ie ook ging gedepORTEERD of wat ook worden
dat daar wel over kan gesproken worden

-
-
-

Corpus Gesproken Nederlands CGN: 1^{ste} indrukken (9)

Vlaams vs. Nederlands (6) *deze kleur = 'zou je niet opschrijven'*

Syntaxis: hulpww ~ scheidbare woorden

bijv. hij zegt dat daar iets van moest terechtkomen
u hebt dat laten voorbijgaan
als ik straks d'r wil inzitten dan moet u mij aanduiden
een initiatief ... waar de CVP niet mag aan deelnemen

Syntaxis: hulpww ~ overig

bijv. hoe is dat dan plots zo kunnen komen
die deze agenda kunnen stap voor stap uitwerken
ik zou niet graag hem zien Nederlands geven hier
daar zijn bomen aangeplant geworden,
in de vorige legislatuur is er nooit gezegd geweest door ...

-
-
-

Corpus Gesproken Nederlands CGN: prosodie

door middel van prosodische annotatie:

Klemtoonverschillen tussen Vlaams en Nederlands

Breda, azalea, rijksdaalder, achttien (*Vlaams*)

Klemtoonverschuivingen in (gesproken) Nederlands?

Radio: pianist ; leenwoorden

Verband tussen zinsaccent (prominentie) en betekenis?

Kruiyt (1985), p. 61-62: tentatief verband tussen
betekenis van ww ~ accentuering van ww in de zin

-
-
-

Corpus Gesproken Nederlands CGN: straks (1)

het CGN en de Geïntegreerde Taalbank (GTB) (1)

Wat wordt de Geïntegreerde Taalbank?

Een databank met Nederlands van 8ste - 21ste eeuw

Doel?

Onderzoeksinstrument Nederlandse taal (en cultuur)

Samenstelling : drie componenten

woordenboeken-component

tekstcorpus-component

lexica-component

-
-
-

Corpus Gesproken Nederlands CGN: straks (2)

het CGN en de Geïntegreerde Taalbank (GTB) (2)

Relatie met CGN?

Taalkundige verrijking: woordsoort en lemma (*syntaxis?*)

Links van GTB naar CGN en vv. in retrievalsysteem

Waarom?

Optimaal onderzoeksinstrument Nederlandse taal:

- * verschillen tussen geschreven en gesproken Nederlands
- * verschillen tussen Vlaams en Nederlands 'Nederlands'
- * signaleren van taal- (*en norm-*) verandering o.b.v. 'data'
- * verband prosodie (e.a. fonetische kenmerken) en betekenis?

Wanneer?

?????