

# Why and how we need to incorporate the multi-level nature of the construction into corpus research

Dirk Pijpops, Dirk Spielman, Stefan Grondelaers, Freek Van de Velde

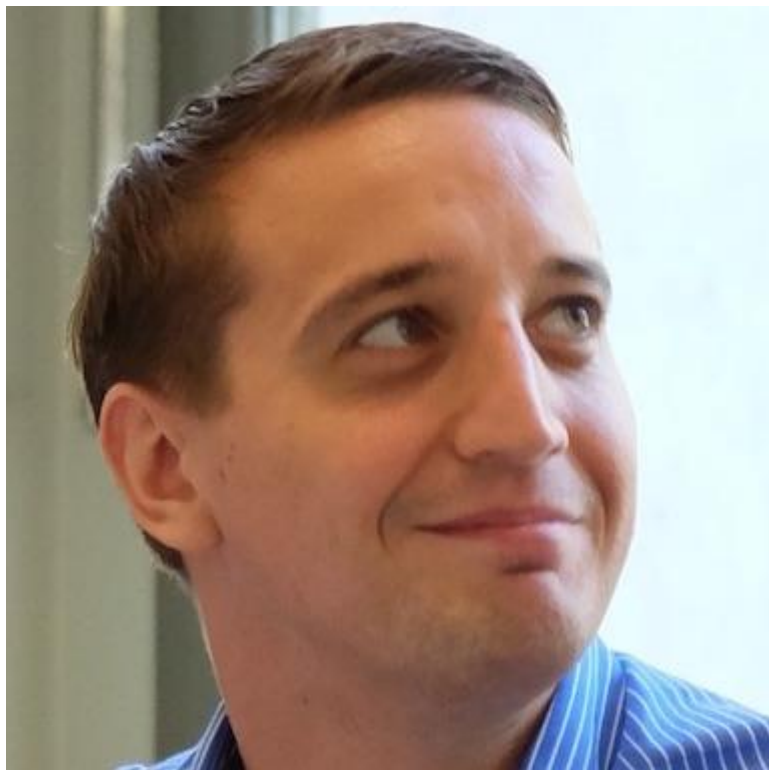
Research Foundation Flanders (FWO)

University of Leuven

Radboud University of Nijmegen

# 3 Problems in Construction Grammar

Drawing from work of



Florent Perek



Ewa Dąbrowska



Hans Boas



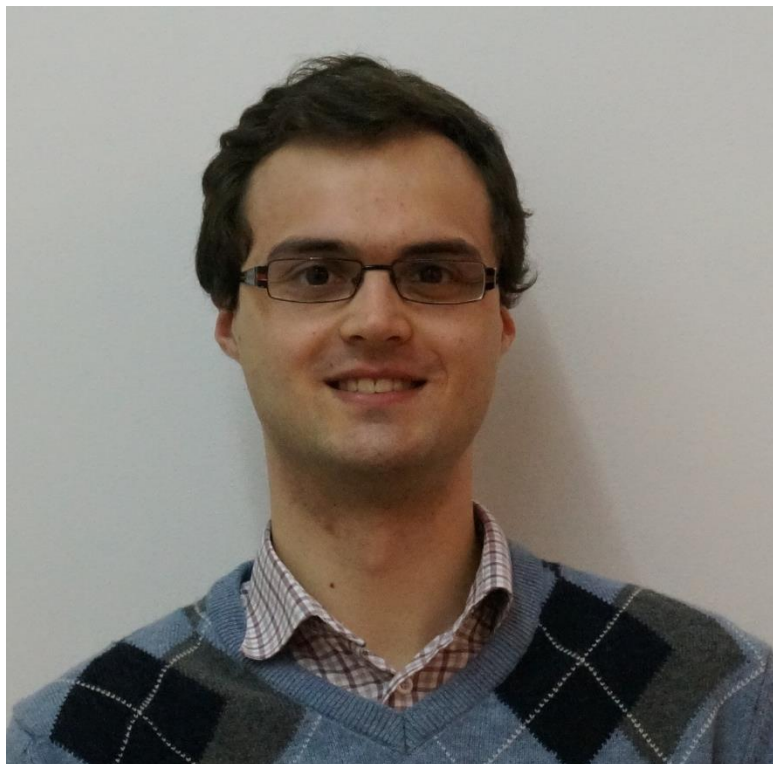
Adele Goldberg

Thanks to





Katrien Beuls



Paul Van Eecke



Fluid Construction Grammar Team

# 3 Problems in Construction Grammar

Transitive construction

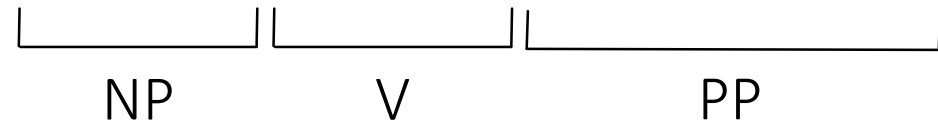
vs.

*naar*-construction



vs.

*naar*



'People desire certainty'

Transitive construction

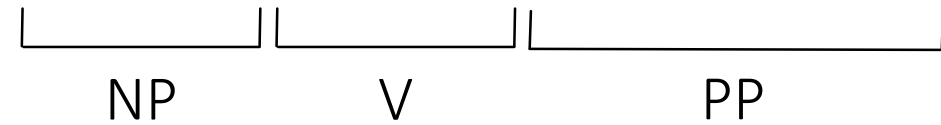
vs.

*naar*-construction

*Ze zoeken slachtoffers*

vs.

*Ze zoeken **naar** slachtoffers*



'They are searching victims'

- **Principle of Semantic Coherence:** verbs that are more semantically similar to an argument construction, will more often appear in that construction
  1. Meanings of the transitive construction and the *naar*-construction?
  2. For each alternating verb: semantic distance to both constructions?
  3. Test the prediction, while controlling for other factors
  
- **Lexical Origin Hypothesis:** abstract argument constructions extract their meaning from the prototypical fillers of their verb slot, e.g. *give* for the English ditransitive construction
  - ⇒ Identify the **prototypical fillers** of the verb slots
  - ⇒ **Collostructional analyses**

### Transitive construction

Frequency transitive construction	Total Frequency	Collostruct. Strength
-----------------------------------	-----------------	-----------------------

<i>doen</i> 'do'	553,960	668,486	10,795,308
<i>krijgen</i> 'get'	511,965	586,452	10,562,871
<i>hebben</i> 'have'	1,280,142	3,396,725	10,063,481
<i>maken</i> 'make'	506,522	644,692	9,306,076

⇒ *doen* 'do' as proxy for the meaning of the transitive construction

### *Naar*-construction

Frequency naar-construction	Total Frequency	Collostruct. Strength
-----------------------------	-----------------	-----------------------

<i>verwijzen</i> 'refer'	23,063	23,787	12,588,214
<i>kijken</i> 'look'	53,954	130,567	12,492,364
<i>gaan</i> 'go'	113,774	953,918	7,456,440
<i>sturen</i> 'send'	24,501	53,963	6,236,835

⇒ *kijken* 'look' as proxy for the meaning of the *naar*-construction



- **Principle of Semantic Coherence:** verbs that are more semantically similar to an argument construction, will more often appear in that construction
1. Meanings of the transitive construction and the *naar*-construction?
  2. For each alternating verb: semantic distance to both constructions?
  3. Test the prediction, while controlling for other factors



⇒ **Distributional Vectors**

*grabbelen* 'scramble'



*graaien* 'grobe'



*happen* 'snap'



*zoeken* 'search'



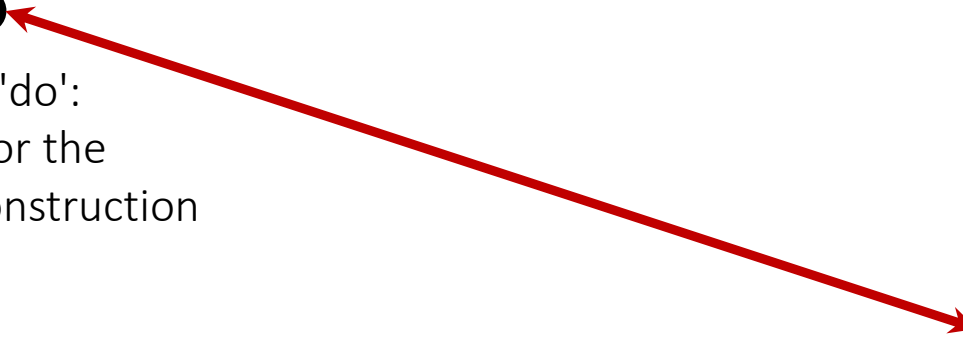
*verlangen* 'desire'



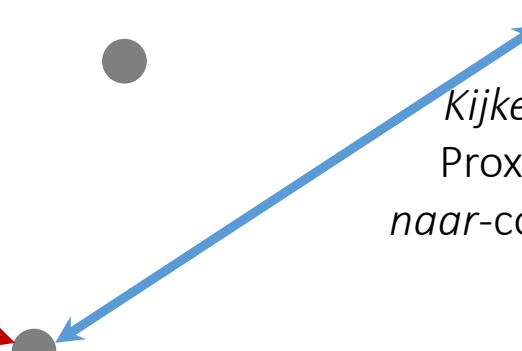
*peilen* 'gauge'



*Doen* 'do':  
Proxy for the  
transitive construction



*Kijken* 'look':  
Proxy for the  
*naar*-construction



- **Principle of Semantic Coherence:** verbs that are more semantically similar to an argument construction, will more often appear in that construction
  1. Meanings of the transitive construction and the *naar*-construction?
  2. **For each alternating verb: semantic distance to both constructions?**
  3. Test the prediction, while controlling for other factors



⇒ **Distributional Vectors**

$$\text{Relative-Similarity-to-Look} = \left( \text{sim}_{cm}(\overrightarrow{kijken}, \overrightarrow{verb}) - \text{sim}_{cm}(\overrightarrow{doen}, \overrightarrow{verb}) \right) 10$$

- **Principle of Semantic Coherence:** verbs that are more semantically similar to an argument construction, will more often appear in that construction
  1. Meanings of the transitive construction and the *naar*-construction?
  2. For each alternating verb: semantic distance to both constructions?
  3. **Test the prediction, while controlling for other factors**



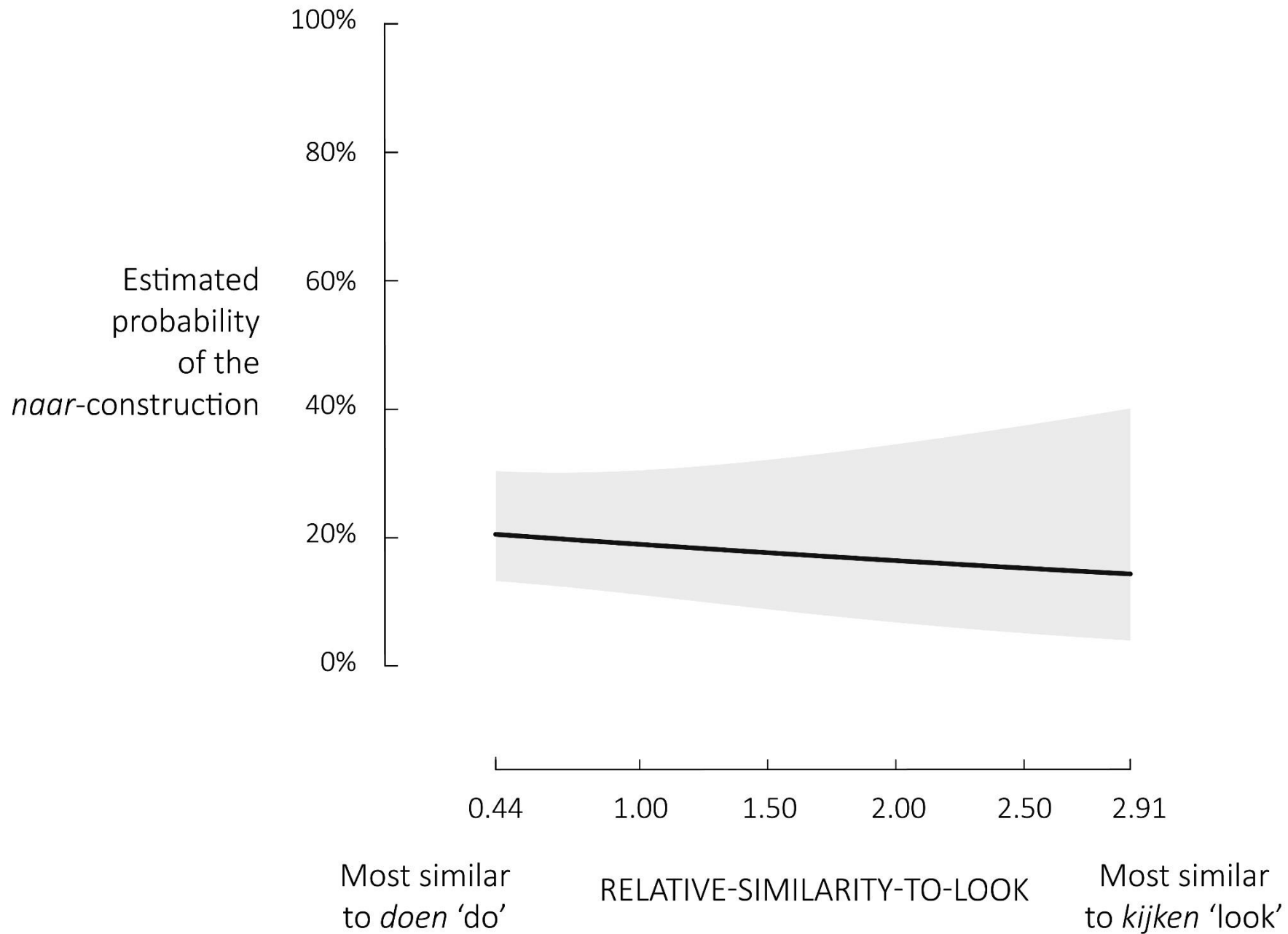
⇒ **Mixed Regression Modelling**

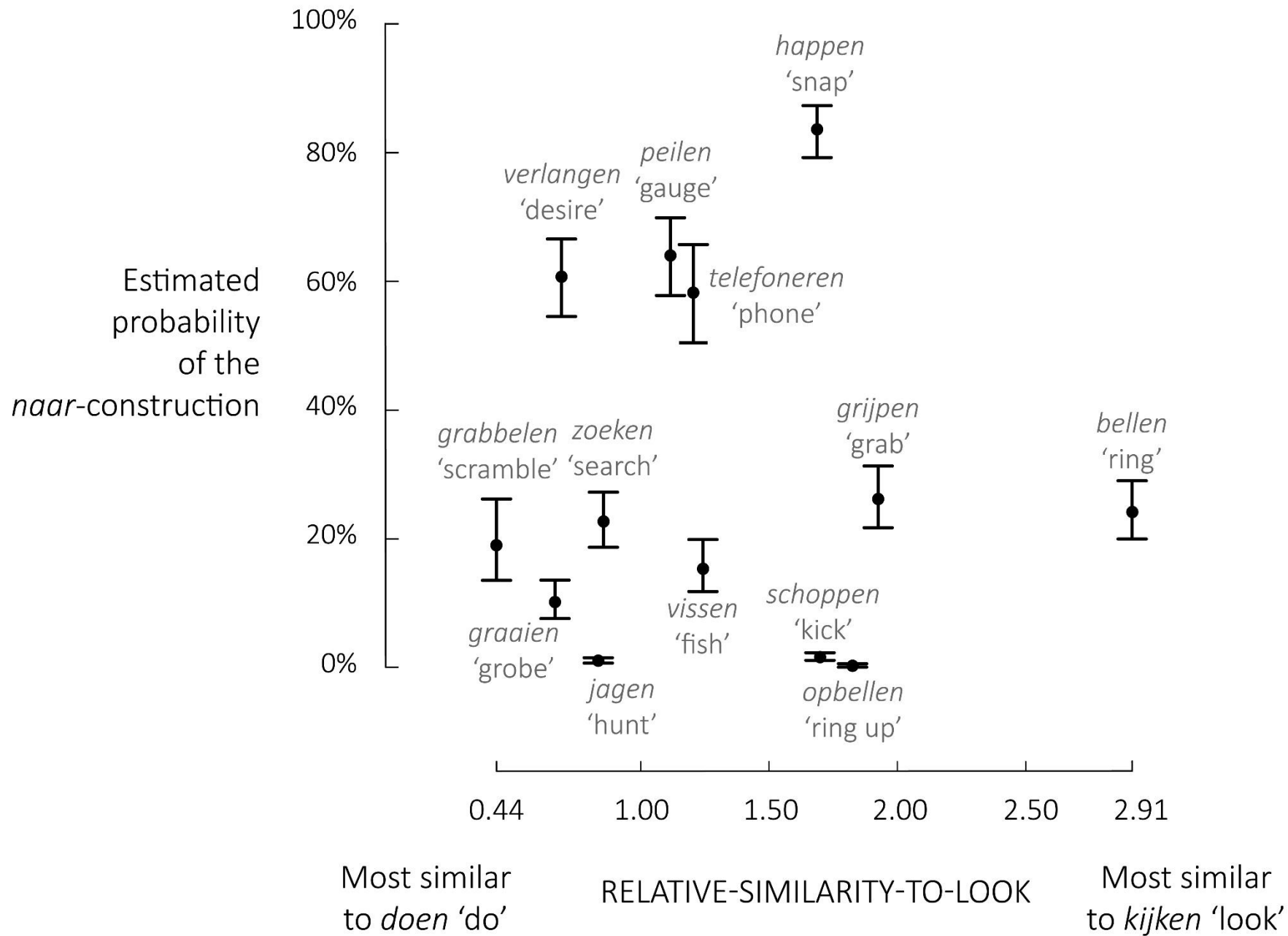
*Argument Construction* ~ *Relative-Similarity-to-Look*

+ *Country* + *Theme-Complexity* + *Verb-Theme-Order*

+ *Theme-Complexity:Verb-Theme-Order*

+ *1|Component* + *1|Verb*





Prediction Failed

# Problem of Prediction



# Problem of Prediction

How to predict concrete meaning distinctions from abstract constructions?

Lenci (2012: 13-15): *decidere (sull')*, *rimproverare per/a*

Perek (2015: 90-144): *chip (at)*, *kick (at)*

Dąbrowska (2017: 21-40)

⇒ *Use concrete constructions, not abstract constructions*

# Problem of Proliferation

# Problem of Proliferation

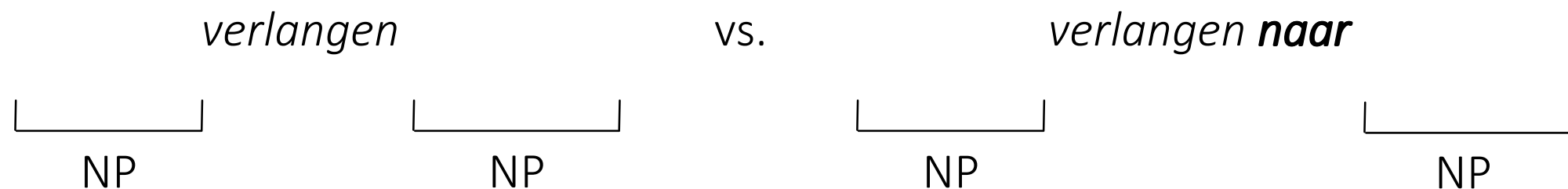
Does a proliferation of concrete constructions decrease the parsimony of construction grammar?

⇒ *No. Measure parsimony in number of mechanisms*

# Problem of Proliferation

What mechanism can explain the emergence of the constructional meaning of concrete constructions?

- ⇒ **Lexical Origin Hypothesis:** argument constructions extract their meaning from the prototypical fillers of their verb slot
- ⇒ What if the verb slot is fixed?



# Problem of Precedence

# Problem of Precedence

At what level in the construction does an alternation express meaning?

- **Problem of Prediction:** Use this level to make successful predictions
- **Problem of Proliferation:** Prune all other levels from the network and explain the emergence of constructional meaning at this level

# Problem of Precedence

At what level in the construction does an alternation express meaning?

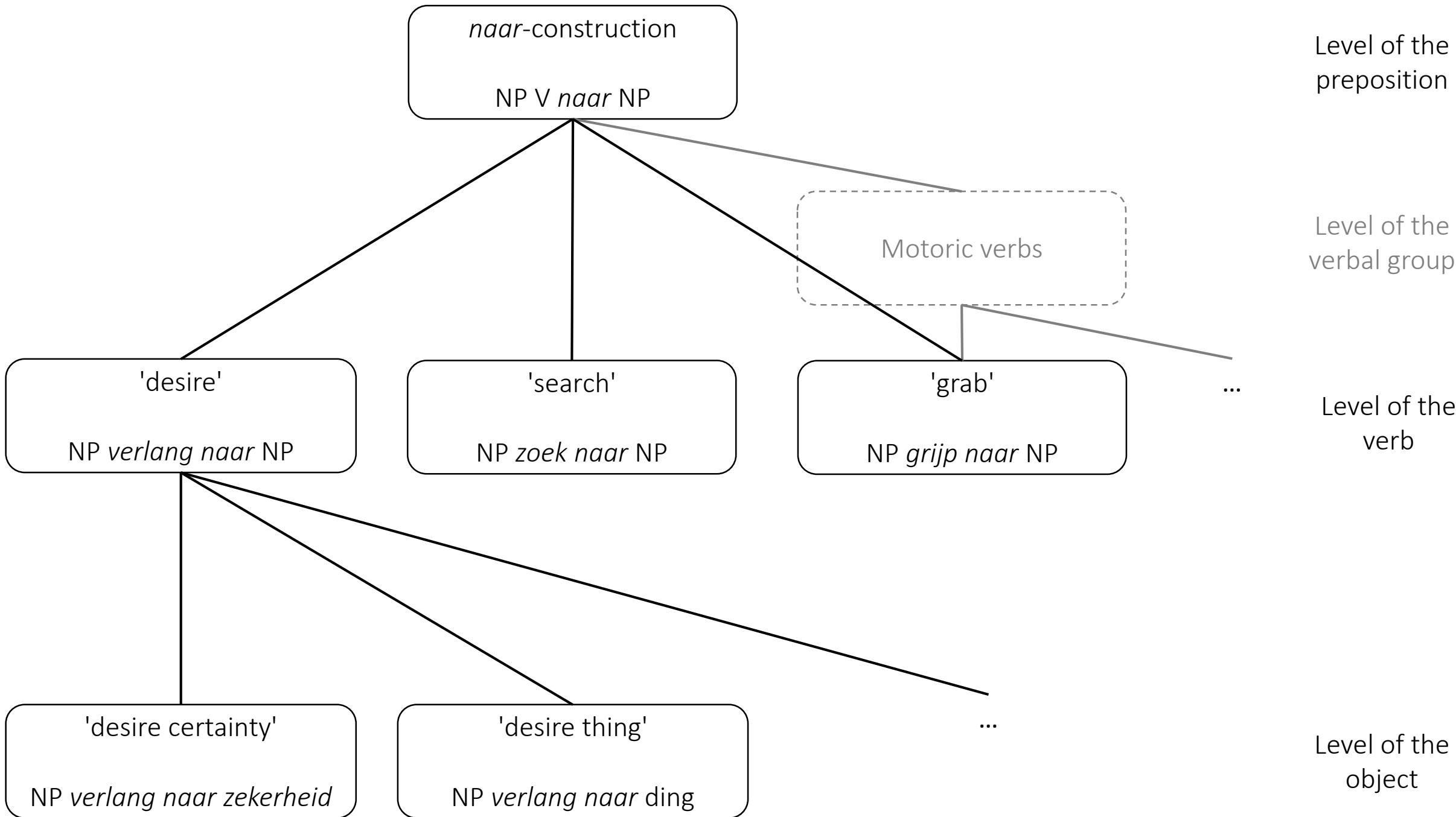
⇒ *Treat this as an empirical question*

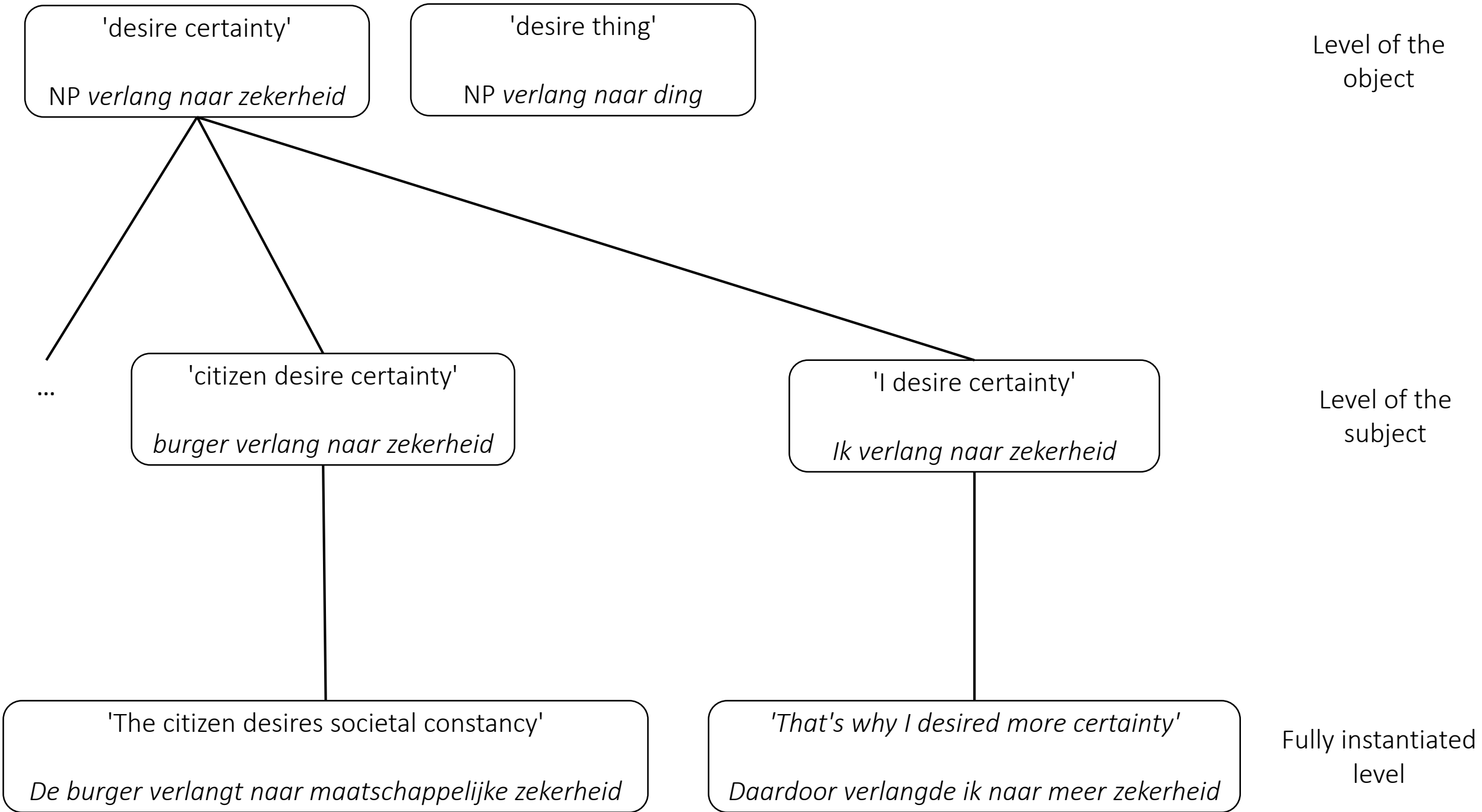
⇒ *Develop a methodological approach to answer this question*

# How?

1. Map out the constructional network
2. Traverse the network







*naar*-construction

NP V *naar* NP

Level of the  
preposition

Prediction Failed

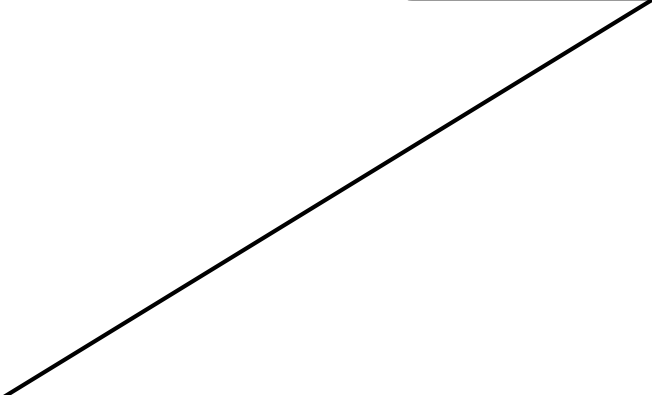
Descend one level

*naar*-construction  
NP V *naar* NP

Level of the  
preposition

'desire'  
NP *verlang naar* NP

Level of the  
verb



*verlangen*



NP



NP

vs.

*verlangen naar*



NP



NP

# Level of the verb

1. Identify the prototypical fillers of the underlying slot: **Collostructional Analyses**
2. Measure the semantic distance to these prototypical fillers: **Distributional Vectors**
3. Test the prediction, while controlling for other factors: **Mixed Regression Modelling**



### Transitive *verlang*-construction

Frequency transitive construction	Total Frequency	Collostruct. Strength
-----------------------------------	-----------------	-----------------------

<i>bewijs</i> 'proof'	18	24,734	32.00
<i>tegenprestatie</i> 'countereffort'	10	901	29.93
<i>offer</i> 'sacrifice'	10	3,122	24.52
<i>bijdrage</i> 'contribution'	14	26,136	23.21
<i>garantie</i> 'garantuee'	12	12,060	23.20
<i>teken</i> 'sign'	12	15,962	21.75

⇒ *bewijs* 'proof' as proxy  
desire as 'demand'

### *verlang-naar*-construction

Frequency naar-construction	Total Frequency	Collostruct. Strength
-----------------------------	-----------------	-----------------------

<i>leven</i> 'life'	35	169,360	43.60
<i>rust</i> 'rest'	26	52,595	42.31
<i>kind</i> 'child'	38	251,250	42.25
<i>dood</i> 'death'	24	51,814	38.44
<i>huis</i> 'house'	26	145,482	31.00
<i>bed</i> 'bed'	14	26,942	23.44

⇒ *leven* 'life' as proxy  
desire as 'long for'

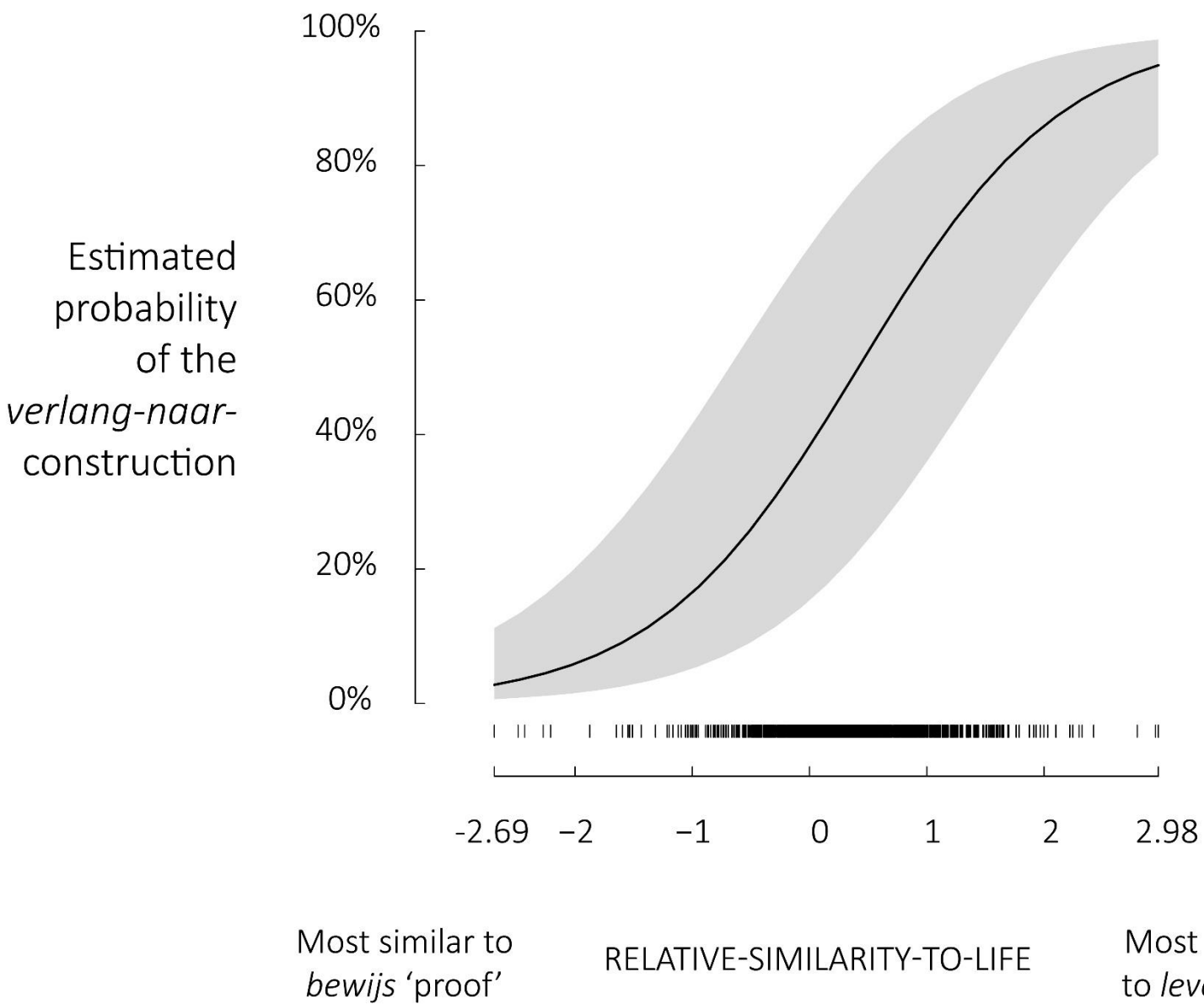
# Level of the verb

1. Identify the prototypical fillers of the underlying slot: *bewijs* 'proof' vs. *leven* 'life'

2. Measure the semantic distance to these prototypical fillers: **Distributional Vectors**

$$\text{Relative-Similarity-to-Life} = \left( \text{sim}_{cm}(\overrightarrow{\text{leven}}, \overrightarrow{\text{object}}) - \text{sim}_{cm}(\overrightarrow{\text{bewijs}}, \overrightarrow{\text{object}}) \right) 10$$

3. Test the prediction, while controlling for other factors: **Mixed Regression Modelling**



*Argument Construction* ~ ***Relative-Similarity-to-Life***

+ *Relative-Similarity-to-House*  
 + *Country + Theme-Complexity*  
 + *Verb-Theme-Order*  
 + *Theme-Complexity:Verb-Theme-Order*  
 + *1|Component + 1|Object-lemma*

- Only full nominal theme lemma's
- *Leven* 'life', *bewijs* 'proof', *huis* 'house', *kans* 'chance' excluded to avoid circularity

Prediction Successful

Precedence level?

Descend one level

*naar*-construction  
NP V *naar* NP

Level of the  
preposition

'desire'  
NP *verlang naar* NP

Level of the  
verb

'desire certainty'  
NP *verlang naar zekerheid*

'desire thing'  
NP *verlang naar ding*

Level of the  
object

# Level of the object

- Use the higher level distinction 'demand' vs. 'long for' to predict the variation at this level
  - *Verlang ding* 'desire thing': 9 transitive instances vs. 12 prepositional instances
  - *Verlang zekerheid* 'desire certainty': 7 transitive instances vs. 5 prepositional instances



### Transitive: predicted 'demand'

*De Raad en het Parlement kunnen natuurlijk niet van de Commissie dingen verlangen waarvoor zij haar geen bevoegdheid verlenen.* (WR-P-P-I-0000000069.p.161.s.1)

‘The Council and the Parliament of course cannot demand things from the Commission for which they do not grant it the authority.’

*Straks gaat hij weer dingen van mij verlangen.* (WR-P-E-G-0000008459.p.377.s.1)

‘He’ll be demanding things from me again.’

### Prepositional: predicted 'long for'

*Het soort liefde dat ik voor haar voelde, deed haar **naar** nieuwe dingen verlangen.*

‘The kind of love that I felt for her, made her long for new things’ (WR-P-P-B-0000000132.p.1726.s.15)

*Stelt u zich eens voor dat u **naar** al die dingen vurig zou verlangen...* (WR-P-P-B-0000000056.p.1379.s.2)

‘Imagine yourself that you are passionately longing for all those things...’

### **Transitive: predicted 'demand'**

*In amendement 2 staat dat de douaneautoriteiten een zekerheid kunnen verlangen om hun aansprakelijkheid (...) te dekken.*

(WR-P-P-I-0000000312.p.1169.s.14)

'In amendment 2, it is stated that the customs authorities can demand an assurance that their liability (...) is covered.'

*Terecht wordt door de gemeenteraad zekerheid verlangd dat de met de aanleg gemoeide kosten binnen de perken blijven.*

(WR-P-P-G-0000003464.p.1.s.3)

'The town council rightly demands the assurance that the costs involved in the construction work remain within certain limits.'

### **Prepositional: predicted 'long for'**

*Zelf hebt u ook hevig verlangd naar onwrikbare zekerheden.*

(WR-P-P-G-0000213898.p.24.s.1)

'You have also yourself longed for unshakeable certainties.'

*Men verlangt naar zekerheid, naar houvast.*

(WR-P-P-H-0000097808.p.22.s.4)

'One longs for security, for something to hold on to.'

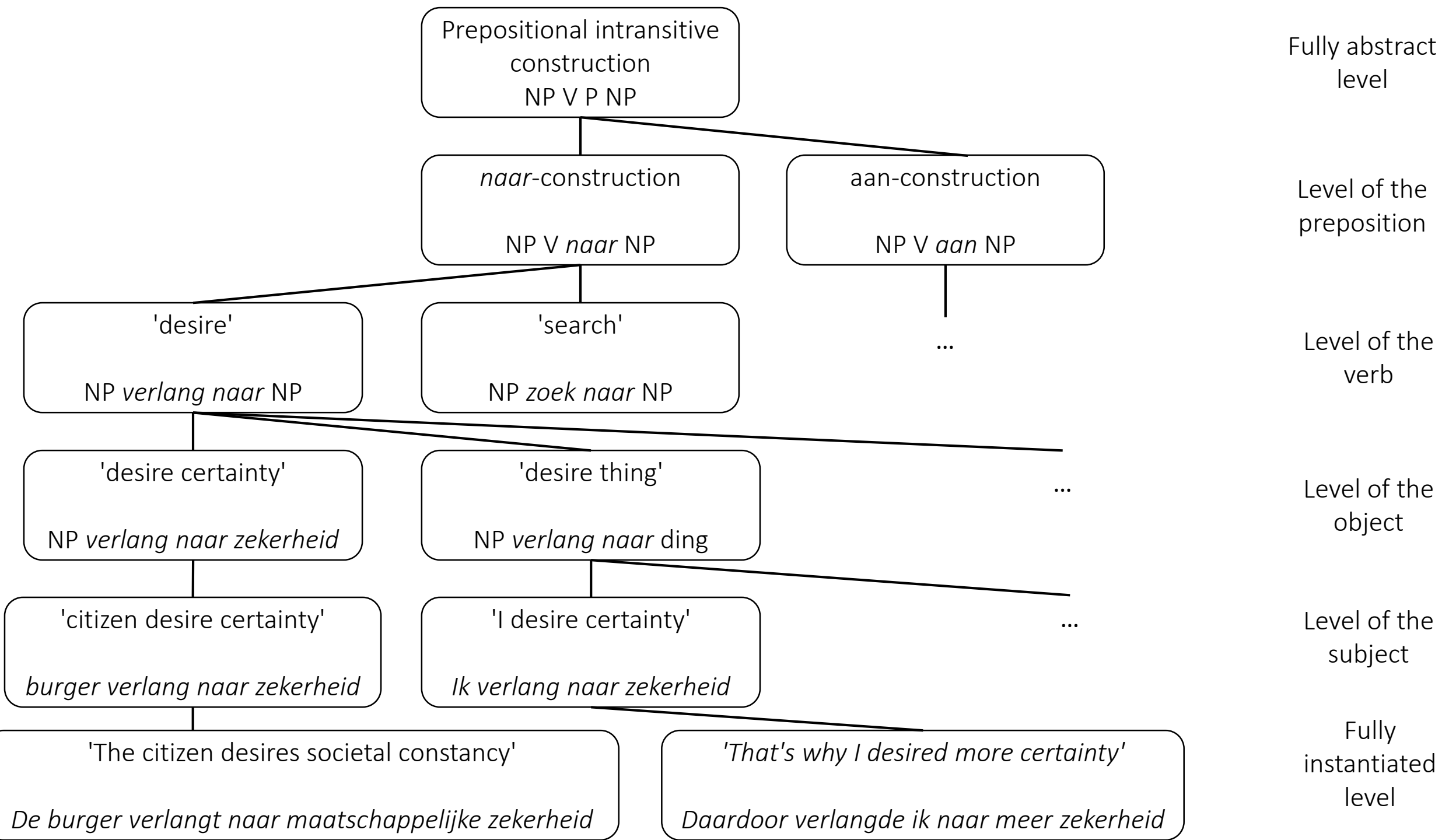
Prediction Successful

# How?

1. Map out the constructional network
2. Traverse the network

# Why?

- **Problem of Precedence:** At what level in the construction does an alternation express meaning?
  - ⇒ Identify this precedence level by **incorporating the multi-level nature** of the construction into corpus research. For this case study: the level of the verb.
- **Problem of Prediction:** How to predict concrete meaning distinctions from abstract constructions?
  - ⇒ Use the **precedence level**
- **Problem of Proliferation:** Does a proliferation of concrete constructions decrease the parsimony of construction grammar?
  - ⇒ **Prune all other levels** from the network



Prepositional intransitive construction  
NP V P NP

Fully abstract level

Level of the preposition

'desire'  
NP *verlang naar* NP

'search'  
NP *zoek naar* NP

Level of the verb

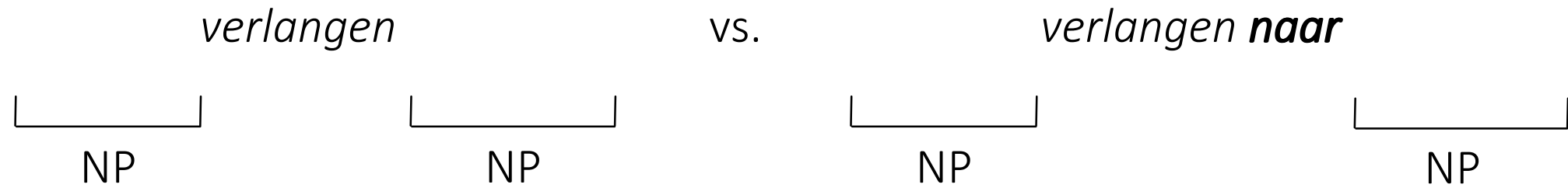
Level of the object

Level of the subject

'The citizen desires societal constancy'  
*De burger verlangt naar maatschappelijke zekerheid*

*'That's why I desired more certainty'*  
*Daardoor verlangde ik naar meer zekerheid*

Fully instantiated level



**Lexical Origin Hypothesis:** argument constructions at any level of abstraction extract their meaning from the prototypical fillers of their underlying slot



# Why?

- **Problem of Precedence:** At what level in the construction does an alternation express meaning?
  - ⇒ Identify this precedence level by **incorporating the multi-level nature** of the construction into corpus research. For this case study: the level of the verb.
- **Problem of Prediction:** How to predict concrete meaning distinctions from abstract constructions?
  - ⇒ Use the **precedence level**
- **Problem of Proliferation:** Does a proliferation of concrete constructions decrease the parsimony of construction grammar?
  - ⇒ **Prune all other levels** from the network
  - ⇒ Account for the **emergence of constructional meaning at this level.**

# Thanks!

Dirk Pijpops, Dirk Spielman, Stefan Grondelaers, Freek Van de Velde

Research Foundation Flanders (FWO)

University of Leuven

Radboud University of Nijmegen

# References

- Bates, Douglas, Martin Maechler, Ben Bolker and Steven Walker. 2013. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.4.
- Boas, Hans. 2008. Determining the structure of lexical entries and grammatical constructions in Construction Grammar. *Annual Review of Cognitive Linguistics* 6(1). 113–144.
- Boas, Hans. 2010. The syntax-lexicon continuum in Construction Grammar. A Case study of English communication verbs. *Belgian Journal of Linguistics* 24(1). 54–82.
- Boas, Hans. 2014. Lexical and phrasal approaches to argument structure: Two sides of the same coin. *Theoretical Linguistics* 40(1–2). 89–112.
- Boyd, Jeremy, Erin Gottschalk and Adele Eva Goldberg. 2009. Linking Rule Acquisition in Novel Phrasal Constructions. *Language Learning* 59(s1). Malden, USA: Blackwell Publishing Inc. 64–89.
- Broccias, Cristiano. 2001. Allative and ablative at-constructions. In Mary Adronis, Christopher Ball, Elston Heide & Sylvain Neuvel (eds.), *CLS 37: The Main Session. Papers from the 37th Meeting of the Chicago Linguistic Society*, 67–82. Chicago: Chicago Linguistic Society.
- Broekhuis, Hans. 2004. Het voorzetelvoorwerp. *Nederlandse Taalkunde* 9. 31–97.
- Cappelle, Bert. 2006. Particle placement and the case for “allostructions.” In D. Schönefeld (ed.), *Constructions Special Volume 1 — Constructions all over: case studies and theoretical implications*, vol. 4.
- Croft, William. 2003. Lexical rules vs. constructions. A false dichotomy. In Hubert Cuyckens, Thomas Berg, René Dirven & Klaus-Uwe Panther (eds.), *Motivation in language: studies in honor of Günter Radden*, 49–68. Stanford: CSLI Publications.
- Croft, William. 2012. *Verbs: aspect and causal structure*. (Ed.) William Croft. Oxford: Oxford university press.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.
- Dunn, Jonathan. 2017. Learnability and falsifiability of Construction Grammars. *Proceedings of the Linguistic Society of America*, vol. 2, 1–15.
- Forsberg, M, R Johansson, L Bäckström, L Borin, B Lyngfelt, J Olofsson and J Prentice. 2014. From construction candidates to construction entries: An experiment using semi-automatic methods for identifying constructions in corpora. *Constructions and Frames* 6(1). John Benjamins Publishing Company. 114–135.
- Fox, John, Sanford Weisberg, Michael Friendly, Jangman Hong, Robert Andersen, David Firth and Steve Taylor. 2016. Effect Displays for Linear, Generalized Linear, and Other Models. R package version 3.2.
- Goldberg, Adele Eva. 1992. The inherent semantics of argument structure: the case of the English ditransitive. *Cognitive Linguistics* 3. 37–74.
- Goldberg, Adele Eva. 1995. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago press.
- Goldberg, Adele Eva. 1999. The emergence of the semantics of argument structure constructions. In Brian MacWhinney (ed.), *Emergence of Language*, 197–212. Hillsdale: Lawrence Erlbaum Associates.
- Goldberg, Adele Eva. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*.
- Goldberg, Adele Eva. 2005. Argument realization. The role of constructions, lexical semantics and discourse factors. *Construction grammars: cognitive grounding and theoretical extensions*, 17–43. Amsterdam/Philadelphia: John Benjamins.
- Goldberg, Adele Eva. 2006. *Constructions at work: the nature of generalization in language*. Oxford: Oxford University Press.
- Goldberg, Adele Eva. 2009. The nature of generalization in language. *Cognitive Linguistics* 20(1). 93–127.
- Goldberg, Adele Eva. 2013. Argument Structure Constructions versus Lexical Rules or Derivational Verb Templates. *Mind & Language* 28(4). 435–465.
- Goldberg, Adele Eva, Devin Casenhiser and Nitya Sethuraman. 2004. Learning Argument Structure Generalizations. *Cognitive Linguistics* 15(3). 289–316.
- Gries, Stefan Thomas. 2007. Coll.analysis 3.2a.
- Gries, Stefan Thomas. 2012. Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics. Some necessary clarifications. *Studies in Language* 36(3). 477–510.
- Grondelaers, Stefan. 2000. De distributie van niet-anaforisch er buiten de eerste zinsplaats: sociolexicologische, functionele en psycholinguïstische aspecten van er’s status als presentatief signaal. Dissertation University of Leuven.
- Hopper, Paul and Sandra Annear Thompson. 1980. Transitivity in Grammar and Discourse. *Language* 56(2). 251–299.
- Hout, Anna Maria Henrica van. 1996. Event semantics of verb frame alternations: a case study of Dutch and its acquisition. Dissertation Tilburg University.
- Langacker, Ronald Wayne. 2009. Constructions and constructional meaning. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New Directions in Cognitive Linguistics [Human Cognitive Processing]*, 225–267. Amsterdam/Philadelphia: John Benjamins.
- Lenci, Alessandro. 2012. Argument alternations in Italian verbs: a computational study. In Valentina Bambini, Irene Ricci & Pier Marco Bertinetto (eds.), *Linguaggio e cervello - Semantica/Language and the Brain - Semantics. Atti del XLII Congresso Internazionale di Studi della Società di Linguistica Italiana*, 1–26. Rome: Bulzoni.
- Levshina, Natalia and Dirk Geeraerts. 2010. Constructing the Constructicon Empirically: Experiments with Dutch causatives. *International Conference on Construction Grammar*. September 5, Prague.
- Levshina, Natalia and Kris Heylen. 2014. A radically data-driven Construction Grammar: Experiments with Dutch causative constructions. In Ronny Boogaart, Timothy Colleman & Gijssbert Rutten (eds.), *Extending the Scope of Construction Grammar*, 17–46. Berlin: Mouton de Gruyter.
- Müller, Stefan. 2006. Phrasal or Lexical Constructions? *Language* 82(4). 850–883.
- Müller, Stefan and Phrasen Wechsler. 2014. Lexical approaches to argument structure. *Theoretical Linguistics* 40(1–2). 1–76.
- Norde, Muriel. 2014. On parents and peers in constructional networks. *CoglingDays* 6. December 12, Ghent.
- Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste and Ineke Schuurman. 2013a. 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.
- Pedersen, Johan. 2016. Spanish constructions of directed motion. A quantitative study. In Jiyoung Yoon & Stefan Thomas Gries (eds.), *Corpus-based approaches to Construction Grammar*, 105–144. Amsterdam/Philadelphia: John Benjamins.
- Perek, Florent. 2012. Alternation-based generalizations are stored in the mental grammar: Evidence from a sorting task experiment. *Cognitive Linguistics* 23(3). 601–635.
- Perek, Florent. 2015. *Argument structure in usage-based construction grammar: experimental and corpus-based perspectives*. Amsterdam/Philadelphia: John Benjamins.
- Perek, Florent and Adele Eva Goldberg. 2015. Generalizing beyond the input: The functions of the constructions matter. *Journal of Memory and Language* 84. 108–127.
- Perek, Florent and Martin Hilpert. 2014. Constructional tolerance: Cross-linguistic differences in the acceptability of non-conventional uses of constructions. *Constructions and Frames* 6(2). John Benjamins Publishing Company. 266–304.
- Perek, Florent and Maarten Lemmens. 2010. Getting at the meaning of the English at-construction: the case of a constructional split. *CogniTertes* 5. Association française de linguistique cognitive (AFLiCo).
- Pijpops, Dirk, Isabeau De Smet and Freek Van de Velde. Constructional contamination in morphology and syntax. Four case studies. *Constructions and Frames*.
- Pijpops, Dirk and Dirk Speelman. 2017. Alternating argument constructions of Dutch psychological verbs. A theory-driven corpus investigation. *Folia Linguistica* 51(1). 207–251.
- Pijpops, Dirk, Dirk Speelman, Stefan Grondelaers and Freek Van de Velde. Comparing explanations for the Complexity Principle. Evidence from argument realization.
- Pijpops, Dirk and Freek Van de Velde. 2016. Constructional contamination: How does it work and how do we measure it? *Folia Linguistica* 50(2). 543–581.
- Pijpops, Dirk and Freek Van de Velde. 2018. Lectal contamination. How language-external variation becomes language-internal through language contact. *Variationist Linguistics meets Contact Linguistics*. 21 May, Ascona.
- R Core Team. 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna.
- Rohdenburg, Günter. 1996. Cognitive Complexity and Increased Grammatical Explicitness in English. *Cognitive Linguistics* 7(2). 149–182.
- Speelman, Dirk. 2014. Logistic regression: A confirmatory technique for comparisons in corpus linguistics. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*, 487–533. (Human Cognitive Processing [HCP]). Amsterdam: John Benjamins.
- Speelman, Dirk, Kris Heylen and Dirk Geeraerts. 2018. Introduction. In Dirk Speelman, Kris Heylen & Dirk Geeraerts (eds.), *Mixed-Effects Regression Models in Linguistics*, 1–10. Cham: Springer.
- Stefanowitsch, Anatol and Stefan Thomas Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–244.
- De Swart, Peter. 2014. Prepositional Inanimates in Dutch: A Paradigmatic Case of Differential Object Marking. *Linguistics* 52(2). 445–468.
- Turney, Peter and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37. 141–188.
- van Noord, Gertjan van. 2006. At Last Parsing Is Now Operational. *TALN*. 20–42.
- van Trijp, Remi. 2015. Cognitive vs. generative construction grammar. The case of coercion and argument structure. *Cognitive Linguistics* 26(4). 613–632.
- Weeds, Julie, David Weir and Diana McCarthy. 2004. Characterising Measures of Lexical distributional similarity. *COLING ’04: Proceedings of the 20th international conference on Computational Linguistics*, 1015.
- Wellens, Pieter. 2011. Organizing constructions in networks. In Luc Steels (ed.), *Design Patterns in Fluid Construction Grammar*, 182–201. Amsterdam/Philadelphia: John Benjamins.
- Wible, David and Nai-Lung Tsao. 2017. Constructions and the problem of discovery: A case for the paradigmatic. *Corpus Linguistics and Linguistic Theory*. Published online, ahead of print.
- Zeldes, Amir. 2013. Productive argument selection: Is lexical semantics enough? *Corpus Linguistics and Linguistic Theory* 9(2). 263–291.

# Why

Solve the:

- Problem of Precedence
- Problem of Prediction
- Problem of Proliferation

# How

Incorporate the multi-level nature of the construction into corpus research:

- Map out the constructional network
- Systematically traverse it