

De Spaanse inchoatiefconstructie in beeld. Clusteranalyse als antwoord op het quasi-synonymie vraagstuk.

Sven VAN HULLE & Renata ENGHELS

Abstract

This study about the productivity and the near-synonymy of the inchoative construction in Peninsular Spanish reveals that at least 25 verbs can enter the auxiliary slot. These auxiliaries are recruited from six semantic domains: (1) superlexical inherent inchoative verbs, (2) put verbs, (3) throw verbs, (4) movement verbs, (5) manual manipulation verbs and (6) verbs of destruction. This study confirms that a hierarchical cluster analysis is suited to investigate the near-synonymy of semantically related verbs as it unfolds interesting clusters for the inchoative auxiliaries. Concretely, the hierarchical cluster analysis divides the auxiliaries into three clusters. A first one with the superlexical auxiliaries (*comenzar*, *empezar* and *iniciar*), a second cluster with the put verbs (*meter* and *poner*) and a third cluster with the less grammaticalized and more lexicalized auxiliaries *echarse* and *romper*.

1. INLEIDING: INCHOATIVITEIT EN PRODUCTIVITEIT IN HET SPAANS

De inchoatiefconstructie drukt het begin uit van een actie of gebeurtenis en is opgebouwd uit vier constituenten of ‘slots’: een onderwerp, een hulpwerkwoord, een voorzetsel en een infinitief (1).

(1)	Juan	empieza	a	trabajar.
	[O]	[hulpww.]	[vz.]	[inf]
	“Jan	begint	te	werken”

Deze studie over de inchoatiefconstructie in het Iberisch Spaans bestudeert de productiviteit van het hulpwerkwoordslot en het infinitiefslot. In tegenstelling tot andere Romaanse talen is de weinig bestudeerde constructie opvallend productief geworden in het Spaans (Fernández 2012, Garachana 2017, Fernández Martín 2019). Een uitgebreid gamma aan werkwoorden, behorend tot verschillende semantische klassen, kan dienst doen als hulpwerkwoord in de constructie, onder andere gooi-werkwoorden (*Se echó a reír*, letterlijk: “hij/zij gooide zichzelf te lachen”), plaatsings-werkwoorden (*Se mete a escribir*, letterlijk: “hij/zij plaatst zichzelf te schrijven”) of destructie-werkwoorden (*Se rompe a llorar*, letterlijk “hij/zij breekt te wenen”) (Enghels & Van Hulle 2018). Het eerste doel van deze studie is om de hoge graad van productiviteit van het hulpwerkwoordslot in kaart te brengen en beter te begrijpen (sectie 2).

De hoge productiviteitsgraad doet ook de cruciale vraag rijzen in welke mate deze hulpwerkwoorden kunnen beschouwd worden als (quasi-)synoniemen (sectie 3). De graad van quasi-synonymie zal onderzocht worden op twee verschillende niveaus. Enerzijds analyseren we het collocatieve gedrag van de hulpwerkwoorden in kwestie om gelijkenissen en verschillen op te sporen. Verbinden de hulpwerkwoorden zich met dezelfde of eerder met verschillende infinitieven? En behoren deze infinitieven tot éénzelfde semantische veld of niet? Anderzijds gaan we dieper in op de productiviteit van deze hulpwerkwoorden: vertonen semantisch verwante hulpwerkwoorden eveneens dezelfde graad van productiviteit?

Het collocatieve gedrag zal worden verkend door middel van een hiërarchische clusteranalyse (sectie 3.3.). Die berekent welke types, in ons geval hulpwerkwoorden, samen clusteren op basis van de infinitieven die ze bij voorkeur selecteren. De clusteranalyse zal onthullen of en welke semantische clusters onderscheiden kunnen worden en welke collocatieve patronen er bestaan tussen de hulpwerkwoorden en de infinitieven.

Een tweede bron van variabiliteit binnen de constructie situeert zich in het infinitiefslot, waar, afhankelijk van het hulpwerkwoord, veel of weinig verschillende werkwoorden, al dan niet van diverse semantische velden, tot worden aangetrokken. Bijgevolg bestuderen we in dit artikel ook de productiviteit van het infinitiefslot van een bepaalde micro-constructie¹. De mate van productiviteit van het infinitiefslot laat ook toe om semantische gelijkenissen en verschillen tussen de hulpwerkwoorden te detecteren.

Productiviteit is een complex en moeilijk te definiëren taalkundig concept. In deze studie hanteren we drie productiviteitsmaten: *tokenfrequentie*, *typefrequentie* en *hapaxfrequentie*. Tokenfrequentie is het totale aantal attestaties van een of alle types van een constructie in een tekst of corpus; typefrequentie is het totale aantal verschillende lexicaal instanties die kunnen voorkomen in een constructie; hapaxfrequentie is het totale aantal lexicaal types die slechts één keer voorkomen in een tekst of corpus (Barðdal 2008).

In het algemeen wordt verondersteld dat een hogere typefrequentie wijst op een hogere graad van productiviteit. Met andere woorden, het is waarschijnlijk dat een constructie met een hoog aantal verschillende werkwoorden in de toekomst nóg meer werkwoorden zal aantrekken. In dat opzicht wordt typefrequentie als belangrijke maatstaf voor productiviteit beschouwd (Bybee 1995, Goldberg 1995, Clausner & Croft 1997, Tomasello 2003, Barðdal 2008, Suttle & Goldberg 2011).

Een *hapax* (*legomenon*) is, gezien zijn lage frequentie, een meer atypisch type of werkwoord. Bijgevolg wordt aangenomen dat een constructie met een

¹ Als we het hebben over de inchoatiefconstructie in het algemeen, dan verwijzen we naar de constructie op macro-niveau, inclusief alle hulpwerkwoorden of types. De term 'micro-constructie' verwijst naar een structuur ingeleid door een concreet hulpwerkwoord.

hoog aantal hapaxen vermoedelijk steeds meer atypische werkwoorden zal incorporeren. Om die reden is hapaxfrequentie een indicatie voor de *potentiële* productiviteit van de constructie, of de mate waarin de constructie zich zal uitbreiden naar nieuwe semantische domeinen (Hilpert 2013: 128). In tegenstelling tot hapaxfrequentie, brengen token- en typefrequentie de *gerealiseerde* productiviteit van de constructie in beeld. De gerealiseerde productiviteit beschrijft de huidige productiviteitsgraad van een constructie op basis van haar tokenfrequentie en het aantal verschillende *fillers* (typefrequentie), in ons geval infinitieven, zonder voorspellingen te maken (Baayen 2009).

Hoe meer verschillende infinitieven geobserveerd worden in een micro-constructie, hoe productiever die is. Hierbij analyseren we ook de semantische velden waartoe de infinitieven behoren. In navolging van algemene studies over productiviteit beschouwen we een micro-constructie met veel infinitieven uit diverse semantische domeinen als productiever dan een micro-constructie met veel infinitieven uit dezelfde semantische velden (Perek & Hilpert 2017). De productiviteitsstudie zal bijgevolg verder toelaten om gelijkenissen en verschillen te bepalen tussen de verschillende hulpwerkwoorden voor wat betreft hun graad van grammaticalisatie² of lexicalisatie³. De verwachting is dat sterk gegrammaticaliseerde hulpwerkwoorden, zoals *poner* (“plaatsen”, “zetten”) meer infinitieven uit verschillende semantische domeinen toelaten (Comer & Enghels 2017), in tegenstelling tot meer gelexicaliseerde hulpwerkwoorden, zoals *romper* (“breken”) (Enghels & Van Hulle 2018).

De concrete onderzoeksvragen voor deze studie luiden bijgevolg als volgt: (1) Welke werkwoorden komen voor in de inchoatiefconstructie in het hedendaagse Spaans en dragen bij tot de productiviteit ervan? (2) Is het mogelijk om aan de hand van een productiviteitsanalyse en een hiërarchische clusteranalyse de quasi-synonymie van de inchoatieve hulpwerkwoorden beter te begrijpen? (3) In welke mate laat de clustering toe verschillende gradaties van lexicalisatie of grammaticalisatie van de hulpwerkwoorden te onderscheiden?

Deel 2 behandelt de productiviteit van het hulpwerkwoordslot, wat nodig is om een globaal beeld te krijgen van de verschillende inchoatieve hulpwerkwoorden. Deel 3 gaat dieper in op de quasi-synonymie van die werkwoorden. Eerst bespreken we de methodologie (3.1) en de dataset (3.2) om vervolgens over te gaan tot de hiërarchische clusteranalyse (3.3).

² Grammaticalisatie wordt gedefinieerd als een proces waarbij lexemen veranderen in grammaticale vormen, en dus een grammaticale functie opnemen, en waarbij grammaticale vormen nog grammaticaler worden (Lehmann 1995: 305). Lexicalisatie kan beschouwd worden als de tegenpool van grammaticalisatie en wordt gedefinieerd als “het gebruik van een syntactische groep of woord-formatie als een nieuwe inhoudsvolle vorm die semantisch niet volledig compositioneel en relatief idiosyncratisch is.” (Traugott 2010: 276).

³ Voor een overzicht van de verschillende invullingen die aan beide termen werden toegeschreven, zie Enghels & Garachana (2021).

2. PRODUCTIVITEIT VAN HET HULPWERKWOORDENSLOT

In eerste instantie wenst deze studie na te gaan welke werkwoorden in het Spaans als inchoatief hulpwerkwoord kunnen gebruikt worden. De lijst van potentiële werkwoorden werd geconstrueerd op basis van tien ‘ingeburgerde’ inchoatieve hulpwerkwoorden, die herhaaldelijk geciteerd worden in de literatuur (Roca Pons 1958, Fente Gómez 1972, Gómez Torrego 1988, Fernández de Castro 1990, Morera 1991, García González 1992, García Fernández 2006, Garachana 2017 en Fernández Martín 2019), met name: *arrancar*, *comenzar*, *echar*, *empezar*, *explotar*, *liar*, *meter*, *poner*, *romper* en *soltar*. Zij vormden het vertrekpunt van een exhaustieve synoniemencheck die gerealiseerd werd met behulp van drie woordenboeken⁴.

Voor elk van deze 10 werkwoorden werden alle synoniemen in een lijst opgenomen, wat leidde tot een inventaris van 1000 potentiële werkwoorden. Vervolgens werd voor elk item op deze lijst nagegaan of het voorkomt in de Spaanse inchoatiefconstructie. Concreet werd aan de hand van het esTen-Ten18-corpus (cf. *infra* sectie 3.1) gecontroleerd of een bepaald werkwoord tokens kan hebben die beantwoorden aan de syntactische structuur [HULPWERKWOORD] + [a] + [INFINITIEF]⁵. Als dit niet het geval was, werd het werkwoord geschrapt als potentieel inchoatief hulpwerkwoord. Bij een positieve match, werden de exacte betekenis van het voorbeeld en het werkwoord verder handmatig geanalyseerd. Enkel als de betekenis overeenkwam met de inchoatieve lading (‘beginnen te’), werd het werkwoord in kwestie beschouwd als inchoatief hulpwerkwoord. Als resultaat van deze uitgebreide synoniemencheck werden vijftien nieuwe inchoatieve hulpwerkwoorden aan de bestaande lijst van reeds tien gekende hulpwerkwoorden toegevoegd, wat hun totaal op 25 bracht. Deze lijst vormt het vertrekpunt voor een beter begrip van de semantische relaties tussen de individuele items.

Op basis van hun originele lexicale betekenis kunnen deze inchoatieve hulpwerkwoorden worden teruggebracht tot zes semantische domeinen (Tabel 1), namelijk (1) de superlexicale⁶ inherente inchoatieve werkwoorden *comenzar*, *empezar*, *iniciar*, *principiar*, *recomenzar* en *reempezar*, (2) de plaatsings-werkwoorden *poner* en *meter*, (3) de gooi-werkwoorden *arrojar*, *echar(se)*, *lanzar* en *tirar*, (4) de bewegings-werkwoorden *embarcar* (“*inschepen*”), *largar* (“*weggaan*”), *saltar* (“*springen*”) en *zambullir* (“*duiken*”), (5)

⁴ De woordenboek van de Real Academia Española (*Diccionario de la Real Academia de la lengua Española*, www.dle.rae.es), *The Thesaurus dictionary* (https://www.lenguaje.com/herramientasV2/sinonimos_ES.html) en de *Van Dale woordenboeken* (Van Dale Woordenboeken, Nederlands – Spaans / Spaans-Nederlands).

⁵ De zoeksyntax die gebruikt werd om de constructie op te sporen, was de volgende: “[lemma="empezar"] [tag="R.*"]{0,3} "a" [tag="V.*"] within <s/>” (cf. *infra* Sectie 3.1).

⁶ Inherent inchoatieve of superlexicale werkwoorden drukken altijd de betekenis van “*beginnen*” uit, zowel binnen de constructie als hulpwerkwoord, als buiten de constructie als hoofdwkwoord. De term ‘superlexicaal’ benadrukt het feit dat hun originele semantische waarde ook aanwezig blijft in hun gebruik als inchoatief hulpwerkwoord, wat niet het geval is met de andere (meer grammaticaliseerde) hulpwerkwoorden.

de grijp-werkwoorden *agarrar* (“vastgrijpen”), *liar* (“samenbinden”) en *soltar* (“loslaten”), en (6) de destructie-werkwoorden *arrancar* (“losrukken”), *estallar* (“ontploffen”), *explotar* (“ontploffen”), *prorrumpir* (“losbarsten”), *reventar* (“openbarsten”) en *romper* (“breken”). Die indeling is gebaseerd op Levins (1993) semantische classificatie van Engelse werkwoorden.

Tabel 1. Overzicht van de 25 inchoatieve hulpwerkwoorden per semantische groep

Inherent inchoatieve werkwoorden	Plaatsings-werkwoorden	Gooi-werkwoorden	Bewegings-werkwoorden	Grijp-werkwoorden	Destructie-werkwoorden
comenzar “beginnen”	meter “plaatsen”	arrojar “gooien”	embarcar “inschepen”	agarrar “vastgrijpen”	arrancar “losrukken”
empezar “beginnen”	poner “plaatsen”	echarse “gooien”	largar “weggaan”	liar “samenbinden”	estallar “ontploffen”
iniciar “beginnen”		lanzar “lanceren”	saltar “springen”	soltar “loslaten”	explotar “ontploffen”
principiar “beginnen”		tirar “gooien”	zambullir “duiken”		prorrumpir “losbarsten”
recomenzar “herbeginnen”					reventar “openbarsten”
reempezar “herbeginnen”					romper “breken”

Tabel 1 toont aan dat het hulpwerkwoordsslot van de inchoatiefconstructie (op macro-niveau) in hoge mate productief is. We tellen niet alleen 25 verschillende werkwoorden, ook worden zij gerekruteerd uit zes verschillende semantische velden. Die vaststelling leidt tot de vraag in welke mate deze hulpwerkwoorden synoniemen zijn van elkaar, en vergelijkbaar zijn in termen van productiviteit en collocatieel gedrag.

Enerzijds drukken deze werkwoorden als inchoatief hulpwerkwoord allemaal het inchoatieve aspect uit, namelijk “beginnen te”, anderzijds is het onwaarschijnlijk dat ze perfecte synoniemen zijn. Er is in de literatuur eerder een consensus dat bepaalde varianten dezelfde semantisch-pragmatische nuances uitdrukken (Fogsgaard 2001). Volgens Glynn (2014) kunnen twee lexemen als synoniem beschouwd worden als ze vervangen kunnen worden zonder dat de ‘ware semantische’ betekenis van de zin verandert. Ook Divjak (2010) merkt op dat quasi-synoniemen twee verschillende lexemen zijn met dezelfde, of op zijn minst een zeer gelijkaardige, betekenis en dat de uitwisseling van het ene door het andere de interpretatie van de zin niet zou mogen veranderen. Tegelijkertijd voegt ze toe dat perfecte, volledige of absolute synoniemen niet bestaan omdat alle contextuele relaties van de verschillende lexicale units nooit identiek kunnen zijn. Om deze reden definieert ze synoniemen als “lexicale items met identieke betekenissen en met een aantal identieke centrale semantische kenmerken, maar die verschillen in een aantal kleinere of perifere kenmerken” (Divjak 2010: 3). Bovendien specificeert ze dat synoniemen geen vrije variatie noch complementaire distributie vertonen. Vandaar rijst de relevante vraag of de inchoatieve hulpwerkwoorden vervangen kunnen worden door elkaar, zonder de ‘ware semantische’ betekenis van de gehele zin te wijzigen. In de volgende sectie gaan we dieper in op we de mate van quasi-synonymie van de hulpwerkwoorden, en meer bepaald hun collocatieel gedrag.

3. QUASI-SYNONYMIE VAN DE INCHOATIEVE HULPWERKWOORDEN

3.1. Data en methodologie

De empirische analyse is gebaseerd op gedownloade data van het *Spanish Web corpus*⁷ (esTenTen18) (Kilgariff & Renau 2013). Dat corpus bevat internetdata, met voorbeelden die voornamelijk van fora, blogs, Wikipedia, enzoverder afkomstig zijn. Enkel het subcorpus van het Europees Spaanse subcorpus werd geconsulteerd. De zoeksyntax die gebruikt werd om de constructie op te sporen, was de volgende: “[lemma="empezar"]⁸ [tag="R.*"]

⁷ Het Spanish Web Corpus bestaat uit 20,3 miljard woorden, waarvan 3,5 miljard tot het Europees Spaanse domein behoren.

⁸ Of een ander lexicaal item uit de lijst van 25 kandidaten.

{0,3}"a"[tag="V.*"] within <s/>". Na het downloaden werden de voorbeelden manueel gefilterd. In totaal bevat de inchoatieve dataset, na het verwijderen van vals-positieve voorbeelden⁹, 5514 tokens met een maximum van 500 tokens per hulpwerkwoord. Van de hulpwerkwoorden met minder dan 500 relevante tokens in het esTenTen-corpus werden alle tokens in de dataset opgenomen; voor de hulpwerkwoorden waarvan er meer dan 500 inchoatieve tokens waren in het esTenTen-corpus werden enkel de eerste 500 geselecteerd. Dat was concreet het geval voor *comenzar*; *empezar*; *meter*; *poner*; *echar(se)*; *liar*; *arrancar* en *romper*. Tabel 2 geeft een overzicht van de tokenfrequentie per hulpwerkwoord in de dataset.

Tabel 2. Overzicht van de tokenfrequenties van de 25 hulpwerkwoorden in de dataset

Hulpwerkwoord	Tokenfrequentie
Inherente werkwoorden	1626
comenzar ("beginnen")	500
empezar ("beginnen")	500
iniciar ("beginnen")	454
principiar ("beginnen")	140
recomenzar ("herbeginnen")	28
reempezar ("herbeginnen")	4
Plaatsings-werkwoorden	1000
meter ("plaatsen")	500
poner ("plaatsen")	500
Gooi-werkwoorden	1054
arrojar ("gooien")	28
echar(se) ("gooien")	500
lanzar ("lanceren")	445
tirar ("werpen")	81
Bewegings-werkwoorden	443
embarcar ("inschepen")	30
largar ("weggaan")	176
saltar ("springen")	234
zambullir ("duiken")	3
Grijp-werkwoorden	618

⁹ Vals positieve voorbeelden bestonden, onder andere, uit taggingsfouten waarbij zelfstandige naamwoorden zoals *Superman*, *Pokémon*, *Irán* en dergelijke als infinitieven beschouwd werden, en uit voorbeelden waarbij het hulpwerkwoord in combinatie met de infinitief niet het inchoatief aspect maar de originele semantische betekenis uitdrukte, zoals bijvoorbeeld *saltar a nadar*, dat "[in het water] springen om te zwemmen" betekent en niet "beginnen te zwemmen".

Hulpwerkwoord	Tokenfrequentie
agarrar ("vastgrijpen")	23
liar ("samenbinden")	500
soltar ("loslaten")	95
Destructie-werkwoorden	773
arrancar ("losrukken")	217
estallar ("ontplofffen")	16
explotar ("ontplofffen")	10
prorrumpir ("losbarsten")	11
reventar ("openbarsten")	19
romper ("breken")	500
Totaal	5514

Deze dataset vormt de input voor een hiërarchische clusteranalyse. Die methode kan gedefinieerd worden als een collectie van verschillende algoritmes die objecten in clusters plaatst volgens duidelijk afgebakende regels van gelijkens (Divjak & Fieller 2014: 406; Manning & Schütze 1999: 498). Divjak (2010: 130) stelt dat de hiërarchische clustermethode als een van de meest fundamentele verkennende technieken geldt om grote datasets te analyseren en in betekenisvolle structuren te organiseren. De hiërarchische clusteranalyse is geen enkelvoudige techniek, maar omvat een familie van technieken om data te clusteren en in een boomstructuur of dendrogram voor te stellen (Baayen 2008: 138). De clusteranalyse die wordt toegepast op onze dataset, wordt uitgevoerd in R (R Development Core Team 2013) door middel van het package *pvcust* (Suzuki & Shimodaira 2006). We kiezen voor de hiërarchische clustermethode *hclust* en *Ward*¹⁰ als afstandsmaat.

Het dendrogram als output van deze clustermethode en afstandsmaat groepeer eenheden met een sterke gelijkens in dezelfde cluster en verdeelt eenheden met een zwakke gelijkens in afzonderlijke clusters (Divjak & Gries 2006; 2009; Divjak 2010; Baayen 2011; Gries 2013; Levshina 2015). Bovendien stelt Speece (1994: 35) dat *Ward* duidelijk interpreteerbare resultaten genereert. Daarnaast is *Ward* een frequent gebruikte strategie, die goed lijkt te werken bij een laag aantal types (Divjak 2010: 133). In ons geval zijn er acht hulpwerkwoorden die 500 keer voorkomen in onze dataset, en die we zullen opnemen in de clusteranalyse (cf. *infra* sectie 3.2).

De eerste clustering wordt uitgevoerd op basis van de lexicale invulling van het infinitiefsloot. In navolging van de stelling van Firth (1957: 11), "you shall know a word by the company it keeps", kan gesteld worden dat de infinitief

¹⁰ De afstandsmaat *Ward* test significante verschillen tussen gemiddelden van de clusters en linkt de clusters met de minste variatie (Divjak 2010: 132).

die de gebeurtenis uitdrukt die ‘begonnen wordt’ een indicatie is voor de semantiek van het hulpwerkwoord. Bijgevolg kan als hypothese gesteld worden dat de clusters op basis van de concrete infinitieven de quasi-synonymie van de verschillende inchoatieve hulpwerkwoorden in beeld brengen.

Vervolgens wordt een tweede clustering gecreëerd, ditmaal op basis van de lexicale categorie waartoe de infinitieven behoren. Die analyse is gebaseerd op de bestaande lexicale classificatie van ADESSE, dat Spaanse werkwoorden in semantische groepen opdeelt (<http://adesse.uvigo.es/data>) (García-Miguel & Albertuz 2005). De classificatie bevat zes macro-categorieën en 56 subcategorieën. Het zijn die subcategorieën waarmee de tweede clustering is uitgevoerd¹¹. De clustering op basis van de lexicale subcategorieën biedt een algemeen beeld van welke semantische domeinen voorkomen bij bepaalde hulpwerkwoorden en brengt op deze manier ook de eventuele semantische voorkeuren van de hulpwerkwoorden aan het licht.

Teneinde een antwoord te bieden op de onderzoeksvragen, analyseren we of clustering op basis van de collocaten waarmee de hulpwerkwoorden gecombineerd worden semantisch gemotiveerde clusters opleveren. De clusters worden als significant beschouwd als ze de semantische intuïties van groepen van hulpwerkwoorden die tot éénzelfde semantisch veld behoren, bevestigen (cf. *supra*, Tabel 1). Anderzijds is het ook informatief als de clusters wijzen op groepen van hulpwerkwoorden die intuïtief niet samen horen. Het is immers mogelijk dat hulpwerkwoorden die op het eerste zicht niet semantisch verwant zijn, zich toch met dezelfde infinitieven verbinden.

3.2. Overzicht van de dataset

Op een totaal van 5514 voorbeelden, observeren we 1090 verschillende werkwoorden die voorkomen in het infinitiefslot. De data toont aan dat meerdere infinitieven meermaals voorkomen in de constructie en dat er bijgevolg geen unilaterale verhouding bestaat tussen het hulpwerkwoordslot en het infinitiefslot. Het is immers niet zo dat een bepaald hulpwerkwoord zich op een unieke manier verbindt met een bepaalde infinitief, of zelfs met een semantische klasse. De verhouding, of de mate van aantrekking tussen beide slots van de constructie is trouwens een complex gegeven binnen constructiegrammatica (Goldberg 1995). Dit is een fundamentele kwestie, die verband houdt met hoe de sprekers de infinitief of het hulpwerkwoord selecteren. In de dataset

¹¹ Als een bepaald werkwoord niet in de classificatie van ADESSE opgenomen is, wordt de classificatie van een synoniem overgenomen. Als de voorgestelde classificatie niet overeenkomt met de betekenis van het werkwoord in het specifieke voorbeeld, wordt de categorie overeenkomstig die betekenis geannoteerd. In bepaalde contexten wijkt een werkwoord af van zijn standaard betekenis en wordt het figuurlijk of metaforisch gebruikt. In dat geval wordt er geannoteerd volgens die precieze betekenis in het specifieke voorbeeld, en niet volgens de standaard betekenis van het werkwoord. Ook bij polyseme werkwoorden, waarbij ADESSE verschillende categorieën voorstelt, wordt de classificatie volgens het precieze gebruik van het werkwoord in het voorbeeld overgenomen.

komen minimale paren voor waarbij identieke infinitieven zich verbinden met verschillende hulpwerkwoorden. Dit fenomeen lijkt erop te wijzen dat bepaalde hulpwerkwoorden blijkbaar kunnen alterneren zonder een te grote semantische impact. Het meest extreme voorbeeld hiervan is de infinitief *hablar* (“spreken”), die gecombineerd wordt met 16 van de 25 inchoatieve hulpwerkwoorden.

- (2) *Después de un rato, el abuelo **arrancó a hablar**.*
“Na een tijdje, **begon** de grootvader **te spreken**.”
- (3) *Ambos **se lanzan a hablar sin escuchar al otro**.*
“Beiden **beginnen te spreken** zonder naar elkaar te luisteren.”
- (4) *El hombre **reventó a hablar** porque su interior era una olla a presión.*
“De man **begon te spreken** omdat zijn binnenste een kookpot onder druk was.”

Verder bevat de dataset 539 *hapax legomena*, i.e. infinitieven die slechts één keer voorkomen in de constructie. Zoals hierboven al werd aangehaald (Sectie 1), zijn hapaxen een belangrijk instrument om de potentiële productiviteit van een constructie te bepalen. Een hoge mate van *hapax legomena* wijst er immers op dat een constructie nieuwe werkwoorden aantrekt (Barðdal 2008; Baayen 2009; Hilpert 2013). Voorbeelden 5-8 illustreren een aantal hapaxen in onze dataset, zoals *codesarrollar* (“mee-ontwikkelen”), *desmovilizar* (“demobiliseren”), *filosofear* (“filosoferen”) of *modernizar* (“moderniseren”).

- (5) *De ahí **saltó a codesarrollar** la Hipótesis Gaia con James Lovelock, postulando que el planeta entero es un ser vivo.*
“Vandaar **begon** hij de Gaia-Hypothese **mee te ontwikkelen** met James Lovelock, die stelde dat de gehele planeet een levend organisme is.”
- (6) *Stalin, [...], **había empezado a desmovilizar** gran parte de sus tropas del escenario europeo, ya desde el final de la Segunda Guerra Mundial.*
“Stalin **was begonnen** een groot deel van zijn toepen van het Europees podium **te demobiliseren**, al vanaf het einde van de Tweede Wereldoorlog.”
- (7) ***Meterse a filosofear** sobre lo ético del comportamiento de un compañero, requiere mirarse primero bien dentro.*
“**Beginnen te filosoferen** over het ethische van het gedrag van een kompaan, vereist eerst grondig in zichzelf te kijken.”
- (8) *En consecuencia, el país **comenzó a modernizar** su aparato militar y de seguridad, tanto en términos de calidad como en cantidad.*

“Bijgevolg, **begon** het land zijn militaire en veiligheidsapparaat **te moderniseren**, zowel in termen van kwaliteit als van kwantiteit.”

In onze dataset vertoont elk hulpwerkwoord ongeveer dezelfde distributie, beantwoordend aan een Zipf-verdeling, met een aantal zeer frequente infinitieven en een lange reeks weinig frequente types of *hapax legomena*. Figuren 1 en 2 brengen deze distributie duidelijk in beeld. Om praktische redenen is het niet mogelijk om het semantische profiel van alle werkwoorden gedetailleerd voor te stellen. Ter illustratie kiezen we voor twee zeer verschillende werkwoorden in de dataset, namelijk *empezar* en *romper*. *Empezar* (“beginnen”) is één van de semantisch meest open en neutrale hulpwerkwoorden, aangezien het een inherent inchoatief werkwoord is, ook als hoofdwkwoord buiten de constructie. Dat betekent dat *empezar* zich met veel verschillende infinitieven verbindt, die tot verschillende semantische klassen behoren. Voorbeelden (9) en (10) illustreren respectievelijk het gebruik van *empezar* als hoofdwkwoord buiten de constructie en als inchoatief hulpwerkwoord. *Romper* (“breken”) laat daarentegen slechts een beperkt aantal infinitieven toe. In de dataset komen bij *empezar* en *romper* respectievelijk 280 en 29 verschillende infinitieven voor, op een totaal van 500 tokens voor elk werkwoord. Op basis van het criterium van typefrequentie, tonen we twee extremen van het semantische spectrum van de constructie: een semantisch zeer open (*empezar*) en zeer gesloten (*romper*) micro-constructie.

- (9) *El programa **empezó** con la colaboración de instituciones como la Universidad de Columbia y Pensilvania.*
 “Het programma begon met de samenwerking van instellingen zoals de Universiteit van Colombia en Pennsylvania.”
- (10) *¿Sabéis por qué se **empezó a celebrar** San Valentín?*
 “Weten jullie waarom men Sint-Valentijn **begon te vieren**?”

Figuur 1 illustreert het frequentiespectrum van de infinitieven bij *empezar*¹².

¹² De X-as van deze grafiek stelt de infinitieven die voorkomen bij *empezar* voor en op de Y-as zien we de absolute frequenties van de respectievelijke infinitieven. Niet alle werkwoorden (280) kunnen worden weergegeven in deze figuur, wat de semantische productiviteit van *empezar* extra duidelijk maakt. Om deze reden hebben we ervoor gekozen om de infinitieven in kwestie niet op te nemen in de grafiek zelf, voor een duidelijke visualisatie.

Romper (“breken”) is één van de werkwoorden die meer recent – namelijk in de 19^e eeuw – tot de constructie werd aangetrokken (Enghels & Van Hulle 2018). Dat kan verklaren waarom het tot op heden minder productief is, en zich dus minder gemakkelijk met verschillende infinitieven verbindt.

We stellen vast dat er twee infinitieven zijn die heel frequent zijn in de dataset, namelijk *llorar* (“huilen”) (n=288) (11) en *hervir* (“koken”) (n=125) (12). Het lijkt er dus op dat *romper* zich gelexicaliseerd heeft, en mogelijk een idiomatische uitdrukking vormt met deze infinitieven.

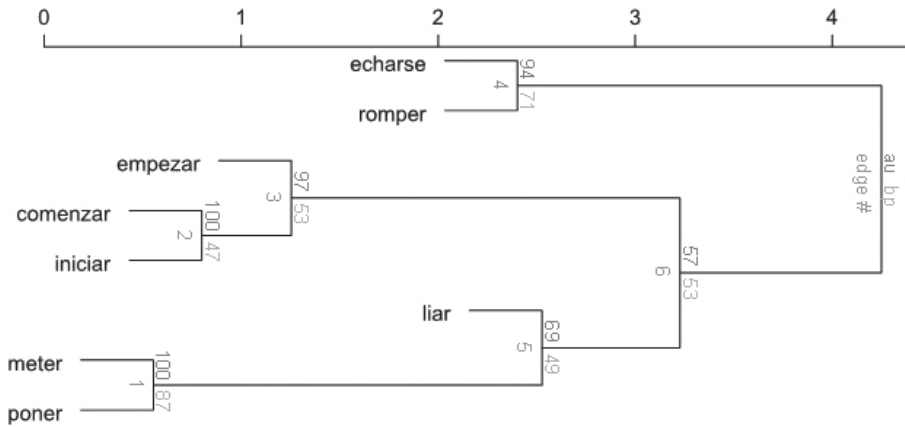
- (11) *Alfonso, un niño de 7 u 8 ocho años, rompió a llorar en un momento determinado de las deliberaciones.*
“Alfonso, een jongen van 7 of 8 jaar, **begon te wenen** op een bepaald moment van de deliberaties.”
- (12) *Cuando rompa a hervir, añadir la mezcla de maizena y remover continuamente.*
“Wanneer **het begint te koken**, de mix van maizena toevoegen en voortdurend roeren.”

Opnieuw zijn er een aantal werkwoorden met een gemiddelde frequentie tussen 2 en 23, zoals *aplaudir* (“applaudisseren”) (n=23), *reír* (“lachen”) (n=23) en *sudar* (“zweeten”) (n=21). Net zoals bij *empezar* vertoont het frequentiespectrum een lange staart van hapaxen. Nu we een beter beeld hebben van de dataset in het algemeen en van de frequentiespectra van de hulpwerkwoorden, kunnen we overgaan tot de hiërarchische clusteranalyse.

3.3. Hiërarchische clusteranalyse

Om een statistisch betrouwbare clusteranalyse te garanderen, worden voor dit deel van het onderzoek enkel die hulpwerkwoorden weerhouden waarvoor minstens 500 tokens in het ruime esTenTen corpus werden gevonden. Het is namelijk zo dat het niet gunstig is om items met een (zeer) laag aantal voorbeelden op te nemen in een clusteranalyse. Ook al is het niet strikt noodzakelijk om enkel items met identieke frequenties te selecteren voor de clusteranalyse, toch willen we vermijden dat frequentieverschillen de resultaten beïnvloeden (Divjak & Gries 2006; 2009; Divjak 2010; Levshina 2015). Daarom verkiezen we om te werken met een stabiel sample van 500 tokens per hulpwoord. Op basis van dit criterium werden de volgende acht werkwoorden geselecteerd voor verdere analyse: *echarse*, *empezar*, *comenzar*, *iniciar*, *liar*, *meter*, *poner* en *romper*. Figuur 3 toont een eerste dendrogram dat de bovenvermelde acht werkwoorden in beeld brengt op basis van hun combinatie met verschillende infinitieven.

Figuur 3. Dendrogram op basis van de infinitieven.



In het algemeen kunnen we drie clusters onderscheiden: (1) de inherent inchoatieve cluster, met *comenzar*, *empezar* en *iniciar*, (2) de cluster met de plaatsings-werkwoorden *meter* en *poner*, en (3) de cluster met *echarse* en *romper*. De inherent inchoatieve werkwoorden splitsen zich duidelijk af van de meer gegrammaticaliseerde hulpwerkwoorden, die door een proces van semantische *bleaching* in de constructie werden geïncorporeerd. Er is ook een onderscheid tussen *meter* en *poner* enerzijds, die door de eeuwen heen neutraler zijn geworden, en *echarse* en *romper* anderzijds.

Alvorens dieper in te gaan op elke cluster, staan we eerst kort stil bij de extra informatie die weergegeven wordt in het dendrogram. De getallen in het rood zijn “approximately unbiased” (AU) p-waarden en de getallen in het groen zijn “bootstrap probability” (BP) p-waarden.¹⁴ De getallen in het grijs (“edge”) kennen een ID-nummer toe aan elke knoop, op basis van de volgorde waarin de *pvclust*-functie de clusters onderscheidt. Ze hebben echter geen hiërarchische waarde: zo is cluster 1 niet sterker dan cluster 2 enzoverder. De p-waarden zijn wel van belang voor de correcte interpretatie. Het betreft hier immers significantieniveaus: hoe hoger de percentages, hoe signifikanter de clusters zijn.

In dit geval kunnen we stellen dat *comenzar* en *iniciar* significant aan elkaar geclusterd worden (100%), en ook dat *empezar* op significante wijze aan deze cluster gekoppeld wordt (97%). Deze cluster bevestigt dus de originele semantische classificatie volgens dewelke deze werkwoorden neutrale inchoatieve hulpwerkwoorden zijn, met meer semantische variatie, en

¹⁴ Wij zullen vooral beroep doen op de “approximately unbiased” (AU) p-waarden, omdat de bootstrap probability” (BP) p-waarde eerder een back-up is als de AU p-waarde niet significant is. Met andere woorden, als de AU p-waarde niet significant (lager dan 95) is, maar de BP p-waarde wel (hoger dan 95), dan kan de cluster toch nog als significant beschouwd worden. De cijfers op de x-as verwijzen naar de hoogte (van de dendrogram) waarop het aantal clusters onderscheiden worden.

dus met meer semantisch diverse infinitieven (*cf.* Tabel 1). *Meter* en *poner*, de plaatsings-werkwoorden, worden eveneens zeer duidelijk aan elkaar geclusterd (100%). Het dendrogram verbindt ook *liar* met deze werkwoorden, maar deze knoop is niet significant (69%). Dit impliceert dat *liar*, origineel een grijp-werkwoord, wel enigszins gelijkenissen vertoont met de plaatsings-werkwoorden, maar deze gelijkenis is eerder zwak, of berust op toeval. Tot slot vormen ook de minder gegrammaticaliseerde en sterker gelexicaliseerde hulpwerkwoorden *echarse* en *romper* een relevante groep (94%)¹⁵. In wat volgt bespreken we in detail de eigenschappen van de items binnen elke cluster.

Empezar, *comenzar* en *iniciar* zijn ‘neutrale’ inchoatieve werkwoorden die een hoge graad aan semantische variatie in het infinitiefslot toelaten. De andere vijf werkwoorden zijn afkomstig uit verschillende lexicale domeinen (zie Tabel 1) en zijn in meer of mindere mate gegrammaticaliseerd tot hulpwerkwoord. *Meter* en *poner*, bijvoorbeeld, zijn plaatsings-werkwoorden (“plaatsen”, “zetten”), die al sinds de 14^{de} eeuw een inchoatief gebruik kennen als hulpwerkwoord (Comer & Enghels 2017). Aangezien het grammaticalisatieproces van deze werkwoorden al vroeg heeft plaatsgevonden, zijn de micro-constructies waartoe ze behoren semantisch meer open geworden. De locatieve component van hun originele semantiek is nog aanwezig, zij het in beperkte mate, zoals duidelijk te zien is in voorbeeld (13). Dat voorbeeld¹⁶ moet eigenlijk gelezen worden als *Pon el aceite en una cacerola a calentar* (“Plaats de olie in een braadpan om ze op te warmen”). *Poner* fungeert hier tegelijkertijd zowel als plaatsingswerkwoord met een locatieve betekenis (“plaatsen”), als als hulpwerkwoord met het inchoatief aspect.

- (13) ***Pon a calentar el aceite en una cacerola.***
 “Begin de olie op te warmen in een braadpan”
 (Letterlijk: “Plaats de olie in een braadpan om ze op te warmen”)

Echarse en *romper* zijn recentere hulpwerkwoorden, die voor het eerst in de 19^{de} eeuw geobserveerd worden (Enghels & Van Hulle 2018). Als zelfstandig hoofdwkwoord betekent *echarse* “gooien” en *romper* “breken”, wat op het eerste zicht weinig compatibel lijkt met de inchoatieve betekenis. Toch kunnen ze wel degelijk optreden als inchoatief hulpwerkwoord, weliswaar met een sterk aspectuele lading, namelijk die van een abrupt, bruusk en onverwacht begin van een actie (Fente Gómez 1972, Gómez Torrego 1988, García

¹⁵ Hoewel deze cluster statistisch net niet significant is (94%), beschouwen we deze groep toch als theoretisch relevant, omdat het steek houdt dat de hulpwerkwoorden *romper* en *echarse* gelijkaardig zijn, zoals hieronder zal worden aangetoond.

¹⁶ Dit voorbeeld onthult in feite de overgangsfase of *bridging context* van de originele semantische betekenis van het hulpwerkwoord, in dit geval *poner* als locatief werkwoord, naar de aspectuele betekenis, in dit geval het inchoatief aspect (Comer & Enghels 2017).

González 1992 en Fernández de Castro 1999). Die betekeniscomponent wordt in voorbeelden (14) en (15) aangegeven door de bijwoorden *de pronto* en *de repente* ‘plots’.

- (14) *De pronto echó a correr en la dirección del montículo.*
 “Plotseling **begon** hij/zij **te rennen** in de richting van het heuveltje.”
- (15) *Puedo ser la chica más feliz de la tierra y de repente romper a llorar.*
 “Ik kan het gelukkigste meisje op aarde zijn en plotseling **beginnen te wenen.**”

Echarse en *romper* zijn bijgevolg in mindere mate grammaticaliseerd dan *poner* en *meter* omdat ze die sterke aspectuele waarde van abruptheid nog steeds uitdrukken. Dat blijkt ook uit het feit dat ze met minder verschillende infinitieven gecombineerd worden. De mate van grammaticalisatie en de semantische productiviteit van bovenvermelde werkwoorden wordt weerspiegeld in het dendrogram, dat geconstrueerd is op basis van de infinitieven die ze selecteren in het corpus.

Hoewel deze werkwoorden tot verschillende semantische groepen behoren, respectievelijk de gooi-werkwoorden en destructie-werkwoorden, hoeft het niet te verwonderen dat ze samen clusteren op basis van de infinitieven die ze selecteren. Verschillende studies bevestigen dat *echarse* en *romper* heel gelijkaardige hulpwerkwoorden zijn (Fente Gómez 1972, Gómez Torrego 1988, Fernández de Castro 1990 en García González 1992). Alvorens het collocatieve gedrag van *echarse* en *romper* meer in detail te bekijken op basis van onze data, geven we een kort overzicht van wat vorige studies over deze werkwoorden schreven.

Op basis van onze bronnen, laat *echar(se)* enkel de volgende infinitieven toe: *andar* “wandelen”, *caminar* “stappen”, *correr* “lopen”, *nadar* “zwemmen”, *temblar* “trillen”, *volar* “vliegen”, *llorar* “wenen” en *reír* “lachen” (Fente Gómez 1972, Gómez Torrego 1988, Fernández de Castro 1990 en García González 1992). De infinitieven waarmee *romper* zich verbindt, zijn volgens Fente Gómez (1972) ook heel beperkt, vooral met *llorar* “wenen” en *reír* “lachen”, maar in beperkte mate en in specifieke contexten ook met *cantar* “zingen”, *hablar* “spreken”, *andar*¹⁷ “wandelen” en *llover* “regenen”. Gómez Torrego (1988) voegt toe dat ook *gritar* “roepen” en *chillar* “gillen” en, in zeer uitzonderlijke gevallen, ook *escribir* “schrijven” bij *romper* waargenomen worden. Fernández de Castro (1999) vermeldt dat ook *romper a her-*

¹⁷ *Romper* in combinatie met de infinitieven *hablar* en *andar* drukt de specifieke betekenis uit dat een baby voor de eerste keer begint te spreken of te stappen, of dat een persoon met een fysieke beperking of na een ongeval, na een periode van revalidatie, terug in staat is om deze handelingen uit te voeren. Bijvoorbeeld: *Cuando rompió a hablar, todos nos llevamos una gran alegría.* (“Wanneer hij/zij begon te spreken, waren wij allen in grote vreugde gebracht.”) en *Hasta los quince meses no rompió a andar; estábamos muy preocupados.* (“Tot zijn 15 maanden begon hij niet te stappen; we waren zeer ongerust”) (Gómez Torrego 1988: 114).

vir mogelijk is, waarmee bedoeld wordt dat iets, meestal een vloeistof, zoals water, het kookpunt bereikt (cf. *supra* Sectie 3.2). Deze auteur voegt echter toe dat het nagenoeg onmogelijk is om een volledige lijst van infinitieven te geven, aangezien er altijd nieuwe creatieve gebruiken mogelijk zijn.

In onze dataset komen *echarse* en *romper*, op een totaal van 500 tokens, respectievelijk met 15 en 29 verschillende infinitieven voor. Dit bevestigt dat het aantal mogelijke infinitieven bij deze hulpwerkwoorden zeer beperkt zijn, in overeenstemming met de bovenvermelde bronnen. Ter vergelijking, *comenzar* verbindt zich met 287, *empezar* met 280, *iniciar* met 223, *meter* met 210 en *poner* met 179 verschillende infinitieven, op een totaal van 500 tokens. Tabel 3 illustreert de infinitieven die in onze dataset met *echarse* en *romper* geobserveerd worden¹⁸. Voor *echarse* specificeren we alle vijftien infinitieven, en voor *romper* beperken we ons, omwille van praktische redenen, tot de twaalf meest frequente types en drie hapaxen in alfabetische volgorde¹⁹.

Tabel 3. Infinitieven in combinatie met *echarse* en *romper*

echarse	vertaling	#	romper	vertaling	#
llorar	"wenen"	189	llorar	"wenen"	288
reír	"lachen"	117	hervir	"koken"	125
temblar	"trillen"	88	aplaudir	"applaudiseren"	23
dormir	"slapen"	60	reír	"lachen"	23
correr	"lopen"	17	sudar	"zweeten"	21
andar	"lopen"	14	cantar	"zingen"	4
volar	"vliegen"	3	hablar	"spreken"	3
morir	"sterven"	3	chillar	"gillen"	2
caminar	"wandelen"	3	embestir	"aanvallen"	2
navegar	"navigeren"	1	jugar	"spelen"	2
leer	"lezen"	1	sollozar	"wenen"	2
gemir	"zuchten"	1	tocar	"aanraken"	2
descansar	"rusten"	1	abjurar	"afzweren"	1
bailar	"dansen"	1	andar	"lopen"	1
arder	"branden"	1	arder	"branden"	1

Het is vooral de hoge frequentie van *llorar* bij beide hulpwerkwoorden die verklaart waarom de werkwoorden samen geclusterd worden. Andere gemeenschappelijke infinitieven zijn *andar*, *reír* en *arder*. De geobserveerde infinitieven bij *echarse* en *romper* komen in hoge mate overeen met de op-

¹⁸ Gezien hun hoge aantallen, kunnen we de infinitieven die voorkomen bij de andere hulpwerkwoorden, niet op dezelfde gedetailleerde manier weergeven.

¹⁹ De andere hapaxen bij *romper* worden vermeld in voetnoot 20.

gesomde werkwoorden in bovenvermelde studies. Onze dataset vertoont een iets hogere graad aan variatie, vermoedelijk doordat de micro-constructies recent productiever geworden zijn, en dus een hoger aantal verschillende infinitieven aantrekken. Deze infinitieven behoren wel tot dezelfde semantische categorieën (cf. *infra*). *Romper* is sterk gelexicaliseerd met *llorar* en *hervir*, en kan met deze infinitieven als idiomatische uitdrukking beschouwd worden. Anderzijds laat *romper* nog behoorlijk wat semantische variatie toe. In Tabel 3 worden veertien werkwoorden weggelaten, die bovendien allemaal hapaxen zijn²⁰. De relatief frequente aanwezigheid van hapaxen wijst erop dat er creatieve gebruiken met *romper* mogelijk zijn, zoals Fernández de Castro (1999) ook vermeldt.

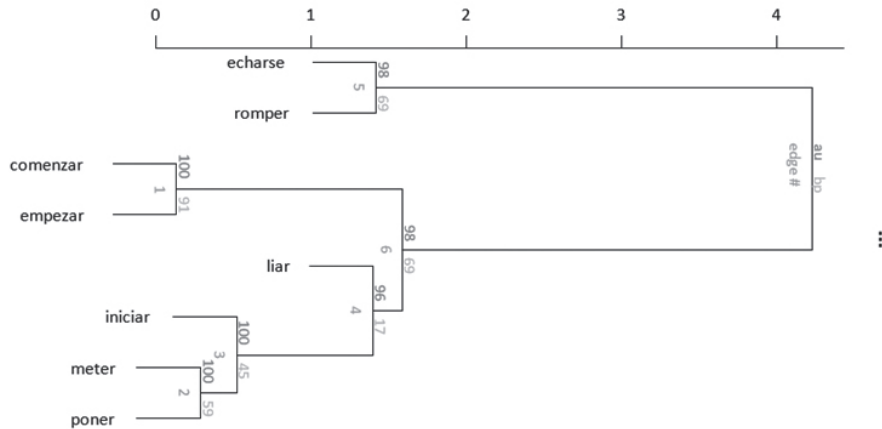
Tot nu toe hebben we de clustering op basis van de individuele infinitieven besproken. Dit heeft ons inzage gegeven in een eerste aspect van het collocatieve gedrag van de verschillende micro-constructies, namelijk met hoeveel en welke infinitieven elk hulpwerkwoord zich verbindt. Echter, om een beter beeld te krijgen van de semantische productiviteit van de inchoatiefconstructie, moeten we de semantische ‘spreiding’ van elke micro-constructie onderzoeken: bevinden de infinitieven – per micro-constructie – zich vooral binnen dezelfde semantische domeinen of niet? Als een bepaalde micro-constructie veel verschillende infinitieven aantrekt die tot hetzelfde semantische veld behoren, is die minder ‘open’ (en dus minder productief) dan een andere micro-constructie die infinitieven rekruteert uit diverse semantische velden (Perek & Hilpert 2017). Om de semantische spreiding in kaart te brengen, doen we beroep op de classificatie van ADESSE.

Op basis van de bestaande lexicale classificatie van ADESSE (cf. *supra* Sectie 3.1), hebben we de infinitieven gelabeld volgens de semantische categorie waartoe ze behoren. Door te documenteren tot welke categorie elke infinitief hoort, krijgen we een duidelijker beeld van de semantische variatie per hulpwerkwoord.

Na de toekenning van de semantische categorie waartoe elke infinitief behoort, voeren we een tweede clusteranalyse uit. Zoals eerder vermeld, zijn er in totaal 56 subcategorieën in de lexicale classificatie van ADESSE. Bijgevolg is het aantal verschillende waarden, op basis waarvan geclusterd wordt, gereduceerd tot slechts 56. Figuur 4 geeft het resultaat van die analyse.

²⁰ Deze werkwoorden zijn: *burbujea* (“uitblazen”), *callar* (“zwijgen”), *corear* (“uitroepen”), *croar* (“kwaaken”), *galopar* (“galopperen”), *granizar* (“hagelen”), *leer* (“lezen”), *llover* (“regenen”), *marcar* (“markeren”), *querer* (“willen”), *realiza* (“realiseren”), *recitar* (“reciteren”), *trabajar* (“werken”) en *volar* (“vliegen”).

Figuur 4. Dendrogram op basis van de semantische klassen.



Het bovenstaande dendrogram toont in eerste instantie dezelfde clusters als bij het eerste dendrogram (Figuur 3). Zo zien we bijvoorbeeld dat opnieuw de minder gegrammaticaliseerde en meer gelexicaliseerde hulpwerkwoorden *echarse* en *romper*, de plaatsings-werkwoorden *meter* en *poner*, en de inherent inchoatieve hulpwerkwoorden *comenzar* en *empezar* op significante wijze samen clusteren. Deze drie paren zijn in die zin dus gelijkaardig dat ze ook samen gegroepeerd worden op basis van de semantische klassen van de infinitieven die ze selecteren. Toch duiken er ook een aantal interessante verschillen op in vergelijking met het vorig dendrogram. Zo wordt de knoop tussen *liar* enerzijds en *iniciar*, *meter* en *poner* anderzijds nu ook significant (96%), terwijl dat in Figuur 3 niet het geval was (69%). Een tweede verschil is dat *iniciar*, een inherent inchoatief hulpwerkwoord, in dit dendrogram gelinkt wordt aan *poner* en *meter*, de gegrammaticaliseerde plaatsings-werkwoorden, in tegenstelling tot de eerdere indeling bij *comenzar* en *empezar*. Dat resultaat is evenwel niet geheel onverwacht.

Het is namelijk zo dat, op basis van de semantische klassen, de inherente cluster en de plaatsingscluster ook significant geclusterd worden (98%). Dit impliceert dat er op basis van de semantische spreiding van de infinitieven in feite slechts twee grote clusters onderscheiden kunnen worden: (1) de cluster met de meer gelexicaliseerde werkwoorden *echarse* en *romper* en (2) een overkoepelende cluster, opgebouwd uit de inherent inchoatieve werkwoorden *comenzar*, *empezar* en *iniciar*, en de plaatsings-werkwoorden *liar*, *meter* en *poner*, die samen één grote semantische groep vormen en zich verbinden met infinitieven uit overwegend dezelfde semantische domeinen. Om deze reden is het niet merkwaardig dat *iniciar* gekoppeld wordt aan *poner* en *meter* in plaats van aan *comenzar* of *empezar*: alle zes werkwoorden van deze grote cluster komen immers in grote mate voor met identieke lexicale categorieën.

De dominante categorie bij *echarse* en *romper* is *Fysiologie* (cfr. infra). De frequente categorieën bij de andere zes hulpwerkwoorden zijn *Activiteit* (16), *Communicatie* (17), *Creatie* (18), *Modificatie* (19), *Kennis* (20) en *Perceptie* (21). Dat zijn dan ook de semantische velden waartoe de meeste infinitieven in de inchoatiefconstructie behoren. Deze kunnen we bijgevolg als het prototypische semantische gebruik van de constructie beschouwen. In wat volgt gaan we dieper in op de lexicale categorieën die elke cluster selecteert. Op die manier brengen we de semantische voorkeuren van de verschillende clusters en hulpwerkwoorden in beeld. We zoomen eerst in op de inherent inchoatieve cluster. Tabel 4 geeft de 20 meest frequente klassen bij *comenzar*, *empezar* en *iniciar* weer.

- (16) *El doctor Álvaro Gándara **comenzó a trabajar** en esta área hace unos 12 años.*
 “Dokter Álvaro Gándara **begon** ongeveer 12 jaar geleden in dat domein **te werken**”
- (17) *Alexander dejó de cantar y **se puso a hablar** como siempre.*
 “Alexander stopte met zingen en **begon te spreken** zoals altijd.”
- (18) *Tony Stark **se lía a construir** Iron Men que luego acuden a el con una simple llamada.*
 “Tony Stark **begon Iron Men te bouwen** die daarna naar hem snellen met een simpel bevel”
- (19) *Con un ligero crujido la compuerta de la cúpula del telescopio **empezó a abrirse**.*
 “Met een licht gekraak **begon** de sluis van de koepel van de telescoop **open te gaan**.”
- (20) *Tendría que **meterme a analizarlo** más profundamente.*
 “Ik zou moeten **beginnen het** grondiger **te analyseren**.”
- (21) *Al mes y medio ya puede **iniciarse a ver** un poco de resultado.*
 “Na een anderhalve maand kan men al **beginnen** een beetje resultaat **te zien**.”

Tabel 4. De 20 meest frequente klassen bij *comenzar*, *empezar* en *iniciar*.

comenzar	#	empezar		iniciar	#
Verplaatsing	55	Perceptie	50	Activiteit	56
Activiteit	52	Verplaatsing	46	Creatie	43
Creatie	40	Modificatie	41	Modificatie	32
Modificatie	38	Activiteit	39	Verplaatsing	31
Perceptie	38	Communicatie	37	Kennis	28
Transfer	23	Kennis	21	Communicatie	27
Gebruik	19	Creatie	21	Hulpwerkwoorden	24

comenzar	#	empezar		iniciar	#
Attributie	18	Sensatie	20	Perceptie	23
Kennis	18	Attributie	19	Inname	21
Communicatie	17	Hulpwerkwoorden	18	Transfer	15
Lokalisatie	14	Bestaan	16	Gebruik	15
Bestaan	13	Transfer	15	Verandering	13
Acquisitie	12	Bezit	12	Fase	13
Emissie	10	Acquisitie	11	Acquisitie	11
Fase	10	Gebruik	11	Dispositie	11
Fysiologie	10	Emissie	10	Lokalisatie	9
Bezit	10	Fysiologie	10	Attributie	8
Hulpwerkwoorden	10	Lokalisatie	9	Bezit	8
Controle	9	Contact	8	Emissie	7
Unie	9	Relatie	8	Unie	7

Eerder stelden we dat *comenzar*, *empezar* en *iniciar*, de inherent inchoatieve hulpwerkwoorden, neutraler zijn en een hoge graad van semantische variatie vertonen. Deze stelling wordt bevestigd door onze data. De infinitieven die geobserveerd worden bij *comenzar* behoren tot 43 verschillende semantische subcategorieën, bij *empezar* tot 47 en bij *iniciar* tot 43, van de 56 in totaal. Het aantal gemeenschappelijke semantische klassen is aanzienlijk: bij *comenzar* is er maar één klasse die niet voorkomt bij *empezar* en *iniciar*, namelijk *Controle* (bv. *bevrijden*, *domineren*). Bij *empezar* zijn de klassen van *Sensatie* (bv. *voelen*, *genieten*), *Contact* (bv. *slaan*, *bijten*) en *Relatie* (bv. *verwijzen*, *afhangen*) uniek. En voor *iniciar* zijn de klassen van *Inname* (bv. *drinken*, *eten*), *Verandering* (bv. *bewerken*, *configureren*) en *Dispositie* (bv. *proberen*, *presteren*) specifiek. Alle andere klassen overlappen bij de drie inherent inchoatieve hulpwerkwoorden. De meest frequente klassen die bij alle drie waargenomen worden, zijn *Activiteit* (bv. *werken*, *spelen*), *Communicatie* (bv. *spreken*, *antwoorden*), *Creatie* (bv. *produceren*, *uitvinden*), *Kennis* (bv. *weten*, *kennen*), *Modificatie* (bv. *veranderen*, *oplossen*), *Perceptie* (bv. *zien*, *ontdekken*) en *Verplaatsing* (bv. *rennen*, *val-len*). Logischerwijs zijn de acties waarvan het begin uitgedrukt wordt, meestal van duratieve aard en dat blijkt bij bovenvermelde klassen ook het geval te zijn. Een activiteit (22), creatie (23), verplaatsing (24), en perceptie (25) zijn allemaal acties die een bepaalde tijdsduur impliceren.

(22) *La empresa comenzó a trabajar para laboratorios de agua y medio ambiente.*

“Het bedrijf begon te werken voor laboratoria van water en milieu.”

- (23) *El templo se **inició a construir** alrededor del año 1284 a.C.*
 “De tempel **begon gebouwd te worden** rond het jaar 1284 voor Christus.”
- (24) *Los frutos ya maduraron y **comienzan a caer** del árbol.*
 “De vruchten zijn al gerijpt en **beginnen** van de boom **te vallen.**”
- (25) ***Empecé a notar** el cansancio y necesitaba parar.*
 “Ik **begon** de vermoeidheid **te merken** en ik moest stoppen.”

Meter verbindt zich met 42 verschillende semantische klassen en *poner* met 37. Beide hulpwerkwoorden laten dus nog steeds behoorlijk veel variatie toe. *Echarse* en *romper* komen voor met respectievelijk 9 en 16 klassen en zijn dus semantisch minder productief. Dit is te verklaren doordat deze werkwoorden recenter tot de constructie zijn aangetrokken en minder sterk gegrammaticaliseerd zijn. Het is echter weinig waarschijnlijk dat de micro-constructies met *echarse* en *romper* in de toekomst productiever zullen worden omdat hun productiviteit al meerdere eeuwen niet of weinig is toegenomen. Integendeel, de diachrone data lijken aan te tonen dat de constructies in de 21^{ste} eeuw net dalen in productiviteit (Van Hulle, Enghels & Jansegers in preparation).

Ook bij de plaatsings-werkwoorden *poner* en *meter*, en het grijp-werkwoord *liar*, stellen we veel gemeenschappelijke semantische klassen vast. En opnieuw zijn dezelfde klassen als bij de inherente inchoatieve werkwoorden de meest frequente, met name *Activiteit*, *Communicatie*, *Modificatie* en *Kenis*, maar ook *Creatie* en *Verplaatsing*. De unieke klassen bij *liar* zijn *Emissie* (bv. *uitademen*, *schijnen*), *Substitutie* (bv. *vervangen*, *uitwisselen*) en *Leven* (*overleven*, *redden*), bij *meter* *Sociale relaties* (bv. *bezoeken*, *trouwen*) en *Attributie* (bv. *worden*, *resulteren*) en bij *poner* *Cognitie* (bv. *nadenken*, *overwegen*), *Inname* (bv. *drinken*, *eten*), *Beweging* (bv. *draaien*, *trillen*) en *Gebruik* (bv. *bedienen*, *misbruiken*). *Liar* blijkt een apart collocatieel gedrag te vertonen, wat reeds bleek uit het eerste dendrogram (cf. *Figuur 3*). Zo is de klasse *Hulpwerkwoorden* (bv. *uitvoeren*, *realiseren*) heel frequent bij *liar* (n=68) en komt die klasse in mindere mate voor bij *meter* (n=40) en *poner* (n=26). Ook *Transfer* (bv. *overbrengen*, *uitzenden*) is relatief frequent bij *liar* (n=40) en die klasse duikt niet op bij *meter* en is heel weinig frequent bij *poner* (n=3).

Verder vinden we de meest frequente klassen van *liar* ook terug bij *poner* en *meter*. Dit verklaart waarom *liar* geclusterd werd aan *poner* en *meter*, zij het op een minder significante manier. Enerzijds verbindt *liar* zich met dezelfde klassen als de plaatsings-werkwoorden, maar minder frequent, en anderzijds komen ook klassen voor, zoals *Hulpwerkwoorden* en *Transfer*, die weinig of niet waargenomen worden bij *poner* en *meter*. Tabel 5 bevat de 20 meest frequente klassen bij *liar*, *meter* en *poner*.

Tabel 5. De 20 meest frequente klassen bij *liar*, *meter* en *poner*.

liar	#	meter	#	poner	#
Hulpwerkwoorden	68	Modificatie	53	Activiteit	95
Communicatie	61	Activiteit	52	Modificatie	69
Modificatie	47	Perceptie	52	Communicatie	53
Transfer	40	Communicatie	48	Kennis	46
Verplaatsing	37	Hulpwerkwoorden	40	Perceptie	44
Creatie	35	Kennis	33	Creatie	28
Lokalisatie	28	Creatie	33	Hulpwerkwoorden	26
Leven	16	Verandering	23	Fysiologie	23
Contact	15	Verplaatsing	22	Verplaatsing	20
Verandering	14	Competitie	14	Gebruik	13
Perceptie	14	Acquisitie	13	Cognitie	10
Activiteit	13	Contact	10	Competitie	8
Acquisitie	12	Waardering	9	Geluidsemissie	7
Waardering	9	Fysiologie	8	Verandering	6
Controle	8	Dispositie	7	Inname	5
Kennis	7	Relatie	7	Lokalisatie	5
Substitutie	7	Sociale relaties	6	Beweging	5
Dispositie	6	Transfer	6	Relatie	5
Emissie	5	Attributie	5	Contact	3
Geluidsemissie	5	Controle	5	Transfer	3

Tabel 5 onthult eveneens waarom *meter* en *poner* samen geclusterd worden: beide worden het vaakst gecombineerd met infinitieven die een *Activiteit* (26) of *Modificatie* (27) uitdrukken. Ook *Communicatie* (28), *Creatie* (29), *Kennis* (30) en *Perceptie* (31) zijn gemeenschappelijke frequente semantische klassen. *Meter* en *poner* zijn bijgevolg in hoge mate synoniem: (a) ze behoren beide tot de semantische groep van plaatsings-werkwoorden, (b) ze verbinden zich hoofdzakelijk met dezelfde infinitieven, en (c) ze verkiezen in hoge mate dezelfde semantische klassen.

- (26) *Y se metió a trabajar en el despacho de Burgos.*
 “En hij/zij **begon te werken** in het kantoor van Burgos.”
- (27) *Con este contrato Hergón se mete a arreglar vías urbanas del municipio.*
 “Met dit contract **begon** Hergón de urbane wegen van de gemeente **in orde te brengen.**”

- (28) *Se ponen a hablar entre ellas en catalán.*
 “Ze **beginnen** onder elkaar **te spreken** in het Catalaans.”
- (29) *El gobierno decidió meterse a hacer armas nucleares.*
 “De regering heeft beslist om nucleaire wapens te **beginnen ma-**
ken.”
- (30) *Nos pusimos a investigar la viabilidad técnica del proyecto.*
 “We **begonnen** de technische uitvoerbaarheid van het project **te on-**
derzoeken.”
- (31) *Me he metido a leer esto por curiosidad.*
 “Ik **ben begonnen** dit **te lezen** uit nieuwsgierigheid.”

Tot slot illustreren we de semantische klassen bij *echarse* en *romper* (Tabel 6).

Tabel 6. De semantische klassen bij *echarse* en *romper*.

romper	#	echarse	#
Fysiologie	319	Fysiologie	366
Modificatie	124	Beweging	86
Waardering	23	Verplaatsing	37
Geluidsemisatie	7	Gebruik	3
Communicatie	6	Leven	3
Activiteit	5	Activiteit	2
Competitie	3	Verandering	1
Verplaatsing	2	Geluidsemisatie	1
Emissie	2	Perceptie	1
Meteorologie	2		
Sociale relaties	2		
Verandering	1		
Kennis	1		
Perceptie	1		
Gebruik	1		
Wil	1		

Het valt meteen op dat de klasse *Fysiologie* sterk domineert bij beide werkwoorden. Die klasse bestaat uit werkwoorden zoals *llorar* (“wenen”), *reír* (“lachen”) en *sudar* (“zweten”), die frequent geobserveerd worden bij beide werkwoorden. Zoals verwacht (*cfr.* Tabel 3) treffen we meer klassen en dus meer semantische spreiding aan bij *romper* dan bij *echarse*. De typische klassen bij *romper* zijn *Communicatie* (32), *Geluidsemisatie* (33), *Modificatie* (34) en *Waardering* (35). Dat *Modificatie* zo frequent is, is te wijten aan de

hoge aanwezigheid van *hervir* (“koken”), zoals eerder vermeld. *Waardering* wordt vertegenwoordigd door *aplaudir* (“applaudisseren”). *Geluidsemissie* en *Communicatie* horen samen, aangezien zij beide bestaan uit semantisch verwante werkwoorden, namelijk *gritar* (“roepen”), *chillar* (“schreeuwen”), *cantar* (“zingen”), *hablar* (“spreken”) en *decir* (“zeggen”). De andere twee frequente klassen bij *echarse* zijn *Beweging* en *Verplaatsing* (36), die ook semantisch verwant zijn. Beide categorieën worden onder andere opgevuld door respectievelijk *temblar* (“trillen”), *bailar* (“dansen”), *rodar* (“draaien”) en *correr* (“lopen”), *caminar* (“wandelen”), *andar* (“gaan”), *volar* (“vliegen”).

- (32) *Va perdiendo paulatinamente su cautela y ha roto a hablar en público*
 “Geleidelijk aan verliest hij zijn voorzichtigheid en **begon** hij/zij **te spreken** in publiek.”
- (33) *Rompió a recitar todo el monólogo de Segismundo.*
 “Hij/Zij **begon** de volledige monoloog van Segismundo **te reciteren**.”
- (34) *Cuando rompa a hervir, añadir la mezcla de maizena y remover continuamente.*
 “Wanneer het **begint te koken**, de mix van maizena toevoegen en voortdurend roeren.”
- (35) *Por fin, el público rompió a aplaudir con fuerza y sinceridad.*
 “Ten slotte **begon** het publiek **te applaudisseren** met kracht en oprechtheid.”
- (36) *El niño abrió los ojos y echó a correr de regreso a su casa.*
 “De jongen opende zijn ogen en **begon** terug naar huis **te rennen**.”

4. CONCLUSIE

Deze studie over de productiviteit van de inchoatiefconstructie in het Iberisch Spaans toont aan dat minstens 25 werkwoorden kunnen fungeren als inchoatief hulpwerkwoord. Deze hulpwerkwoorden worden gerekruteerd uit zes semantische velden: (1) superlexicale inherente inchoatieve werkwoorden, (2) plaatsings-werkwoorden, (3) gooi-werkwoorden, (4) bewegings-werkwoorden, (5) grijp-werkwoorden en (6) destructie-werkwoorden. Door middel van een hiërarchische clusteranalyse konden in eerste instantie drie clusters onderscheiden worden: een eerste cluster met de superlexicale werkwoorden (*comenzar*, *empezar* en *iniciar*), een tweede cluster met de grammaticaliseerde plaatsings-werkwoorden (*meter* en *poner*) en een derde cluster met de meer gelexicaliseerde werkwoorden *echarse* en *romper*.

Op basis van deze clusters kan afgeleid worden dat de inherent inchoatieve werkwoorden en de grammaticaliseerde werkwoorden een apart colloca-

tioneel gedrag vertonen. *Poner* en *meter* zijn vroeger gegrammaticaliseerd en door de eeuwen heen neutraler geworden, en leunen dus dichter aan bij de semantiek van de inherente werkwoorden *comenzar*, *empezar* en *iniciar*. *Echarse* en *romper* daarentegen zijn recenter tot de constructie aangetrokken en vormen een aparte cluster, met hun eigen specifiek collocatoneel patroon.

Het dendrogram op basis van de semantische categorieën bevestigt bovendien dat *echarse* en *romper* een ander gedrag vertonen dan zowel de inherente als de gegrammaticaliseerde plaatsings-werkwoorden: ze laten andere of meer dan wel minder semantische velden toe, al dan niet met een andere distributie. Dat leidt tot de conclusie dat, op semantisch vlak, er twee clusters onderscheiden kunnen worden, namelijk (1) een cluster met de gelexicaliseerde micro-constructies met *echarse* en *romper* met hun specifieke semantische voorkeuren, en (2) een overkoepelende cluster met enerzijds de inherent inchoatieve werkwoorden en anderzijds de plaatsings-werkwoorden, die zich hoofdzakelijk verbinden met dezelfde semantische categorieën. Deze categorieën zijn de semantische velden van *Activiteit*, *Communicatie*, *Creatie*, *Kenis*, *Modificatie* en *Perceptie*, die derhalve als het prototypische semantische gebruik van de inchoatiefconstructie kunnen worden beschouwd.

In het algemeen kan gesteld worden dat zowel concrete lexicaal items als de bijhorende semantische categorieën relevante criteria zijn om zinvolle groeperingen te onderscheiden. Het is evenwel nodig te nuanceren dat andere clustermethodes, eventueel met extra parameters, mogelijks meer diepgang kunnen bieden en op die manier meer gedetailleerde, specifiekere clusters onderscheiden kunnen worden. In dat opzicht stelt deze studie een eerder exploratieve analyse voor, die de aanzet geeft voor meer gevorderde (cluster)technieken.

Door de sterke methodologische restricties konden we bovendien slechts een beeld krijgen van acht van de 25 hulpwerkwoorden. Dankzij deze clustering krijgen we een inkijk in de quasi-synonymie op het meest algemene niveau, maar ze legt niet uit waarom er 25 verschillende inchoatieve hulpwerkwoorden zijn in het Spaans. Daarvoor is een meer gedetailleerde analyse nodig, die mogelijks ook rekening houdt met andere eigenschappen van de micro-constructies. Andere clustermethodes, zoals een *Behavioral Profile Analysis* (Divjak & Gries 2008, Divjak 2010 en Gries 2012), lijken in deze context veelbelovend.

BIBLIOGRAFISCHE REFERENTIES

- ADESSE corpus. *Base de datos de verbos, alternancias de diátesis y esquemas sintáctico semánticos del español*: <http://adesse.uvigo.es/> (ADESSE) (09-01-2018).
- Albertuz, F. J. (2007). "Sintaxis, semántica y clases de verbos: clasificación verbal en el proyecto ADESSE". In *Actas del VI Congreso de Lingüística General* (Santiago de Compostela, mayo 2004). Madrid: Arco Libros, pp. 2015-2030.

- Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baayen, R. (2009). "Corpus linguistics in morphology: Morphological productivity", *Corpus linguistics*, 41, 899-919.
- Barðdal, J. (2008). *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam/Philadelphia: John Benjamins.
- Bergs, A., & Diewald, G. (eds.) (2008). *Constructions and language change*. Berlin: Mouton de Gruyter.
- Brinton, L. J., & Traugott, E. (2005). *Lexicalization and language change*. Cambridge: Cambridge University Press.
- Bybee, J. (1995). "Regular morphology and the lexicon", *Language and cognitive processes*, 10(5), 425-455.
- Bybee, J., & Torres Cacoullos, R. (2009). "The role of prefabs in grammaticization." In Corrigan R., Moravcsik E., Ouali H. & K. Wheatley (eds.) *Formulaic Language*. Amsterdam/Philadelphia: John Benjamins, 187-217.
- Clausner, T. C., & Croft, W. (1997). "Productivity and schematicity in metaphors", *Cognitive science*, 21(3), 247-282.
- Comer, M., & Enghels, R. (2017). "La evolución de las perífrasis verbales causativa e incoativa con poner en español: cambios de construcción y extensiones metafóricas", *Bulletin of Hispanic Studies*, 94 (9), 903-924.
- Divjak, D., & Gries, S. (2006). "Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, 2, 23-60.
- Divjak, D., & Gries, S. (2008). "Clusters in the mind?: Converging evidence from near synonymy in Russian". *The Mental Lexicon*, 3(2), 188-213.
- Divjak, D., & Gries, S. (2009). "Corpus-based cognitive semantics: A contrastive study of phasal verbs in English and Russian". In: Dziwirek, K., & Lewandowska-Tomaszyk, B. (eds.) *Studies in cognitive corpus linguistics*. Frankfurt am Main: Peter Lang, 273-296.
- Divjak, D. (2010). *Structuring the lexicon : a clustered model for near-synonymy*. Berlin: De Gruyter Mouton
- Divjak, D., & Fieller, N. (2014). Cluster analysis: Finding structure in linguistic data. In Dylan Glynn & Justyna R. Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*. Amsterdam: John Benjamins, 405-441.
- Enghels, R., & Van Hulle, S. (2018). "El desarrollo de perífrasis incoativas cuasi-sinónimas : entre construccionalización y lexicalización", *Elua*, 32, 91-110.
- Fente Gómez, R., Feijóo, L. G., & Fernández Alvarez, J. (1972). *Perífrasis verbales*. 2a ed. Madrid: Sociedad general española de librería.
- Fernández de Castro, F. (1999). *Las perífrasis verbales en el español actual*. Madrid: Gredos.
- Fernández, L.G. ed. (2006). *Diccionario de perífrasis verbales*. Madrid: Gredos.
- Fernández, L.G. (2012). *Las perífrasis verbales en español*. Madrid: Edhasa.
- Fernández Martín, P. (2019). *Las perífrasis verbales del español: una perspectiva histórica*. Madrid: Arco/Libros.
- Firth, J. R. (1957). *Papers in linguistics 1934-1951*. London: Oxford University Press.
- Garachana, M. ed. (2017). *La gramática en la diacronía: la evolución de las perífrasis verbales modales en español*. Madrid: Iberoamericana Vervuert.

- García González, J. (1992). *Perífrasis verbales*. Madrid: Sociedad general española de librería.
- García-Miguel, J. M., & Albertuz, F. J. (2005). Verbs, semantic classes and semantic roles in the ADESSE project. In *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*, 50-55.
- Goldberg, A. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Gómez Torrego, L. (1988). *Perífrasis verbales : sintaxis, semántica y estilística*. Madrid: Arco/Libros.
- Gries, S. (2006). "Corpus-based methods and cognitive semantics: The many senses of *to run*." In: Gries, S., & Stefanowitsch, A. (eds.), *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis*. Berlin/Boston: De Gruyter Mouton, 57-99.
- Gries, S. (2010). "Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics". *The Mental Lexicon*, 5(3), 323-346.
- Gries, S. (2012). "Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics". In Gonia Libben, Gary Jarema & Chris Westbury (eds.), *Methodological and analytic frontiers in lexical research*, Amsterdam & Philadelphia: John Benjamins, 57–80.
- Gries, S. (2013). *Statistics for Linguistics with R: a practical introduction*. Berlin/Boston: De Gruyter Mouton.
- Hilpert, M. (2013). *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press.
- Hopper, P. J., & Traugott, E. (2008). *Grammaticalization*. 4th pr., 2nd ed. Cambridge: Cambridge University Press.
- Kilgarriff, A., & Renau, I. (2013). esTenTen, a vast web corpus of Peninsular and American Spanish. *Procedia-Social and Behavioral Sciences*, 95, 12-19.
- Lehmann, C. (1995). *Thoughts on grammaticalization*. München: Lincom Europa.
- Lehmann, C. (2002). "New reflections on grammaticalization and lexicalization". In Wischer I. & Diewald, G. (eds.). *New Reflections on Grammaticalization*. Amsterdam: John Benjamins, 1-18.
- Levin, B. (1993). *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago press.
- Levshina, N. (2015). *How to Do Linguistics with R : Data Exploration and Statistical Analysis*. Amsterdam/Philadelphia: John Benjamins.
- Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge: MIT Press
- Morera, M. (1991). *Diccionario crítico de las perífrasis verbales del español*. Puerto del Rosario: Cabildo Insular de Fuerteventura.
- Perek, F., & Hilpert, M. (2017). A distributional semantic approach to the periodization of change in the productivity of constructions. *International Journal of Corpus Linguistics*, 22(4), 490-520.
- R Development Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Viena, Austria, R Foundation for Statistical Computing, 3.3.1.
- Roca-Pons, J. (1958). *Estudios sobre perífrasis verbales del español*. Madrid: C.S.I.C.
- Speece, D. (1994). "Cluster analysis in perspective", *Exceptionality*, 5(1), 31-44.

- Suttle, L., & Goldberg, A. (2011). "Corpus evidence of the viability of statistical pre-emption", *Cognitive Linguistics*, 22(1), 131-153.
- Suzuki, R., & Shimodaira, H. (2006). Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12), 1540-1542.
- Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Boston: Harvard University Press.
- Traugott, E. (2010). "Grammaticalization", *Historical pragmatics*, 4, 97-126.
- Trousdale, G. (2008). "Constructions in grammaticalization and lexicalization. Evidence from the history of a composite predicate construction in English." In Trousdale, G. y N. Gisborne (eds.). *Constructional approaches to English grammar*. Berlin: Mouton de Gruyter, 33-67.
- Trousdale, G. (2014). "On the relationship between grammaticalization and constructionalization." *Folia Linguistica*, 48(2), 557-577.
- Van Hulle, S., Enghels, R., & Jansegers, M. (in preparation). "The near-synonymy of inchoative auxiliaries in Peninsular Spanish. A Behavioral Profile Approach."

