

Heeft scheidbaarheid betekenis?

Nederlandse (on)scheidbare werkwoorden onder een computationele loep

Ann-Sophie Vrielynck (FNRS – Universiteit Luik/Antwerpen)

&

Anthe Sevenants (FWO – KU Leuven)



(On)scheidbare werkwoorden

(On)scheidbare werkwoorden

SCHEIDBAAR

Bv. Hij **belt** zijn moeder op.

- Klemtoon op partikel
- Scheiding
 - V1/V2
 - 'ge' (voltooid deelwoord)
 - 'te' (infinitiefmarkeerder)
 - Hulpwerkwoorden/modale ww.

ONSCHEIDBAAR

Bv. Hij **ondergaat** zijn pijn.

- Klemtoon op basiswerkwoord
- Geen scheiding

De betekenis van scheidbaarheid?

Booij (2002)

Grammaticalisatieproces

resultatief element \longrightarrow scheidbaar partikel \longrightarrow onscheidbaar prefix
semantische verbleking

- Link betekenis & scheidbaarheid
 - scheidbaar = concreet, spatiaal
 - onscheidbaar = abstract, aspectueel

Bv. *doorlopen*

- Iconiciteit

Deze studie

Onderzoeksvraag

- Hebben scheidbare werkwoorden een concretere betekenis dan onscheidbare werkwoorden?
- Hypothese: ja (cf. Booij 2002)

Methodologie

Stap 1: lijst werkwoorden

- Partikels/prefixen *door-*, *om-* en *over-*
 - Productief in scheidbare en onscheidbare ww.
 - Obv Van Dale (2022) en Referentiebestand Nederlands (van der Vliet 2007)

	Scheidbaar	Onscheidbaar	Totaal
<i>Door-</i>	124 (55,36%)	100 (44,64%)	224 (100%)
<i>Om-</i>	129 (64,18%)	72 (35,82%)	201 (100%)
<i>Over-</i>	189 (49,35%)	194 (50,65%)	383 (100%)
Totaal	442 (54,70%)	366 (45,30%)	808 (100%)

Stap 2: dataset Brysbaert et al. (2014)

- Obv experiment: 30 000 Nederlandse woorden beoordeeld op concreetheid (5-puntenschaal)
 - **Concreet** = "[...] verwijst naar iets wat bestaat in de realiteit. Je kunt het ervaren via één van je zintuigen of door een actie uit te voeren."
 - **Abstract** = "[...] verwijst naar iets wat je niet rechtstreeks kunt ervaren. De betekenis ervan hangt af van taal."
- Werkwoorden rechtstreeks koppelen aan dataset: problemen!
1. Heel wat werkwoorden afwezig
 2. Ww. die scheidbaar én onscheidbaar zijn: maar één keer in dataset

Alternatief: directe objecten

- Reden: proxy voor betekenis ww. (cf. Frasinelli en Schulte im Walde 2019) (Pijpops et al. 2021)
 - "Over het algemeen hebben concrete ww. voornamelijk concrete zelfstandige nw. als onderwerp en lijdend voorwerp, terwijl abstracte ww. voornamelijk abstracte zelfstandige nw. hebben als onderwerp en lijdend voorwerp." (eigen vertaling, Frasinelli en Schulte im Walde 2019: 5)
 - Voorbeelden
 - Concreet: *het bos doorlopen, de boom omhakken*
 - Abstract: *de stilte doorbreken, angst doorstaan*

Alternatief: directe objecten (2)

- **Stap 1:** corpusdata verzamelen
 - NITenTen20 corpus via SketchEngine en Sonar corpus (Oostdijk et al. 2013)
 - Voltooide deelwoorden
- **Stap 2:** automatische identificatie van het direct object (actief) of subject (passief)
 - Sonar: Alpino-parses (van Noord 2006)
 - NITenTen20: Spacy's bibliotheek *nl_core_news_sm* (Honnibal & Montani 2017)

Alternatief: directe objecten (3)

- **Stap 3:** linken aan dataset Brysbaert et al. (2014)
 - Voornaamwoorden eruit gelaten
 - Min. 10 attestaties per ww.

	Scheidbaar	Onscheidbaar	Totaal
<i>Door-</i>	52 (67,53%)	25 (32,47%)	77 (100%)
<i>Om-</i>	48 (64,86%)	26 (35,14%)	74 (100%)
<i>Over-</i>	45 (44,55%)	56 (55,45%)	101 (100%)
Totaal	144 (57,37%)	107 (42,63%)	251 (100%)

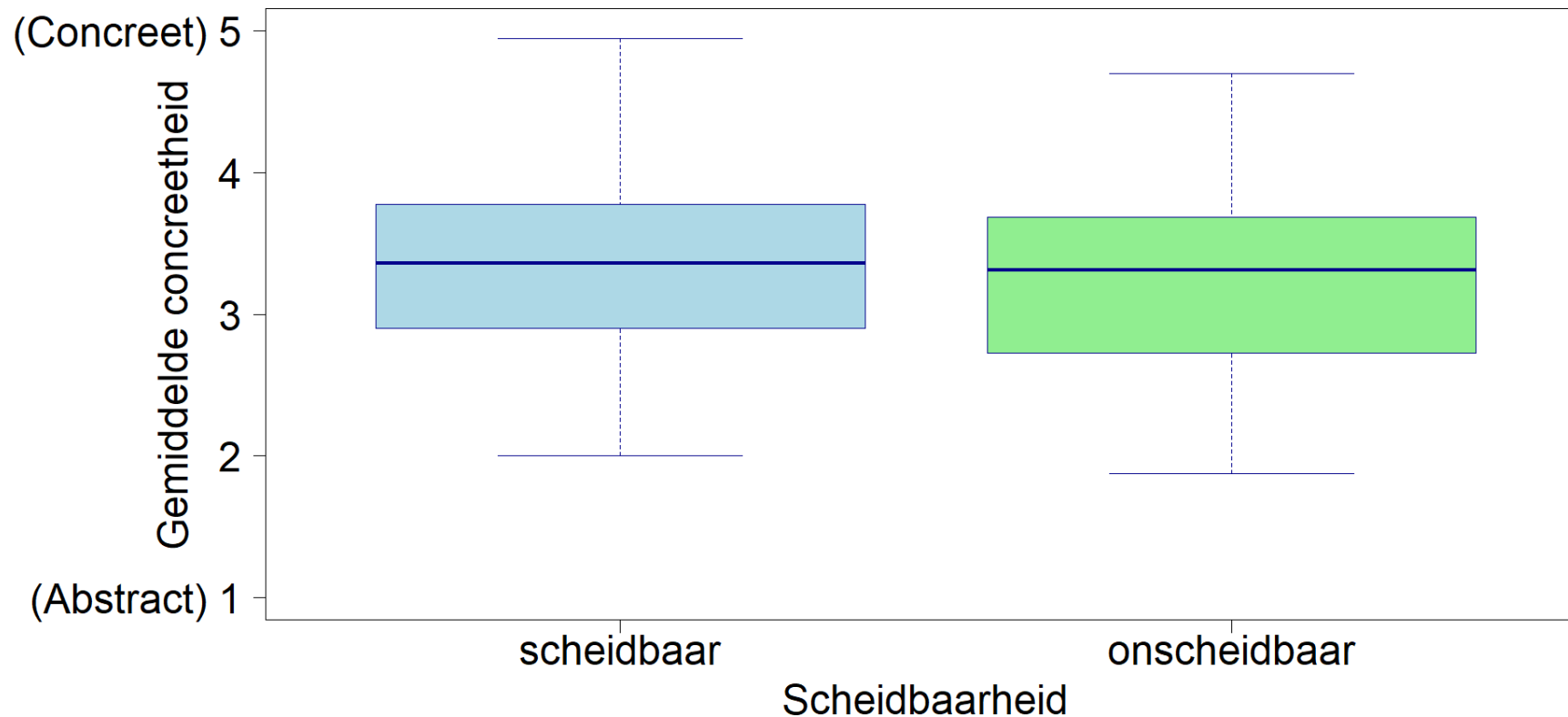
Analyse

- **Stap 4: Binomiale logistische regressie (R pakket 'glm' (R Core Team (2023)))**
 - Gemiddelde "concreetheid" per werkwoord
 - `model <- glm(Scheidbaarheid ~ Concreteness, data = data, family = binomial)`

Resultaten

Alle partikels samen

Concreetheid per scheidbaarheid (door-, om-, over-)



- Weinig verschil ($p = 0.102$)
- Gemiddelden
 - scheidbaar = 3,40
 - onscheidbaar = 3,27
- Veel variatie

Meest concrete werkwoorden

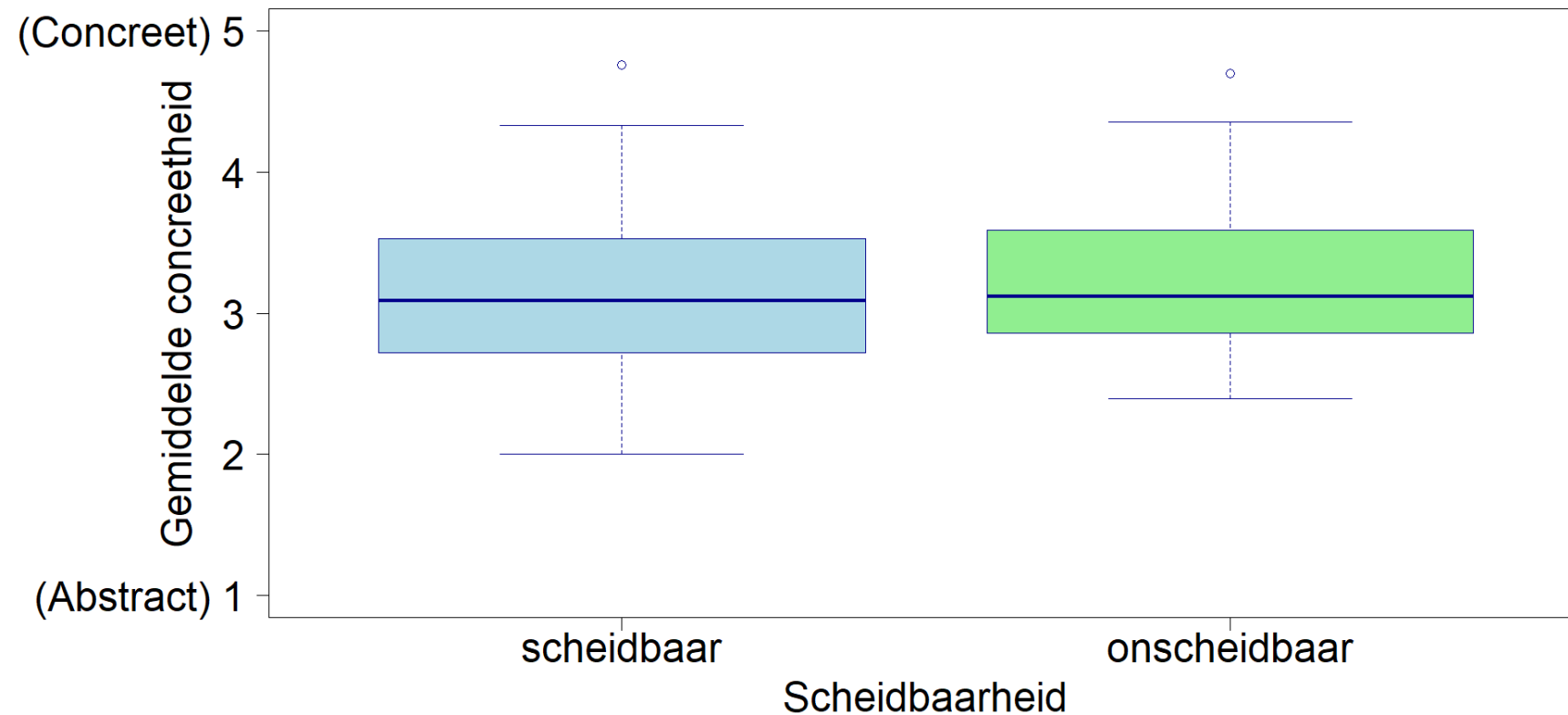
Werkwoord	Scheidbaarheid	Concreetheid	Lemma
Omgedaan	Scheidbaar	4,95	Das
Doorgehakt	Scheidbaar	4,76	Knoop
Doorzwommen	Onscheidbaar	4,70	Water
Omgeplooid	Scheidbaar	4,65	Bord
Omgekapt	Scheidbaar	4,64	Boom
Overspeeld	Onscheidbaar	4,53	Hand
Omgezaagd	Scheidbaar	4,52	Boom
Omgehakt	Scheidbaar	4,51	Boom

Meest abstracte werkwoorden

Werkwoord	Scheidbaarheid	Concreetheid	Lemma
Overtreden	Onscheidbaar	1,88	Wet
Doorgehad	Scheidbaar	2	Opzet
Omgewrongen	Scheidbaar	2,02	Kans
Doorgedreven	Scheidbaar	2,21	Zin
Overschat	Onscheidbaar	2,22	Belang
Doorgemaakt	Scheidbaar	2,22	Ontwikkeling
Overtroffen	Onscheidbaar	2,26	Verwachting
Overroepen	Onscheidbaar	2,26	Probleem

Door-

Concreetheid per scheidbaarheid (door-)



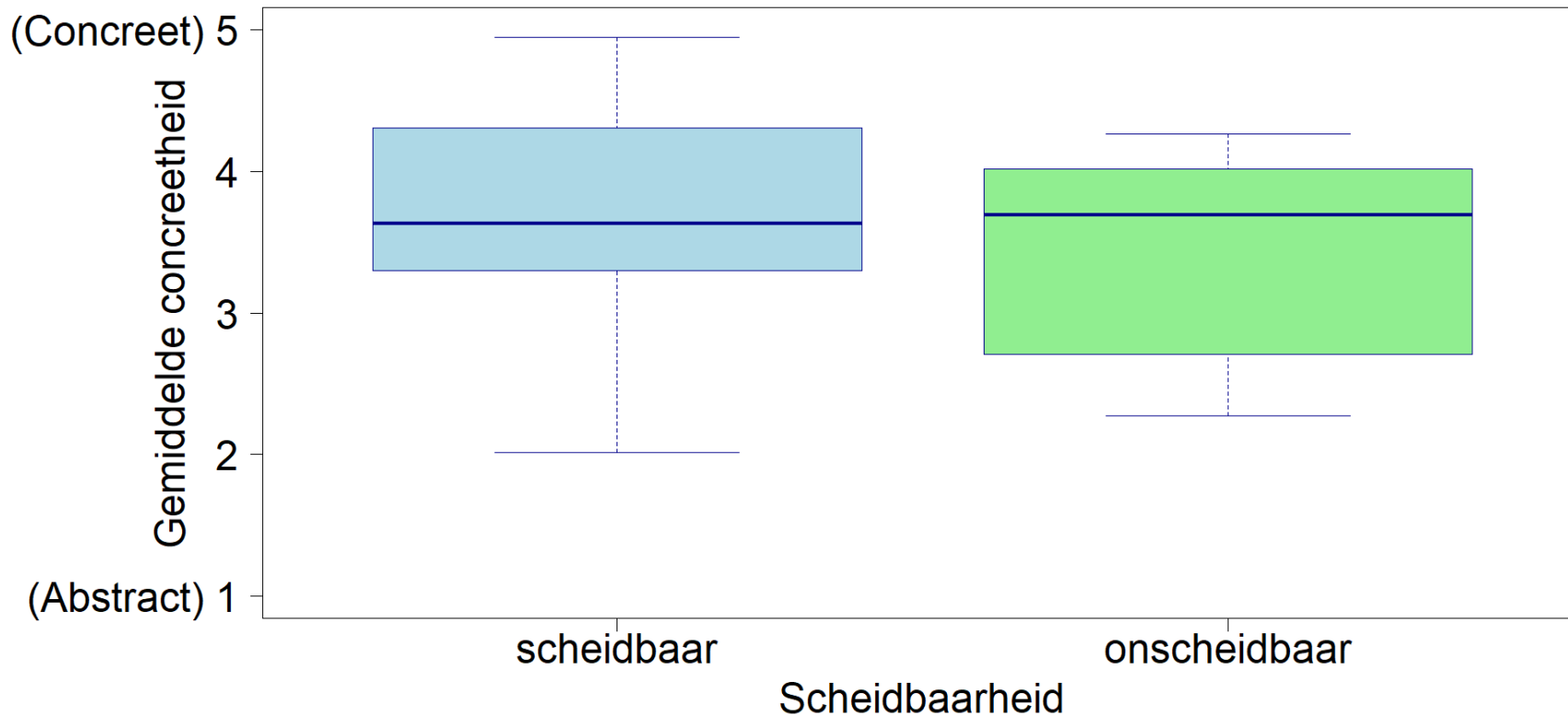
➤ Weinig verschil
($p = 0.27$)

➤ Gemiddelden

- scheidbaar = 3,14
- onscheidbaar = 3,31(!)

Om-

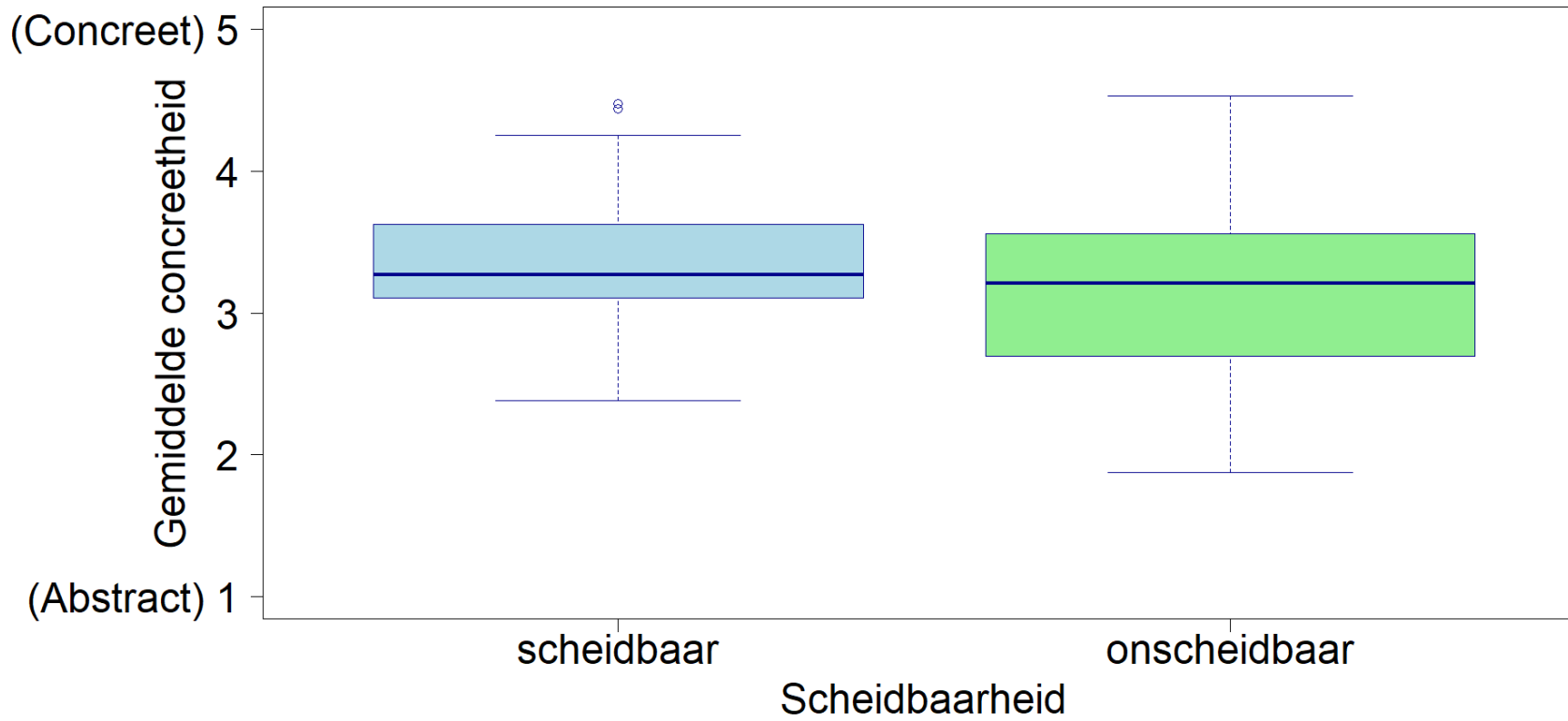
Concreetheid per scheidbaarheid (om-)



- Opnieuw weinig verschil ($p = 0.12$)
- Gemiddelden
 - scheidbaar = 3,71
 - onscheidbaar = 3,46
- Meer variatie bij scheidbaar

Over-

Concreetheid per scheidbaarheid (over-)

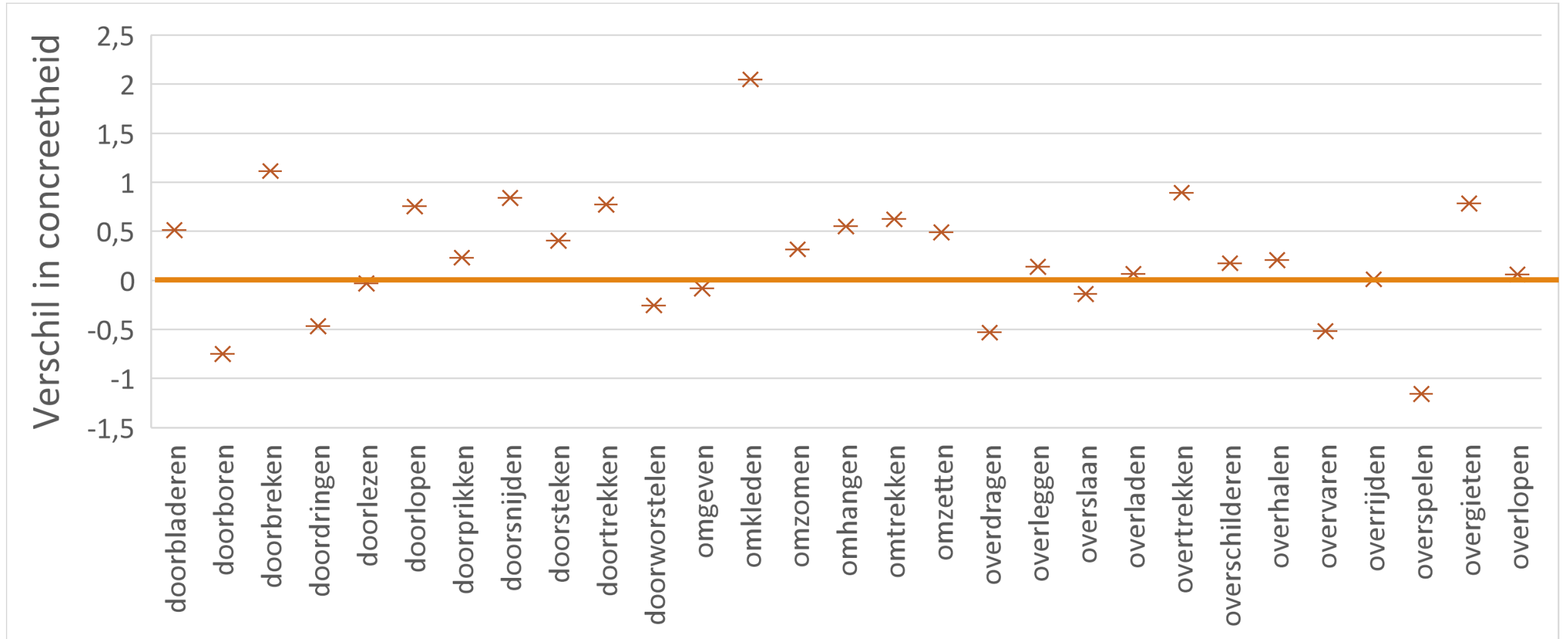


- Opnieuw weinig verschil ($p = 0.07$)
- Gemiddelden
 - scheidbaar = 3,36
 - onscheidbaar = 3,16
- Meer variatie bij onscheidbaar

Interim conclusie

- Link betekenis & scheidbaarheid (cf. Booij 2002)
 - scheidbaar = concreet, spaciaal
 - onscheidbaar = abstract, aspectueel
- Verschillen tussen partikels

Werkwoorden die scheidbaar én onscheidbaar zijn



Ander semantisch verschil?

HYPOTHESEGENERATIE

1) Cornetto

- **Stap 1:** Werkwoorden met *door-*, *om-* en *over-* gekoppeld aan Cornetto (Vossen et al. 2013) (Bedankt Anthe voor de Python-code!)
 - Semantische informatie: Sense, SemanticType, SemanticFeatureSet
- **Stap 2:** Bij polyseme werkwoorden -> obv DutchSemCor (Vossen et al. 2011) meest frequente betekenis geselecteerd
- **Stap 3:** binomiale logistische regressie
 - Geen semantische factoren significant
 - **Transitiviteit** is significant ($p = 0.0281$) => onscheidbaar is significant vaker transitief dan scheidbaar

2) Resultativiteit?

- cf. Blom (2004, 2005): tegenvoorbeelden (inlopen, opzoeken vs. doorzoeken, omgeven)
- link betekenis & scheidbaarheid
 - Resultatief = scheidbaar
 - Bv. *Hij **zaagt** de boom **om**.* / *De zon **gaat onder**.*
 - Niet-resultatief = (on)scheidbaar
 - Bv. *Hij **ondergaat** de pijn.* / *Ze **kijkt** de jongen **aan**.*
- Obv Van Dale 1996
- Vervolgonderzoek: experimenteel testen van deze factor bij non-sense werkwoorden (forced-choice task)

Referenties

Blom, Corrien (2004). On the diachrony of complex predicates in Dutch: predicative and nonpredicative preverbs. *Journal of Germanic Linguistics* 16. 1–75.

Blom, Corrien (2005). *Complex predicates in Dutch: synchrony and diachrony*. Utrecht: Netherlands Graduate School of Linguistics.

Booij, Geert (2002). *The morphology of Dutch*. Oxford: Oxford University Press.

Brysbart, Marc, Stevens, Michaël, De Deyne, Simon, Voorspoels, Wouter & Gert Storms (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica* 150. 80–84.

Van Dale groot woordenboek van de Nederlandse taal (2022). Utrecht: Van Dale Uitgevers.

Frassinelli, Diego & Sabine Schulte im Walde (2019). Distributional interaction of concreteness and abstractness in verb-noun subcategorisation. *Proceedings of the 13th International Conference on Computational Semantics – Short Papers*. 38–43.

Honnibal, Matthew & Ines Montani (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*.

[nlTenTen20] Corpus of the Dutch Web 2020, *Sketch Engine*. <https://www.sketchengine.eu/nltenten-dutch-corpus/> (31 mei, 2024)

van Noord, Gertjan. (2006). At Last Parsing Is Now Operational. In Piet Mertens, Cédric Fairon, Anne Dister & Patrick Watrin (eds.), TALN 2006. *Verbum Ex Machina. Actes de la 13e conférence sur le traitement automatique des langues naturelles*. 20–42. Louvain-la-Neuve: Cental.)

Referenties (2)

Oostdijk, Nelleke, Reynaert, Martin, Hoste, Véronique & Ineke Schuurman. (2013). The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch. In Peter Spyns & Jan Odijk (eds.), *Essential Speech and Language Technology for Dutch, Theory and Applications of Natural Language Processing*. 219–247. Heidelberg: Springer.

Pijpops, Dirk, Speelman, Dirk, Grondelaers, Stef & Freek Van de Velde. 2021. Incorporating the multi-level nature of the constructicon into hypothesis testing. *Cognitive Linguistics* 32(3). 487–528.

R Core Team (2023). R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.

SoNaR-corpus (Version 1.2.1) (2015) [Data set]. Beschikbaar bij het Instituut voor de Nederlandse taal: <http://hdl.handle.net/10032/tm-a2-h5>

van der Vliet, Hennie (2007). The Referentieb Bestand Nederlands as a Multi-Purpose Lexical Database. *International Journal of Lexicography* 20(3). 239–257.

Vossen, Piek, Görög, Attila, Laan, Fons, van Gompel, Maarten, Izquierdo Beviá, Ruben & Antal van den Bosch (2011). DutchSemCor: Building a semantically annotated corpus for Dutch. *Proceedings of eLex 2011*. 286–296.

Vossen, Piek, Maks, Isa, Segers, Roxane, van der Vliet, Hennie, Moens, Marie-Francine, Hoffmann, Katja, Tjong Kim Sang, Erik & Maarten de Rijke (2013). Cornetto: a lexical semantic database for Dutch. In Peter Spyns & Jan Odijk (eds.), *Essential Speech and Language Technology for Dutch: Results by the STEVIN Programme*. 165–184.