

Over corpora, intuities, en idiomen

Tanja Temmerman & Jeroen van Craenenbroeck
Universit  Saint-Louis Bruxelles/CRISP/KU Leuven/KNAW Meertens Instituut

Samenvatting

This paper presents two case studies regarding the syntactic (in)flexibility of idiomatic expressions. We compare the conclusions reached on the basis of corpus studies with native speaker judgements available in the Database of Dutch Dialect Idioms (DaDDI). Based on these two case studies, we conclude that both types of data collection are highly complementary and can mutually reinforce one another. This is in line with the recent rapprochement between the two approaches to linguistic data collection.

1 Inleiding: corpora en intuities

E n van de klassieke datagerelateerde tegenstellingen in de moderne taalkunde is die tussen intuities of oordelen van moedertaalsprekers enerzijds en corpusgegevens van natuurlijke taal anderzijds. Het onderscheid tussen deze twee types van gegevens werd op scherp gesteld in Chomsky's (1964) klassieke tweedeling tussen *competence*, de impliciete kennis die een spreker heeft over haar taal, en *performance*, de manier waarop die kennis gebruikt wordt in concrete taaluitingen. De kritiekpunten voor beide benaderingen zijn genoegzaam bekend (zie bijv. Baldwin et al. (2005:50–52) voor een algemeen overzicht). Corpora zijn per definitie beperkt in de hoeveelheid data die ze bevatten (zeker wanneer het kleine, jonge, of niet-standaardtalen betreft, zie Leivada et al. (2019)), ze kunnen de limieten van variatie in een taal niet aantonen (i.e. ze verschaffen geen negatieve evidentie, zie Henry (2005)), en ze bevatten vaak fouten, versprekingen, etc. Intuities van moedertaalsprekers daarentegen zijn vaak weinig homogeen of categoriaal, worden mogelijk be nvloed door een gebrek aan kennis van meta-taal, zijn soms gebaseerd op de oordelen van een beperkt aantal sprekers, en zijn erg gevoelig voor impliciete *bias* (zie Skilton (2017) en Baggio et al. (2012) voor voorbeelden, en in het bijzonder Sch tze (2016) voor een gedetailleerde bespreking). Hoewel Chomsky (1965) reeds stelde dat “de authentieke gegevens van taalkundige *performance* heel wat bewijs zullen leveren voor het bepalen van de juistheid van hypothesen over onderliggende linguistische structuur, samen met introspectieve oordelen” (p.18, vertaling TT&JvC), leken beide benaderingen lange tijd mijlenver van elkaar te staan. Recent zien we echter een toenadering tussen de twee benaderingen, en groeit het besef dat het bij deze twee soorten van taalgegevens niet om een of-of-, maar om een en-en-verhaal gaat (zie bijv. Leivada et al. 2019:5). Corpusgegevens en oordelen van moedertaalsprekers zijn complementair aan elkaar: de eerste kunnen contexten, gebruikswijzen, en voorbeelden suggereren die ontsnappen aan de verbeeldingskracht van de taalkundige onderzoeker, en bieden frequentiegegevens die genuanceerder en fijnmaziger zijn dan de categoriale zwart-of-wit-oordelen die soms van moedertaalsprekers gevraagd worden. Anderzijds kunnen die intuities wel licht werpen op die fenomenen die (zo goed als) afwezig blijven in het corpus: gaat het om constructies die door moedertaalsprekers afgekeurd worden, of hebben andere factoren geleid tot hun niet-voorkomen in het corpus? Deze *entente* tussen de twee ooit erg vijandig tegenover elkaar staande benaderingen werd nog in de hand gewerkt door het feit dat het verzamelen van oordelen van moedertaalsprekers tegenwoordig ook steeds vaker op een grootschalige en gecontroleerde manier gebeurt (zie bijv. Sprouse and Almeida (2017)).

Dit artikel levert een bijdrage aan de zopas geschetste trend, en wel op het gebied van idiomen. Idiomen zijn uitdrukkingen met een niet-compositionele betekenis: hun betekenis is niet zonder meer voorspelbaar of afleidbaar op basis van (de combinatie van) de letterlijke betekenis van hun samenstellende delen. Enkele voorbeelden zijn *de bloemetjes buiten zetten*, *de kroon spannen*, of *het loodje leggen*. Dit soort constructies vormen het ideale terrein om de interactie tussen corpusgegevens en intuities van moedertaalsprekers te bespreken. Tot voor kort baseerde men zich in het idiomonderzoek immers voornamelijk op een kleine verzameling van steeds terugkerende idiomen (*poster children*), veelal uit het Engels, informeel beoordeeld door enkele sprekers (vaak de onderzoekers zelf). Hier

is recent verandering in gekomen, met meer systematische en uitgebreide corpusstudies (zie, i.a., Fellbaum 2006, 2007, 2014, 2015, 2019; Grégoire 2009; Horvath and Siloni 2009, 2014, 2015, 2016; Sailer 2015; Stone 2017; het PARSEME project: <https://typo.uni-konstanz.de/parseme/>). Zoals kon verwacht worden, hebben deze corpusstudies verschillende van de als vanzelfsprekend beschouwde empirische generalisaties uit de voorgaande literatuur op losse schroeven gezet, door te laten zien dat de corpora meer variatie bevatten dan aanvankelijk door de onderzoekers aangenomen (op basis van hun oordelen). Deze bijdrage laat echter zien dat het corpusonderzoek op haar beurt ook weer vragen oproept, vragen die succesvol beantwoord kunnen worden door een (grootschalige en gecontroleerde) verzameling van oordelen van moedertaalsprekers.

Het artikel is als volgt gestructureerd. In de volgende sectie presenteren we de achtergrond van ons onderzoek, met name het project *The Syntax of Idioms* en de daaruit resulterende online databank. Sectie 3 bevat twee casussen die laten zien dat er een complementariteit bestaat tussen het bestaande corpusonderzoek naar idiomen enerzijds en onze verzameling van intuïties van moedertaalsprekers anderzijds. De twee casussen die in sectie 3 behandeld worden, betreffen beide de (on)mogelijkheid van syntactische manipulaties van idiomen. In subsectie 3.1 wordt nagegaan of er al dan niet verschillende types idiomen (op basis van hun syntactische flexibiliteit) onderscheiden kunnen worden. In subsectie 3.2 worden mogelijke correlaties tussen idioomeigenschappen besproken. Sectie 4 is het besluit van deze bijdrage.

2 Het project *The Syntax of Idioms* en de databank *DaDDI*

Zoals hierboven aangegeven is het empirische uitgangspunt van deze bijdrage een uitgebreid onderzoek naar idiomatische uitdrukkingen. In ons onderzoek gaan wij echter niet uit van Standaardnederlandse idiomen, maar van idiomatische uitdrukkingen in dialecten van het Nederlands. In (1) worden enkele voorbeelden van dialectale idiomen gegeven:

- | | | | |
|-----|----|---|-------------|
| (1) | a. | <i>'t gras kennen heuren wassen</i>
het gras kunnen horen groeien
'in de hoogste mate eigenwijs zijn' | Hogeland |
| | b. | <i>d'n hak schudde</i>
de.sg.nn hiel schudden
'doodgaan' | Groesbeek |
| | c. | <i>in zijn gat gebeten gehad hebben</i>
in zijn achterste gebeten gehad hebben
'nijdig, wrevelig, getergd zijn' | Tessenderlo |

De dialectale data besproken in deze bijdrage werden verzameld in het kader van het onderzoeksproject *The Syntax of Idioms* (De zinsbouw van idiomen). Dat project liep van 2012 tot 2017 en werd uitgevoerd door taalkundigen van de Universiteit Utrecht (UiL-OTS) en de KU Leuven (CRISSP). Het werd gesubsidieerd door de nationale wetenschapsorganisaties van Nederland (NWO) en Vlaanderen (FWO). Het belangrijkste doel van het project was om zowel de interne als de externe syntactische structuur van idiomatische uitdrukkingen in kaart te brengen, dat wil zeggen: zowel de opbouw van idiomen als de manier waarop ze met de rest van de zin interageren. Het project richtte zich dan ook niet op inhoudswoorden, maar op het functionele (grammaticale) materiaal in idiomen.

Er werden twee soorten data verzameld binnen *The Syntax of Idioms*. Eerst werden meer dan 7.000 idiomen geëxcerpeerd uit woordenboeken, grammatica's, en idioticons van dertien dialecten gesproken in Nederland en Vlaanderen. De keuze van de dialecten gebeurde voornamelijk op basis van hun geografische spreiding: de dertien dialecten zijn afkomstig uit de verschillende Nederlandse dialectgebieden, zoals te zien is in Figuur 1. Bovendien werden sommige dialecten geselecteerd omdat ze bepaalde van de standaardtaal afwijkende syntactische eigenschappen bezitten (zoals bijvoorbeeld lidwoordomissie in Hogeland of hulpwerkwoordsverdubbeling in Tessenderlo—een voorbeeld van dit laatste is trouwens te zien in (1c)).

De tweede stap van de dataverzameling bestond uit veldwerk: gegevens werden verzameld op basis van mondelinge formele interviews met moedertaalsprekers van dialecten van het Nederlands. Met het veldwerk wilden de onderzoekers systematisch nagaan enerzijds in hoeverre idiomen afhankelijk zijn van het functioneel materiaal dat in het woordenboeklemma wordt vermeld, en ten tweede in hoeverre idiomen morfosyntactische flexibiliteit toestaan (passivisatie, topicalisatie, modificatie, etc.) Met dit doel werd een aantal idiomen geselecteerd waarop verschillende syntactische manipulaties werden



Figuur 1: De geselecteerde dialecten binnen het project *The Syntax of Idioms*. De indeling in dialectgebieden is overgenomen uit Daan and Blok (1969).

toegepast. Op deze manier zijn vragenlijsten met testzinnen samengesteld voor tien dialecten. Deze vragenlijsten werden voorgelegd aan dialectsprekers die hun oordelen hebben gegeven over de zinnen. Voor iedere testzin hebben zij aangegeven of de idiomatiche betekenis onder de syntactische manipulatie behouden bleef.

Bekijken we bijvoorbeeld het volgende idioom uit Aalst:

- (2) *den heif opgeiten hemmen*
 de zuurdesem opgegeten hebben
 'de zondebok zijn' Aalst

Dit idioom werd in het perfectum opgenomen in het woordenboek. In het veldwerk gingen de onderzoekers na of het perfectum noodzakelijk is voor de idiomatiche interpretatie. Aan de informant werd eerst de woordenboekvorm zoals in (2) voorgelegd: *Kent u de volgende uitdrukking? Wat is de betekenis ervan?* Indien het antwoord op voorgaande vraag positief was, kreeg de informant later zinnen als (3) voorgelegd (waarbij in (3a) de onvoltooid verleden tijd wordt gebruikt, in (3b) passivisatie is toegepast, en in (3c) topicalisatie). Daarbij moest de informant antwoorden op de volgende vragen: *Kan de uitdrukking als volgt worden gebruikt? Wat is de betekenis?* Deze laatste vraag is noodzakelijk aangezien het perfect mogelijk is dat een zin grammaticaal is en door de informant kan worden gebruikt, maar zijn figuurlijke betekenis kwijtgeraakt is.

- (3) a. *A at den heif op.*
 hij at de zuurdesem op
 b. *Den heif is opgeiten geweest deir èm.*
 de zuurdesem is opgegeten geweest door hem
 c. *Den heif eit'n opgeiten.*
 de zuurdesem heeft-hij opgegeten

Op deze manier werden meer dan 13.000 individuele oordelen verzameld. Dit alles resulteerde in de Databank van Nederlandse Dialectidiomen (*Database of Dutch Dialect Idioms*, DaDDI), een online databank waarin alle verzamelde gegevens beschikbaar werden gemaakt. (Corver et al. (2017), zie <http://language.link.let.uu.nl/idioms2>).

Vervolgens kan er ook gezocht worden op deze syntactische kenmerken in de databank. Voor elke testzin wordt in de databank weergegeven welke eigenschap gemanipuleerd wordt (bijv. de aanwezigheid van een modaal hulpwerkwoord, of de bepaaldheid van het direct object), de aard van de manipulatie (bijv. deontische modaliteit wordt vervangen door epistemische modaliteit, of het direct object wordt getopicaliseerd), en de oordelen die moedertaalsprekers gaven aan die gemanipuleerde vorm.

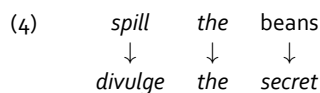
Hiermee ronden we onze beschrijving van het project *The Syntax of Idioms* en de bijhorende database af. In de volgende sectie bespreken we twee casussen die laten zien hoe de gegevens uit deze database een natuurlijke aanvulling vormen op het bestaande corpusonderzoek naar de syntactische flexibiliteit van idiomen.

3 Twee casussen

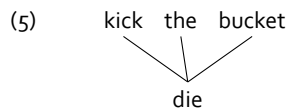
De twee casussen die in deze sectie behandeld worden, betreffen beide de (on)mogelijkheid van syntactische manipulaties van idiomen. In subsectie 3.1 wordt nagegaan of er al dan niet verschillende types idiomen (op basis van hun syntactische flexibiliteit) geattesteerd kunnen worden. In sectie 3.2 worden mogelijke correlaties tussen idioomeigenschappen besproken. Daarbij wordt twee maal vertrokken van recent corpusonderzoek naar idiomatische uitdrukkingen, dat dan aangevuld wordt met observaties uit het hierboven beschreven veldwerk. Gezien het beperkte bestek van deze bijdrage kunnen we niet in grote mate van detail ingaan op de gepresenteerde data, conclusies, en analyses, en beperken we ons tot het kort illustreren van het algemene principe, met name de manier waarop corpusgegevens en een databank met oordelen van moedertaalsprekers complementair zijn aan elkaar. Voor een meer uitvoerige bespreking van de twee casussen verwijzen we naar Temmerman et al. (2016), Stone (2017), Corver et al. (2019), Fellbaum (2019).

3.1 Twee types idiomen?

Volgens de traditionele opvatting in de literatuur (Newmeyer 1974, Nunberg 1978, Wasow et al. 1984, Gazdar et al. 1985, Fillmore et al. 1988, Nunberg et al. 1985, *inter multi alia*) zijn idiomen onder te verdelen in twee types. Enerzijds zijn er wat men noemt *idiomatically combining expressions* (ICEs): dit zijn idiomen waarbij individuele delen van de figuurlijke betekenis gerelateerd kunnen worden aan individuele delen van de letterlijke uiting. Een canoniek voorbeeld van dit type is *spill the beans* ('het geheim onthullen'):



Anderzijds zijn er de zogenoemde *idiomatic phrases* (IdPs): dit zijn idiomen waarbij de figuurlijke betekenis in zijn geheel gerelateerd wordt aan de volledige uiting. Het prototypische voorbeeld van dit type (en ook wel van idiomen in het algemeen) is *kick the bucket* ('sterven'):



ICEs zijn syntactisch flexibel: deze idiomen kunnen syntactische manipulaties zoals bijvoorbeeld topicalisatie en passivisatie ondergaan, zoals geïllustreerd wordt in (6). Deze flexibiliteit zou het gevolg zijn van het feit dat ICEs transparant/ontleedbaar zijn en de onderdelen ervan referentieel, i.e. ze krijgen krijgen een individuele semantische interpretatie (zie (4)).

- (6)
- a. *The beans, Bob has most certainly spilled.*
'Het geheim, dat heeft Bob zeer zeker onthuld.'
 - b. *The beans were spilled (by Bob).*
'Het geheim werd onthuld (door Bob).'

IdPs daarentegen zijn syntactisch niet flexibel, wat geïllustreerd wordt in (7): syntactische manipulaties zoals topicalisatie en passivisatie zijn niet mogelijk (het #-teken geeft het verlies van idiomatische betekenis weer). Dit kan worden gerelateerd aan het feit dat IdPs niet transparant/ontleedbaar zijn, en de onderdelen ervan dus niet-referentieel (zie (5)).

- (7) a. # *The bucket, Bob has gone and kicked.*
 ≠ 'Bob stierf.'
 b. # *The bucket was kicked (by Bob).*
 ≠ 'Bob stierf.'

Hoe aantrekkelijk en intuïtief deze generalisaties ook mogen zijn, de recente uitgebreide corpusstudie naar Engelse en Duitse idiomen van Fellbaum (2019) laat zien dat ze incorrect zijn. Zelfs het meest prototypische voorbeeld van IdPs, i.e. *kick the bucket*, vertoont een zekere mate van syntactische flexibiliteit:

- (8) *Live life to the fullest, you never know when the bucket will be kicked.* Passief
 (9) *No, no kicking of the bucket... not anytime soon.* Nominalisering
 (10) *We are tired of you, so hurry up and kick that bucket.* Lidwoord vervangen door demonstratief

Op basis van waarnemingen als deze besluit Fellbaum (2019:24) dat er geen duidelijke grens is tussen "semantically composing idioms" (ICEs) en "semantically non-composing idioms" (IdPs) en dat "[non-composing] idioms are not as invariant as much of the literature suggests." Zowel de specifieke lexicale samenstelling als de canonieke syntactische vorm van idiomen kan variatie vertonen, zolang het resultaat maar in de context past.

De bijdrage van Fellbaums corpusonderzoek aan het theoretische onderzoek naar idioomtypes en hun syntactische flexibiliteit is substantieel: ze laat zien dat de eenvoudige tweedeling die aanvankelijk op basis van een beperkte (op oordelen gebaseerde) dataset werd aangenomen, inaccuraat is. Het onderscheid tussen ICEs en IdPs is niet categoriaal of zwart-wit. Tegelijk echter roept haar onderzoek ook verdere vragen op. De verschillende idiomen laten immers duidelijke frequentieverschillen zien wat betreft hun syntactische flexibiliteit. Betekent dit dat het toch nog zinvol is om twee (basis)types (ICE vs. IdP) te onderscheiden, en dat er een grote grijze zone tussenin zit? Of is het inderdaad zo dat "anything goes" (Fellbaum 2019:24)?

DaDDI stelt ons in staat om deze vragen verder te exploreren. Door systematisch voor ieder idioom en iedere spreker afzonderlijk na te gaan welke vormen van syntactische flexibiliteit zijn toegestaan, wordt duidelijk dat ieder onderzocht dialect verschillende niet-transparante, niet-flexibele idiomen, i.e. IdPs, bezit. Het Aalsterse idioom dat we in (2)-(3) als voorbeeld gaven, blijkt zo'n IdP te zijn. Dit idioom kan voor geen enkele informant gepassiviseerd of getopicaliseerd worden:

- (11) a. # *Den heif* *wierd opgeiten* *deer mènne collega.*
 de zuurdesem werd opgegeten door mijn collega
 ≠ 'Mijn collega was de zondebok.'
 b. # *Den heif* *eit'n* *(oi) opgeiten.*
 de zuurdesem heeft-hij hij opgegeten
 ≠ 'Hij was de zondebok.' Aalst

Anderzijds laten de DaDDI-gegevens echter ook zien dat Fellbaum (2019) wel degelijk gelijk heeft wanneer ze stelt dat de overgang tussen ICEs en IdPs gradueel is, en niet categoriaal. In elk onderzocht dialect verschijnt namelijk het volgende beeld:

(12)		topicalisatie-ja	topicalisatie-nee
	passivisatie-ja	geattesteerd	geattesteerd
	passivisatie-nee	geattesteerd	geattesteerd

Zo vind je in het Aalsters ook idiomen die zowel getopicaliseerd als gepassiviseerd kunnen worden (cf. (13), *den tol (ni mir) kennen betolen* 'zeer arm zijn'), uitdrukkingen die wel getopicaliseerd maar niet gepassiviseerd kunnen worden (cf. (14), *zenne kees loten* 'sterven'), en uitdrukkingen die wel gepassiviseerd maar niet getopicaliseerd kunnen worden (cf. (15), *ne kemel/metten schieten* 'een flater begaan').

- (13) a. *Den tol wierd/ès zelfs ni mir betoold (deer men broer).*
de.SG.NN tol werd/is zelfs niet meer betaald door mijn broer
'Mijn broer is zeer arm.'
- b. *Den tol kaan'n zeiker ni mir betolen.*
de.SG.NN tol kan-hij zeker niet meer betalen
'Hij is zeer arm.'
- Aalst
- (14) a. *# Noo dat accedent wierd/ès zenne kees geloten (deer Jef).*
na dat ongeval werd/is zijn.SG.NN kaars gelaten door Jef
≠ 'Jef stierf na dat ongeval.'
- b. *Zenne kees liet'n passeede week.*
zijn.SG.NN kaars liet-hij vorige week
'Hij stierf vorige week.'
- Aalst
- (15) a. *Er wierd/ès ne metten/keimel geschoeten (deer mennen besten student).*
er werd/is een.NN kameel/kalf geschoten door mijn.SG.NN beste.SG.NN student
'Mijn beste student beging een flater.'
- b. *# Ne kemel/metten schoet'n (oi) zeiker!*
een.NN kameel/kalf schoot-hij (hij) zeker
≠ 'Hij beging zeker een flater!'
- Aalst

Om dit beeld nog iets concreter te maken, geven we hieronder de cijfers voor twee van de onderzochte dialecten (naast het Aalsters ook het Brugs):

(16)	Aalst (n=35)	topicalisatie–ja	topicalisatie–nee
	passivisatie–ja	20%	14%
	passivisatie–nee	23%	43%
(17)	Brugge (n=43)	topicalisatie–ja	topicalisatie–nee
	passivisatie–ja	23%	19%
	passivisatie–nee	23%	35%

De data in DaDDI bieden dus een verdere verdieping en verfijning van de conclusies van het corpusonderzoek van Fellbaum (2019): ze ondersteunen de hypothese van Espinal and Mateu (2010) dat er behalve 'zuivere' (syntactisch variable) ICEs en 'zuivere' (syntactisch niet-flexibele) IdPs ook idiomen zijn die tussen de twee types in lijken te vallen. Voor een mogelijke analyse van beperkingen op woordvolgordevariatie en syntactische flexibiliteit van de verschillende types idiomen, verwijzen we naar Trotzke (2015).

3.2 Correlaties tussen idioomeigenschappen

Stone (2017) rapporteert over een corpusonderzoek naar onder andere de correlatie tussen modificatie en passivisatie van idiomen. In (18) worden enkele relevante voorbeelden uit Stone (2017:11) gegeven op basis van het idioom *draw the line* ('grenzen stellen, iets niet doen omdat je het te vergaand vindt'). In (18a) wordt het object van het idioom gemodificeerd door een adjectief, in (18b) wordt het lidwoord vervangen door een kwantificerend element, en (18c) is een voorbeeld van passivisatie.

- (18) a. *You're gonna help murder this guy, you draw the **moral** line at extramarital sex?*
b. *We draw **another** line. And they blow up an entire village ...*
c. *The line was drawn, and I think it was a turning point for the president ...*

Uit het onderzoek van Stone blijkt dat de resultaten sterk verschillen per corpus. Op basis van *Google Books* komt ze tot de conclusie dat er geen correlatie is tussen modificatie en passivisatie, terwijl uit COCA (Davies 008) blijkt dat er een implicatieve relatie is: als passivisatie mogelijk is, dan is modificatie dat ook (Stone 2017:31). Opnieuw zien we hoe het corpusonderzoek naar idiomen de theoretische studie naar dit fenomeen verder helpt: om dit soort van correlaties te kunnen onderzoeken zijn grote hoeveelheden taalmateriaal en een systematische aanpak nodig. Tegelijk roepen Stones resultaten echter ook weer verdere vragen op: in hoeverre is de in COCA aangetroffen implicatieve relatie toepasbaar op individuele sprekers/dialecten/registers? En andersom: zou een dergelijke differentiatie (bijv. per spreker) ook in het *Google Books*-corpus een correlatie laten zien?

DaDDI laat toe deze vragen te beantwoorden en levert bovendien positieve resultaten op. In de onderzochte dialecten verschijnt namelijk het volgende beeld:

		modificatie–ja	modificatie–nee
(19)	passivisatie–ja	geattesteerd	onmogelijk
	passivisatie–nee	geattesteerd	geattesteerd

Een voorbeeld van een idiomatische uitdrukking die zowel gepassiviseerd als gemodificeerd kan worden, is het Brugse idioom in (20), *den aap luizen* ('lui zijn'). Een voorbeeld van een idioom dat wel gemodificeerd, maar niet gepassiviseerd kan worden, is de Looise uitdrukking in (21), *den ezel uithangen* ('koppig zijn, tegenwringen'). Een voorbeeld van een idioom (ook uit Tessenderlo) dat gepassiviseerd noch gemodificeerd kan worden, is (22), *ne post pakken* ('een gang gaan, vallen'). Het vierde type idioom (geen modificatie, wel passivisatie) wordt niet gevonden in deze dialecten.

- (20) a. *Den aap wordt geluist door mijn broer.*
'Mijn broer is lui.'
b. *Hij luist den **enormen** aap.*
'Hij is enorm lui.' Brugge
- (21) a. # *Den ezel is/werd uitgehangen (door hem).*
≠ 'Hij was koppig; hij wrong tegen.'
b. *Hij hangt de **koppigen** ezel uit.*
'Hij is koppig; hij wringt tegen.' Tessenderlo
- (22) a. # *Er is/werd ne post gepakt (door hem).*
≠ 'Hij is gevallen.'
b. # *Hij heeft ne **snelle/lelijke/zware/vergeefse** post gepakt.*
≠ 'Hij is gevallen.' Tessenderlo

Om dit beeld nog iets concreter te maken, geven we hieronder ook de cijfers voor deze twee dialecten:

(23)	Brugge (n=18)	modificatie–ja	modificatie–nee
	passivisatie–ja	17%	0%
	passivisatie–nee	44%	39%
(24)	Tessenderlo (n=23)	modificatie–ja	modificatie–nee
	passivisatie–ja	9%	0%
	passivisatie–nee	35%	56%

Ook in deze tweede casus zien we hoe onderzoek gebaseerd op oordelen van moedertaalsprekers ons in staat stelt om gedetailleerde en genuanceerde antwoorden te geven op vragen die voortvloeien uit voorafgaand corpusonderzoek. Op basis van de idiomen in DaDDI komen we tot het besluit dat modificatie van een idioom een noodzakelijke maar geen voldoende voorwaarde is voor passivisatie van dat idioom en dat er dus een implicatieve relatie bestaat tussen deze twee fenomenen. Deze implicatie kan ook theoretisch geïnterpreteerd worden: om 'naar buiten toe' transparant te kunnen zijn (passief), moet het idioom ook intern transparant zijn (modificatie). Terugkoppelend naar de oorspronkelijke onderzoeksvragen van het project *The Syntax of Idioms* (cf. subsectie 2), betekent dit dat de externe syntaxis van een idioom afhankelijk is van zijn interne syntaxis.

4 Besluit

In deze bijdrage werden twee concrete casussen betreffende de (on)mogelijkheid van syntactische manipulaties van idiomen behandeld. Hierbij werd recent corpusonderzoek aangevuld met oordelen van moedertaalsprekers. Op basis van deze twee casussen concluderen we dat beide vormen van datacollectie in hoge mate complementair zijn en een vruchtbare aanvulling op elkaar kunnen vormen.

Dankwoord

We willen graag de volgende mensen bedanken voor hun constructieve opmerkingen, inzichtelijke vragen, en waardevolle suggesties: Eefje Boef, Norbert Corver, Alexis Dimitriadis, Kristel Doreleijers, Will Harwood, Marko Hladnik, Dany Jaspers, Loes Koring, Sterre Leufkens, Cora Pots, Hans Smessaert, Jolijn Sonnaert, Guido Vanden Wyngaerd, de anonieme beoordelaars en editors van dit themanummer, en het publiek van de rondetafelbijeenkomst *Theorie en/in digitale geesteswetenschappen* op 26 april 2018 in Antwerpen. Dit artikel kwam mede tot stand dankzij de steun van FWO (G.Ao49.12N) en NWO (326-78-001).

Referenties

- Baggio, Giosue, Michiel van Lambalgen, and Peter Hagoort. 'Language, linguistics and cognition.' *Handbook of the Philosophy of Science*. R. Kempson, T. Fernando, and N. Asher (red.). Volume 14. Amsterdam: Elsevier, 2012. 325–355.
- Baldwin, Timothy, John Beavers, Emily M. Bender, Dan Flickinger, Ara Kim, and Stephan Oepen. 'Beauty and the Beast: What Running a Broad Coverage Precision Grammar over the BNC Taught Us about the Grammar—and the Corpus.' *Linguistic evidence. Empirical, theoretical, and computational perspectives*. Stephan Kepser and Marga Reis (red.). Berlin: Mouton de Gruyter, 2005. 49–70.
- Chomsky, Noam. *Current issues in linguistic theory*. Mouton & Company, 1964.
- Chomsky, Noam. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press, 1965.
- Corver, Norbert, Jeroen van Craenenbroeck, Eefje Boef, Alexis Dimitriadis, Kristel Doreleijers, William Harwood, Marko Hladnik, Loes Koring, Sterre Leufkens, and Tanja Temmerman. *Database of Dutch Dialect Idioms (DaDDI): Online database*. 2017.
- Corver, Norbert, Jeroen van Craenenbroeck, William Harwood, Marko Hladnik, Sterre Leufkens, and Tanja Temmerman, editors. *Idioms: Linguistics Special Issue*. 2019.
- Daan, Jo and Dirk Peter Blok. *Van Randstad tot Landrand; toelichting bij de kaart: Dialecten en Naamkunde, Vol. XXXVII, Bijdragen en mededelingen der Dialectencommissie van de Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam*. Amsterdam: Noord-Hollandsche Uitgevers Maatschappij, 1969.
- Davies, Mark. *The Corpus of Contemporary American English: 450 million words, 1990-present*. 2008-.
- Espinal, Teresa M. and Jaume Mateu. 'On classes of idioms and their interpretation.' *Journal of Pragmatics* 42 (2010): 1397–1411.
- Fellbaum, Christiane, editor. *Corpus-based studies of German idioms and light verbs. Special issue of the Journal of Lexicography* 19. 2006.
- Fellbaum, Christiane, editor. *Collocations and idioms: Corpus-based linguistic and lexicographic studies*. Birmingham: Continuum Press, 2007.
- Fellbaum, Christiane. 'The syntax and grammar of idioms.' *Handbook of Syntax*. Tibor Kiss & Artemis Alexiadou (red.). Berlin: de Gruyter, 2014. 776–802.
- Fellbaum, Christiane. 'Is there a "grammar of idioms"? *BCGL8: The Grammar of Idioms*, KU Leuven Campus Brussel, 4-5 June 2015, 2015.
- Fellbaum, Christiane. 'Is there a grammar of idioms? A corpus-based study.' *Idioms: Linguistics Special Issue*. Norbert Corver, Jeroen van Craenenbroeck, William Harwood, Marko Hladnik, Sterre Leufkens & Tanja Temmerman (red.). 2019.
- Fillmore, Charles J., Paul Kay, and Mary C. O'Connor. 'Regularity and idiomaticity in grammatical constructions: The case of Let Alone.' *Language* 64 (1988): 501–538.
- Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag. *Generalized phrase structure grammar*. Oxford: Basil Blackwell, 1985.

- Grégoire, Nicole. *Untangling Multiword Expressions: A study on the representation and variation of Dutch multiword expressions*. PhD thesis, Universiteit Utrecht, 2009.
- Henry, Alison. 'Non-standard dialects and linguistic data.' *Lingua* 115 (2005): 1599–1617.
- Horvath, Julia and Tal Siloni. 'Hebrew idioms: The organization of the lexical component.' *Brill's Annual of Afroasiatic Languages and Linguistics* 1 (2009): 283–310.
- Horvath, Julia and Tal Siloni. Idioms: The type-sensitive storage model. *TAU-GU Workshop on Relative Clauses*, Goethe Universität Frankfurt, 27-30 November 2014, 2014.
- Horvath, Julia and Tal Siloni. Idioms: The type-sensitive storage model. *BCGL8: The Grammar of Idioms*, KU Leuven Campus Brussel, 4-5 June 2015, 2015.
- Horvath, Julia and Tal Siloni. Idioms: The type-sensitive storage model. Ms., Tel Aviv University, 2016.
- Leivada, Evelina, Roberta D'Alessandro, and Kleanthes K. Grohmann. 'Eliciting Big Data From Small, Young, or Non-standard Languages: 10 Experimental Challenges.' *Frontiers in Psychology* 10 (2019).
- Newmeyer, Frederick. 'The regularity of idiom behavior.' *Lingua* 34 (1974): 327–342.
- Nunberg, Geoffrey. *The pragmatics of reference*. PhD thesis, City University New York, 1978.
- Nunberg, Geoffrey, Ivan Sag, and Thomas Wasow. 'Idioms.' *Language* 70 (1985): 491–538.
- Sailer, Manfred. Possessive (non-)alternations in idioms. *BCGL8: The Grammar of Idioms*, KU Leuven Campus Brussel, 4-5 June 2015, 2015.
- Schütze, Carson T. *The empirical basis of linguistics: grammaticality judgments and linguistic methodology*. Originally published in 1996 with Chicago University Press. Berlin: Language Science Press, 2016.
- Skilton, Amalia. 'Three speakers, four dialects: Documenting variation in an endangered amazonian language.' *Documenting Variation in Endangered Languages*. K.A. Hildebrandt, C. Jany, and W. Silva (red.). Honolulu: University of Hawai'i Press, 2017. 94–115.
- Sprouse, Jon and Diogo A. de A. Almeida. 'Design sensitivity and statistical power in acceptability judgment experiments.' *Glossa* 2 (2017): 1–32.
- Stone, Megan. Systematic flexibility in verb-object idioms. Ms. University of Arizona, 2017.
- Temmerman, Tanja, Sterre Leufkens, William Harwood, and Marko Hladnik. There's nothing special about the syntax of idioms: With new evidence from Dutch dialects. *The 31st Comparative Germanic Syntax Workshop (CGSW31)*, Universiteit van Stellenbosch, 1-3 December, 2016.
- Trotzke, Andreas. 'How bizarre! A non-syntactic constraint on idiom syntax.' *Yearbook of the German Cognitive Linguistics Association* 3 (2015): 219–230.
- Wasow, Thomas, Geoffrey Nunberg, and Ivan Sag. 'Idioms: An interim report.' *Proceedings of the 13th International Congress of Linguists*. S. Hattori (red.) 1984, 102–115.