The frequent good, the complex bad, and the unacceptable ugly

Ken Ramshøj Christensen
Aarhus University

Workshop on the
Syntax and Morphology of Danish, English, and Related Languages
on the occasion of the PhD defence of Katrine Rosendal Ehlers

2024.04.05

Work in collaboration with Anne Mette Nyvad

GLOSSA a journal of general linguistics

Home Articles Submissions Journal Policies About Search Login Register

Research Article

Complexity, frequency, and acceptability

Abstract

It is sometimes argued that (certain types of) lexical frequency and constructional frequency determine how easy sentences are to process and hence, how acceptable speakers find them. Others have argued that grammatical principles interact with and often override such effects. Here, we present the results from a survey on Danish with more than 200 participants. We asked people to provide acceptability ratings of a number of sentences with varying levels of complexity, with and without extraction, including complement clauses, relative clauses, parasitic gaps, and ungrammatical sentences. We predicted structural complexity and acceptability to be negatively correlated (the more complex, the less acceptable). The results show that construction frequency and acceptability are correlated, but that zero and near-zero frequencies do not predict acceptability. However, there is indeed an even stronger inverse correlation between acceptability and structural complexity, defined as a function of independently motivated factors of syntactic structure and processing, including embedding, adjunction, extraction, and distance between filler and gap. Lexical frequency also affects acceptability, but the effects are small, and, crucially, there is no evidence in our data that ungrammatical sentences are affected by such frequency effects. Furthermore, the acceptability patterns seem to be fairly stable across participants. The results show a pattern that is consistent with an approach based on grammatical principles and processing constraints, rather than based on stochastic principles alone.

Keywords

Syntax, embedding, adjunction, filler-gap dependency, path, extraction, animacy

Share

Authors

Ken Ramshøj Christensen (Aarhus University)
Anne Mette Nyvad (Aarhus University)

Download

Download PDF
Download XML

Issue

2024 • Volume 9

Dates

Accepted	2024-02-23
Published	2024-03-29

Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2023. "Sætningskompleksitet, frekvens og acceptabilitet." *Ny forskning i grammatik* 30 (October): 24–39. https://doi.org/10.7146/nfg.v1i30.137951.

Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2024. "Complexity, Frequency, and Acceptability." *Glossa: A Journal of General Linguistics* 9 (1): 1–44. https://doi.org/10.16995/glossa.10618.

Intro

- We have previously found different effects of lexical frequency on acceptability:
  - Christensen & Nyvad (2014): Positive correlation btw. frequency of matrix verb and acceptability of **RC extraction in Danish**.
  - Christensen & Nyvad (2022): No correlation btw. frequency of matrix verb and acceptability of **RC extraction in English**.
  - Christensen & Nyvad (2019): Neg. corr. btw. freq. of matrix verb and long adjunct **extraction from *wh*-question**.
- We have argued that frequency effects reflect processing ease of grammatical structures only.
- But we haven't checked for freq. effects across different structures…

Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2014. "On the Nature of Escapable Relative Islands." *Nordic Journal of Linguistics* 37 (1): 29–45. https://doi.org/10.1017/S0332586514000055.

Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2022. "The Island Is Still There: Experimental Evidence for the Inescapability of Relative Clauses in English." *Studia Linguistica* 76 (3): 1–25. https://doi.org/10.1111/stul.12192.

Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2019. "No Escape from the Island: On Extraction from Complement *Wh*-Clauses in English." In *The Sign of the V – Papers in Honour of Sten Vikner*, edited by Ken Ramshøj Christensen, Henrik Jørgensen, and Johanna L. Wood, 95–112. Aarhus: Dept. of English, School of Communication & Culture, Aarhus University. https://doi.org/10.7146/aul.348.91.

Frequency: Usage-based approaches

- Linguistic rules are structural regularities emerging from learners' lifetime analysis of the distributional characteristics of the language input. Acceptability judgements reflect the particular frequencies of the speaker's accidental experience (Ellis 2002; see also Bybee and Hopper 2001, Bybee 2007, Reali and Christiansen 2007).

- "When people repeatedly use the same particular and concrete linguistic symbols to make utterances to one another in "similar" situations, what may emerge over time is a pattern of language use, schematized in the minds of users as one or another kind of linguistic category or construction" (Tomasello 2003, 99).

Bybee, Joan. 2007. *Frequency of Use and the Organization of Language*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195301571.001.0001.

Bybee, Joan L., and Paul J. Hopper, eds. 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.

Ellis, Nick C. 2002. "Frequency Effects in Language Processing." *Studies in Second Language Acquisition* 24 (02): 143–88. https://doi.org/10.1017/S0272263102002024.

Reali, Florencia, and Morten H. Christiansen. 2007. "Processing of Relative Clauses Is Made Easier by Frequency of Occurrence." *Journal of Memory and Language* 57 (1): 1–23. https://doi.org/10.1016/j.jml.2006.08.014.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Frequency: Grammar-based approaches

- Others have argued that grammatical principles interact with and often override frequency effects

- e.g. in ambiguity resolution (Bornkessel, Schlesewsky & Friederici 2002; Pickering, Traxler, and Crocker 2000)
- and acceptability of subcategorization (White and Rawlins 2020).

Bornkessel, Ina, Matthias Schlesewsky, and Angela D. Friederici. 2002. "Grammar Overrides Frequency: Evidence from the Online Processing of Flexible Word Order." *Cognition* 85 (2): B21–30. https://doi.org/10.1016/S0010-0277(02)00076-8.

Pickering, Martin J, Matthew J Traxler, and Matthew W Crocker. 2000. "Ambiguity Resolution in Sentence Processing: Evidence against Frequency-Based Accounts." *Journal of Memory and Language* 43 (3): 447–75. https://doi.org/10.1006/jmla.2000.2708.

White, Aaron Steven, and Kyle Rawlins. 2020. "Frequency, Acceptability, and Selection: A Case Study of Clause-Embedding." *Glossa: A Journal of General Linguistics* 5 (1): 105. https://doi.org/10.5334/gjgl.1001.

Frequency: Grammar-based approaches

- Sentences with derived word order are more difficult to comprehend and produce for people with agrammatism, independently of frequency.
    - E.g. V2, fronting, scrambling: In Dutch, V2 is more frequent than V-final, and scrambled order as frequent as base order, but V2 and scrambling are more impaired (Bastiaanse, Bouma, and Post 2009).
    - And agrammatism is also characterised by impaired or non-use of grammatical elements, such as determiners, pronouns, prepositions, complementisers (Damasio 1992) – the most frequent words…

Bastiaanse, Roelien, Gosse Bouma, and Wendy Post. 2009. "Linguistic Complexity and Frequency in Agrammatic Speech Production." *Brain and Language* 109 (1): 18–28. https://doi.org/10.1016/j.bandl.2008.12.004.
Damasio, Antonio R. 1992. "Aphasia." *The New England Journal of Medicine* 236 (8): 531–39. https://doi.org/10.1056/nejm199202203260806.

Frequency: Grammar-based approaches

- In language acquisition, children show production patterns that deviate robustly from the input
  - ignoring speech errors and corrections, producing overgeneralizations and syntactic structures not in the target grammar (Pinker 2004, Thornton and Crain 1994, Yang 2004).

- And children prefer structurally simpler, but less frequent forms
  - E.g. Norwegian *min bil* (preferred; simplex: low freq.) vs. *bil$_1$-en min t$_1$* (complex, high freq.) (Anderssen and Westergaard 2010)

- "To account for the things speakers cannot say, the role of input frequency seems either minimal or insufficient: ungrammatical forms would rarely if ever appear in the input, and crucial disconfirming data may not be robustly represented to be useful to the learner" (Yang 2015, 290).

Anderssen, Merete, and Marit Westergaard. 2010. "Frequency and Economy in the Acquisition of Variable Word Order." *Lingua* 120 (11): 2569–88. https://doi.org/10.1016/j.lingua.2010.06.006.

Pinker, Steven. 2004. "Clarifying the Logical Problem of Language Acquisition." *Journal of Child Language* 31 (4): 949–53. https://doi.org/10.1017/S0305000904006439.

Thornton, Rosalind, and Stephen Crain. 1994. "Successful Cyclic Movement." In *Language Acquisition Studies in Generative Grammar*, edited by Teun Hoekstra and Bonnie D. Schwartz, 215–52. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lald.8.11tho.

Yang, Charles D. 2004. "Universal Grammar, Statistics or Both?" *Trends in Cognitive Sciences* 8 (10): 451–56. https://doi.org/10.1016/j.tics.2004.08.006.

Yang, Charles D. 2015. "For and against Frequencies." *Journal of Child Language* 42 (02): 287–93. https://doi.org/10.1017/S0305000914000683.

Structural complexity and graded acceptability.

Five factors:

- Many things increase "complexity", incl. structure [constituency (size of the tree in number of XPs), XP and head movement, length of chains], islands, binding, semantics [argument structure, number of propositions], pragmatics, finiteness, etc. We focus on the following five factors:

1. **Clausal embedding.** Constituency: more structure requires more processing (and WM) (Hawkins 1994, Pallier et al. 2011).
   - In particular finite clauses (phases and propositions, cf. the *Phase-Impenetrability Condition* (PIC) (Chomsky 2001, 13).
   - People with agrammatism have problems with finiteness and with embedded clauses (Friedmann 2003)

Chomsky, Noam. 2001. "Derivation by Phase." In *Ken Hale: A Life in Language*, edited by Michael J. Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Friedmann, Na'ama. 2003. "The Fragile Nature of the Left Periphery: CP Deficits in Agrammatic Aphasia." In *Proceedings of the 18th IATL Conference*. http://linguistics.huji.ac.il/IATL/18/Friedmann.pdf.

Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Pallier, Christophe, Anne-Dominique Devauchelle, and Stanislas Dehaene. 2011. "Cortical Representation of the Constituent Structure of Sentences." *Proceedings of the National Academy of Sciences* 108 (6): 2522–27. https://doi.org/10.1073/pnas.1018711108.

Structural complexity and graded acceptability.

Five factors:

2. **Adjunction**: Adjunction always increase the number of XPs, it increase the WM load (in particular right-adjunction).
   - With extraction, the base-position/gap is not selected (cf. the *Empty Category Principle* (ECP), Haegeman 1994, 442). Adjuncts are often assumed to be islands (cf. the *Condition on Extraction Domains* (CED), Huang 1982, 505).
   - (Agrammatism again: an adjunction deficit, Lee and Thompson 2011)

3. **Move-Out:** Movement out of an embedded clause is particularly costly (fMRI: Christensen, Kizach, and Nyvad 2013b).
   - This is most likely due to the PIC (Chomsky 2001, 13)

Christensen, Ken Ramshøj, Johannes Kizach, and Anne Mette Nyvad. 2013b. "The Processing of Syntactic Islands – An fMRI Study." *Journal of Neurolinguistics* 26 (2): 239–51. https://doi.org/10.1016/j.jneuroling.2012.08.002.

Chomsky, Noam. 2001. "Derivation by Phase." In *Ken Hale: A Life in Language*, edited by Michael J. Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Haegeman, Liliane. 1994. *Introduction to Government and Binding Theory*. 2nd ed. 1. Oxford: Blackwell.

Huang, Cheng-Teh James. 1982. "Logical Relations in Chinese and the Theory of Grammar." PhD dissertation, Cambridge, MA: MIT. http://pubman.mpdl.mpg.de/pubman/item/escidoc:403106:7/component/escidoc:403105/chinese_huang1982_o.pdf.

Lee, Jiyeon, and Cynthia K. Thompson. 2011. "Real-Time Production of Arguments and Adjuncts in Normal and Agrammatic Speakers." *Language and Cognitive Processes* 26 (8): 985–1021. https://doi.org/10.1080/01690965.2010.496237.

Structural complexity and graded acceptability.

Five factors:

4. **Path (structural distance of movement):** The number of **overt XPs** ("those that are actually perceived and processed") between the filler and the gap in the base-position ('The Filler-Gap Domain', Hawkins 1999, 248-249).
   - (See also Collins 1994, 56 and O'Grady, Lee, and Choo 2003, 435; Christensen, Kizach, and Nyvad 2013a).

5. **Number of fillers:**

- Cf. ***Economy of Derivation***: "make derivations as short as possible, with links as short as possible" (Chomsky 1995, 91)
  - "In fact, this unification of WM and theoretical syntax also seems to call for a revival of the much vilified **Derivational Theory of Complexity** (DTC) in some form, as also suggested by Marantz [2005]" (Christensen 2005, 308) (see also Hornstein 2014).

Christensen, Ken Ramshøj. 2005. "Interfaces: Negation - Syntax - Brain." PhD dissertation, Aarhus: Aarhus University. http://tildeweb.au.dk/au572/Papers/krc-phd.pdf.

Christensen, Ken Ramshøj, Johannes Kizach, and Anne Mette Nyvad. 2013a. "Escape from the Island: Grammaticality and (Reduced) Acceptability of *Wh*-Island Violations in Danish." *Journal of Psycholinguistic Research* 42 (1): 51–70. https://doi.org/10.1007/s10936-012-9210-x.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Collins, Chris. 1994. "Economy of Derivation and the Generalized Proper Binding Condition." *Linguistic Inquiry* 25 (1): 45–61. https://www.jstor.org/stable/4178848.

Hawkins, John A. 1999. "Processing Complexity and Filler-Gap Dependencies across Grammars." *Language* 75 (2): 244–85. https://doi.org/10.2307/417261.

Hornstein, Norbert. 2014. "DTC Redux." *Faculty of Language* (blog). February 24, 2014. http://facultyoflanguage.blogspot.com/2014/02/dtc-redux.html.

O'Grady, William, Miseon Lee, and Miho Choo. 2003. "A Subject-Object Asymmetry in the Acquisition of Relative Clauses in Korean as a Second Language." *Studies in Second Language Acquisition* 25 (3): 433–48. https://doi.org/10.1017/S0272263103000172.

Predictions

- **Prediction 1: Acceptability decreases as complexity increases.** Because structural complexity, which is a function of (at least) embedding, adjunction, Move-Out, path/distance, and number of fillers, increases processing cost, it is negatively correlated with acceptability.

  - Long movement from **an embedded complement clause** is predicted to reduce acceptability compared to the corresponding sentences without movement (Christensen, Kizach, and Nyvad 2013a,b).
  - Extraction from **relative clauses**, which score highly on structural complexity, are predicted to have intermediate acceptability. Christensen and Nyvad (2014) have argued that Danish RC extractions are grammatical, but acceptability, and may be affected by lexical frequency.
  - Likewise, **parasitic gaps** (Engdahl 1983), which are reported to be very rare, score high on structural complexity, and hence, are predicted to show intermediate acceptability.

Christensen, Ken Ramshøj, Johannes Kizach, and Anne Mette Nyvad. 2013a. "Escape from the Island: Grammaticality and (Reduced) Acceptability of *Wh*-Island Violations in Danish." *Journal of Psycholinguistic Research* 42 (1): 51–70. https://doi.org/10.1007/s10936-012-9210-x.

Christensen, Ken Ramshøj, Johannes Kizach, and Anne Mette Nyvad. 2013b. "The Processing of Syntactic Islands – An fMRI Study." *Journal of Neurolinguistics* 26 (2): 239–51. https://doi.org/10.1016/j.jneuroling.2012.08.002.

Christensen, Ken Ramshøj, and Anne Mette Nyvad. 2014. "On the Nature of Escapable Relative Islands." *Nordic Journal of Linguistics* 37 (1): 29–45. https://doi.org/10.1017/S0332586514000055.

Engdahl, Elisabet. 1983. "Parasitic Gaps." *Linguistics and Philosophy* 6 (1): 5–34. https://doi.org/10.1007/BF00868088.

Predictions

- **Prediction 2:** Acceptability is also predicted by construction frequency, but the correlation is weaker. (Low acceptability is not correlated with zero frequency, or vice versa.)
  - Both RC-extraction and parastic gaps (PGs) are **very rare in corpus data**, parasitic gaps presumably almost non-existent, but both are assumed to be **grammatical**. That is, the relative frequency (maybe zero) of the constructions is predicted not to predict the acceptability level.
  - Like PGs, ungrammatical 'word salad' has a zero probability (non-occurrence in a corpus), but unlike PGs, it's clearly ungrammatical and unacceptable.

Predictions

- **Prediction 3:** The level of acceptability is predicted to be somewhat but not dramatically **affected by lexical frequency.**
  - Crucially, **ungrammatical sentences** are predicted to be immune to such effects. Only grammatical sentences can be modulated by frequency and trial (repetition), and the 'baseline' acceptability is determined by complexity.

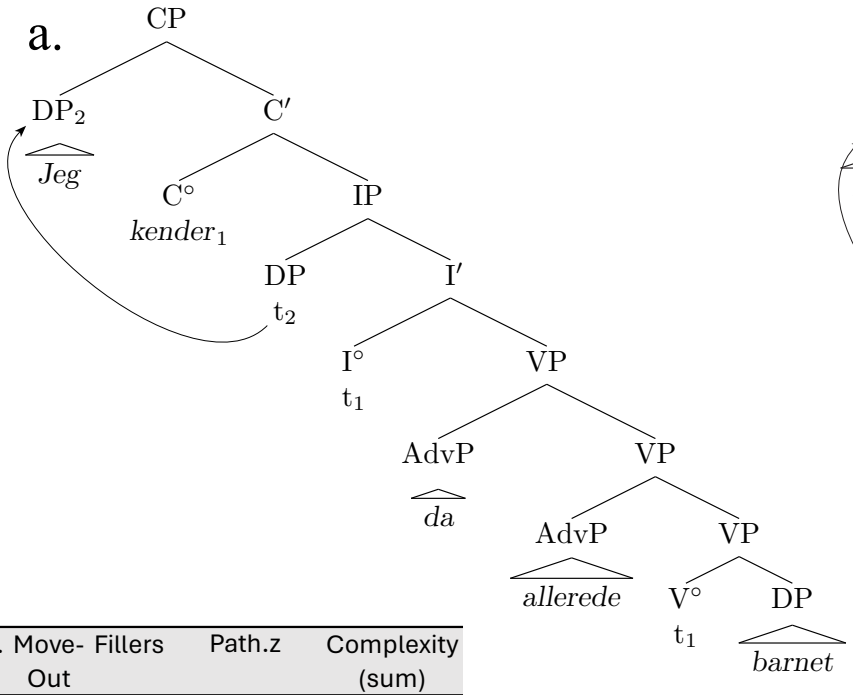
Stimuli:

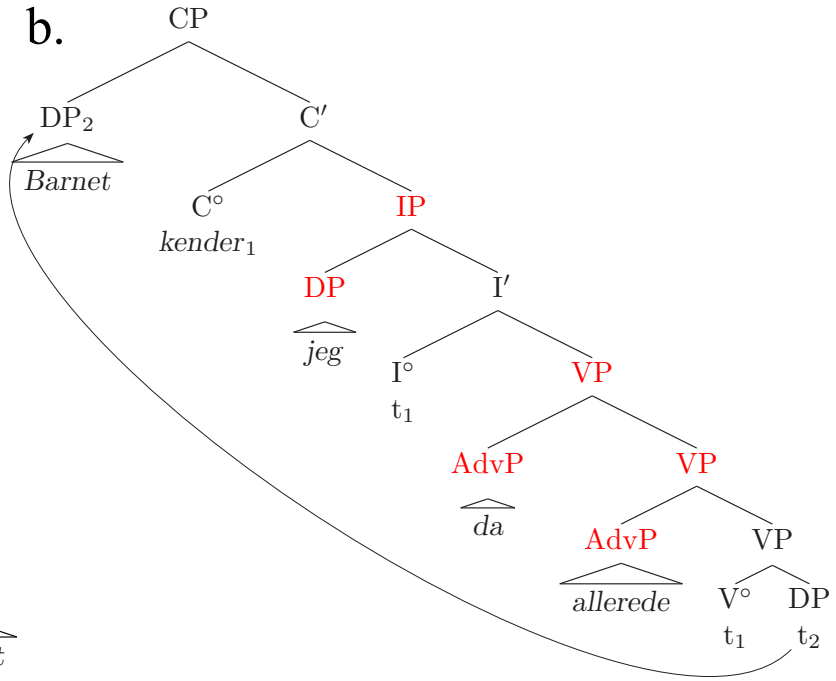
Increasing complexity, ±Obj. extraction, varying frequency

- Context: *Min bedste vens nye kæreste vil gerne have at jeg skal møde hans datter og hans hund i weekenden.*

Men jeg kender da allerede barnet.	**A1**	Simplex
Men barnet kender jeg da allerede __.	**A2**	[+Extr]
Men jeg tror da allerede [at jeg kender barnet].	**B1**	Complement clause
Men barnet tror jeg da allerede [at jeg kender __].	**B2**	[+Extr]
Men jeg kender da allerede barnet [uden at have mødt hende].	**C1**	Adjunct clause
Men barnet kender jeg da allerede __ [uden at have mødt __].	**C2**	[+Extr] [Parasitic Gap]
Men jeg kender da allerede én [OP der __ har mødt barnet].	**D1**	Relative clause
Men barnet kender jeg da allerede én [OP der __ har mødt __].	**D2**	[+Extr]
Men jeg kender da allerede hunden og barnet.	**E1**	Grammatical controls
*Men barnet jeg kender hunden og da allerede.	**E2**	Ungrammatical controls (*V2, *Coord, *adv. place.)

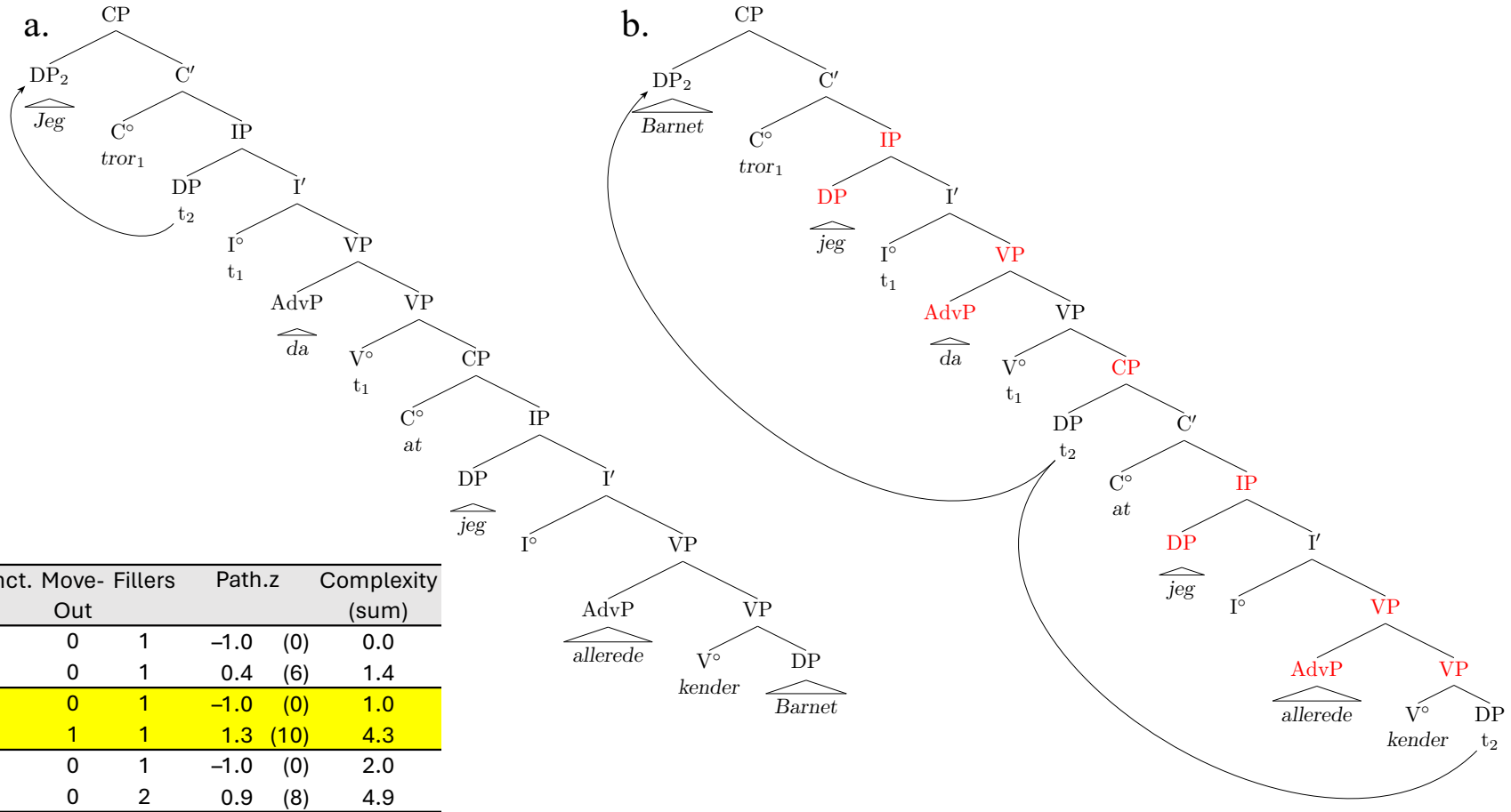
Simplex: A1 vs A2





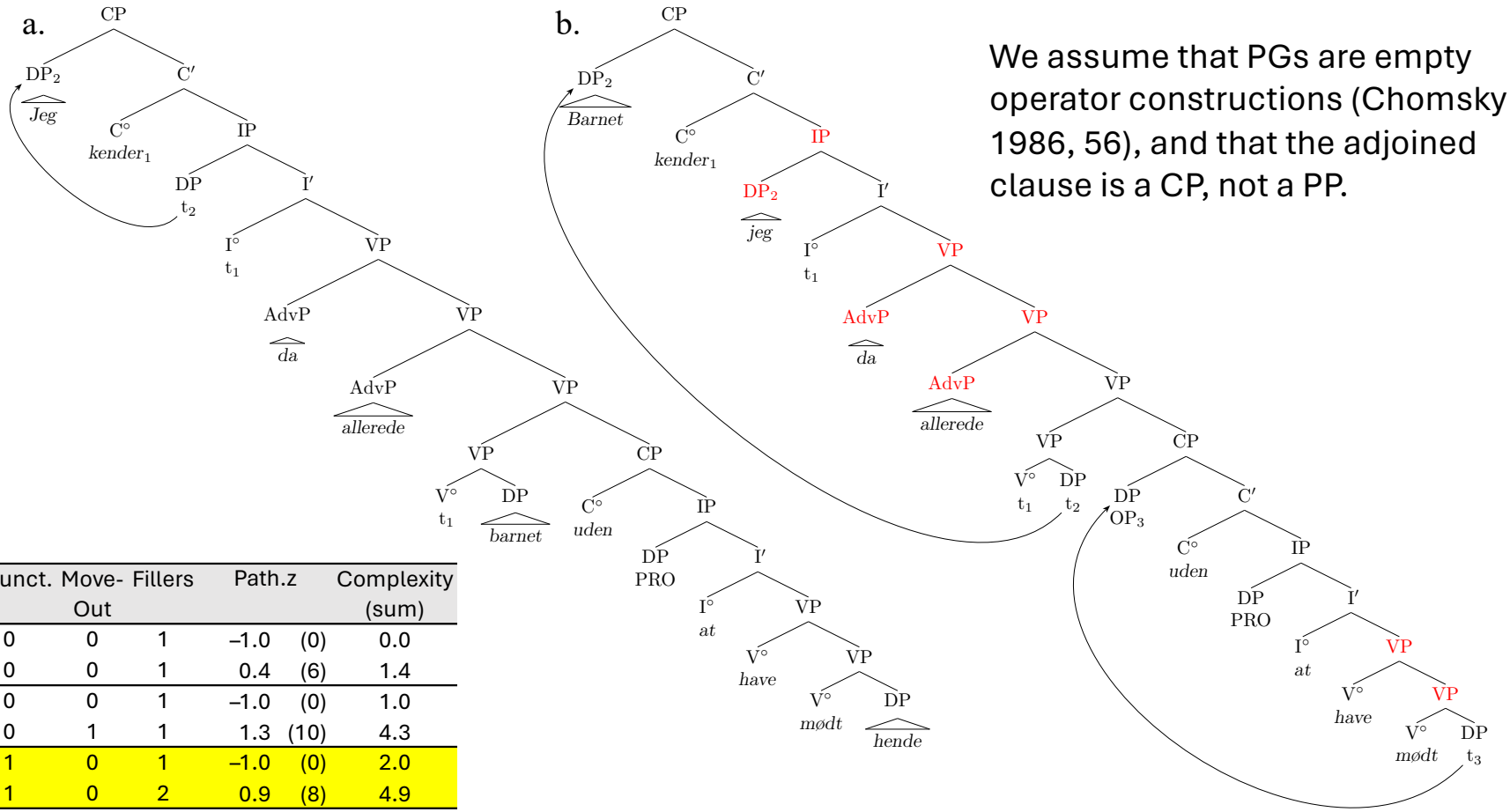
Condition	Embed.	Adjunct.	Move-Out	Fillers	Path.z		Complexity (sum)
A1: Simplex [–Ex]	0	0	0	1	–1.0	(0)	0.0
A2: Simplex [+Ex]	0	0	0	1	0.4	(6)	1.4
B1: Compl. [–Ex]	1	0	0	1	–1.0	(0)	1.0
B2: Compl. [+Ex]	1	0	1	1	1.3	(10)	4.3
C1: Adjunct [–Ex]	1	1	0	1	–1.0	(0)	2.0
C2: Adjunct [+Ex] (PG)	1	1	0	2	0.9	(8)	4.9
D1: RC [–Ex]	1	1	0	2	–1.0	(0)	3.0
D2: RC [+Ex]	1	1	1	2	1.3	(10)	6.3

Complement clause: B1 vs. B2



Condition	Embed.	Adjunct.	Move-Out	Fillers	Path.z		Complexity (sum)
A1: Simplex [–Ex]	0	0	0	1	–1.0	(0)	0.0
A2: Simplex [+Ex]	0	0	0	1	0.4	(6)	1.4
B1: Compl. [–Ex]	1	0	0	1	–1.0	(0)	1.0
B2: Compl. [+Ex]	1	0	1	1	1.3	(10)	4.3
C1: Adjunct [–Ex]	1	1	0	1	–1.0	(0)	2.0
C2: Adjunct [+Ex] (PG)	1	1	0	2	0.9	(8)	4.9
D1: RC [–Ex]	1	1	0	2	–1.0	(0)	3.0
D2: RC [+Ex]	1	1	1	2	1.3	(10)	6.3

Adjunction: C1 vs. C2 [Parasitic Gap]

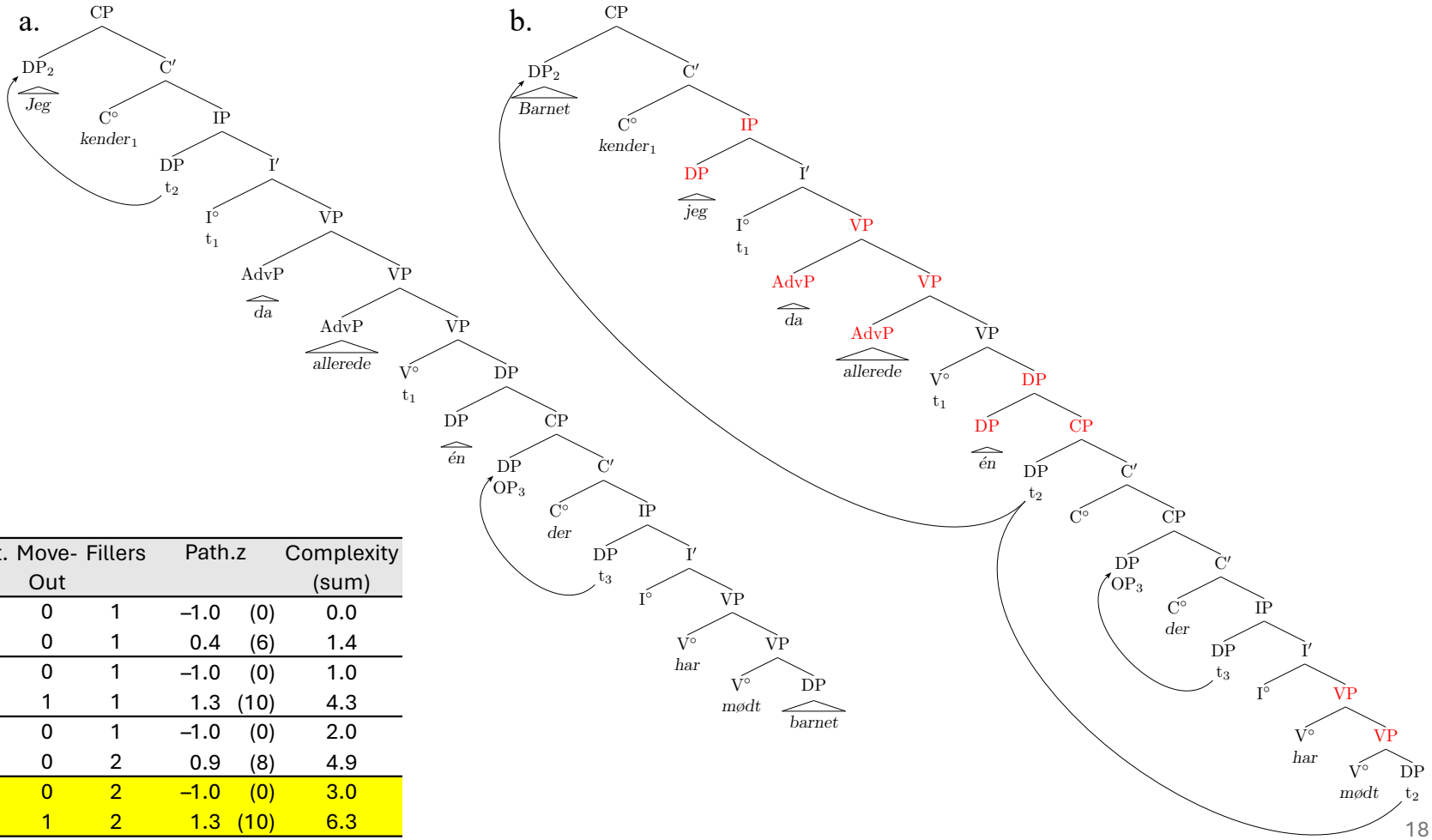


We assume that PGs are empty operator constructions (Chomsky 1986, 56), and that the adjoined clause is a CP, not a PP.

Condition	Embed.	Adjunct.	Move-Out	Fillers	Path.z		Complexity (sum)
A1: Simplex [–Ex]	0	0	0	1	–1.0	(0)	0.0
A2: Simplex [+Ex]	0	0	0	1	0.4	(6)	1.4
B1: Compl. [–Ex]	1	0	0	1	–1.0	(0)	1.0
B2: Compl. [+Ex]	1	0	1	1	1.3	(10)	4.3
C1: Adjunct [–Ex]	1	1	0	1	–1.0	(0)	2.0
C2: Adjunct [+Ex] (PG)	1	1	0	2	0.9	(8)	4.9
D1: RC [–Ex]	1	1	0	2	–1.0	(0)	3.0
D2: RC [+Ex]	1	1	1	2	1.3	(10)	6.3

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Relative clauses: D1 vs. D2



Condition	Embed.	Adjunct.	Move-Out	Fillers	Path.z		Complexity (sum)
A1: Simplex [–Ex]	0	0	0	1	–1.0	(0)	0.0
A2: Simplex [+Ex]	0	0	0	1	0.4	(6)	1.4
B1: Compl. [–Ex]	1	0	0	1	–1.0	(0)	1.0
B2: Compl. [+Ex]	1	0	1	1	1.3	(10)	4.3
C1: Adjunct [–Ex]	1	1	0	1	–1.0	(0)	2.0
C2: Adjunct [+Ex] (PG)	1	1	0	2	0.9	(8)	4.9
D1: RC [–Ex]	1	1	0	2	–1.0	(0)	3.0
D2: RC [+Ex]	1	1	1	2	1.3	(10)	6.3

Finding construction frequencies in KorpusDK

Con.	Search frame	Hits	Log_{10}
A1	Coordinator (‘but/and’) NOM.Pronoun (‘I/you/he/she/we/you/they’) Verb Verb ACC.Pronoun (‘me/you/him/her/us/you/them’)	1,165	3.07
A2	Coordinator (‘but/and’) ACC.Pronoun (‘me/you/him/her/us/you/them’) Verb NOM.Pronoun (‘I/you/he/she/we/you/they’) Verb	120	2.08
B1	Complementiser NOM.Pronoun (‘I/you/he/she/we/you/they’) Verb Comp (‘that’) NOM.Pronoun (‘I/you/he/she/we/you/they’) Verb Verb ACC.Pronoun (‘me/you/him/her/us/you/them’)	98	1.99
B2	Complementiser ACC.Pronoun (‘me/you/him/her/us/you/them’) VERB NOM.Pronoun (‘I/you/he/she/we/you/they’) Comp (‘that’) NOM.Pronoun (‘I/you/he/she/we/you/they’) Verb Verb	2	0.30
C1	Complementiser (‘without/after’) Inf. (‘to’) Verb Verb ACC.Pronoun (‘me/you/him/her/us/you/them’)	157	2.20
C2	–	1	0.00
D1	Pron (‘one/anyone/everybody/someone’) REL (‘$that_{Subj}$’) Verb Verb ACC.Pronoun (‘me/you/him/her/us/you/them’)	20	1.30
D2	ACC.Pronoun (‘me/you/him/her/us/you/them’) Verb NOM.Pronoun (‘I/you/he/she/we/you/they’) Pron (‘one/anyone/everybody/someone’) REL (‘$that_{Subj}$’)	1	0.00
E1	=A1	1,165	3.07
E2	–	1	0.00

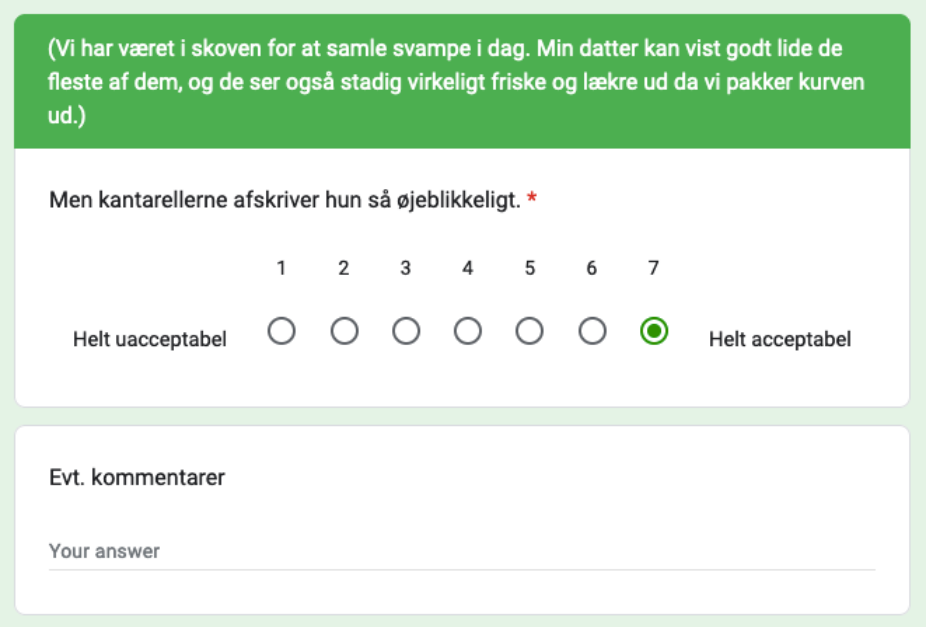
Lexical frequencies in KorpusDK

- We constructed 24 sets (contexts) of 10 sentences (A1-E2), 240 in total (Subj = pron [*jeg* / *han* / *hun*], 50% animate Obj).
  - The stimuli was balanced over three (relative) **frequency zones.**
  - E.g. the same V, ADV, and object N were used to construct a set in the high freq. zone, etc.
  - The main V in cond. A (simplex) was used as either matrix or embedded main V in cond. B-D (complex).
  - All V were simple present tense, all N were definite and non-compound.
  - All sentences were controlled for cohesion.
  - Discourse particle (e.g. *jo*, *da*, *bare*) were inserted to make sentences more natural.

	Low frequency	**Middle frequency**	**High frequency**
Verb	*fotografere* (1,102) 'photograph' *forgude* (102) worship'	*vurdere* (7,936) 'estimate' *bage* (1,723) 'bake'	*finde* (60,402) 'find' *elske* (8,981) 'love'
Noun	*kanin* (587) 'rabbit' *mandolin* (22) 'mandolin'	*kat* (2,315) 'cat' *ur* (833) 'watch'	*barn* (62,538) 'child' *cykel* (3,073) 'bike'
Adverb	*utvivlsomt* (762) 'undoubtedly' *formodentligt* (23) 'presumably'	*selvfølgelig* (15,269) 'of course' *umiddelbart* (3,490) 'offhand'	*også* (187,096) 'also' *alligevel* (19,875) 'nonetheless'
Log_{10} mean	2.33	3.47	4.31
Log_{10} range	1.92–2.63	3.30–3.69	4.02–4.85

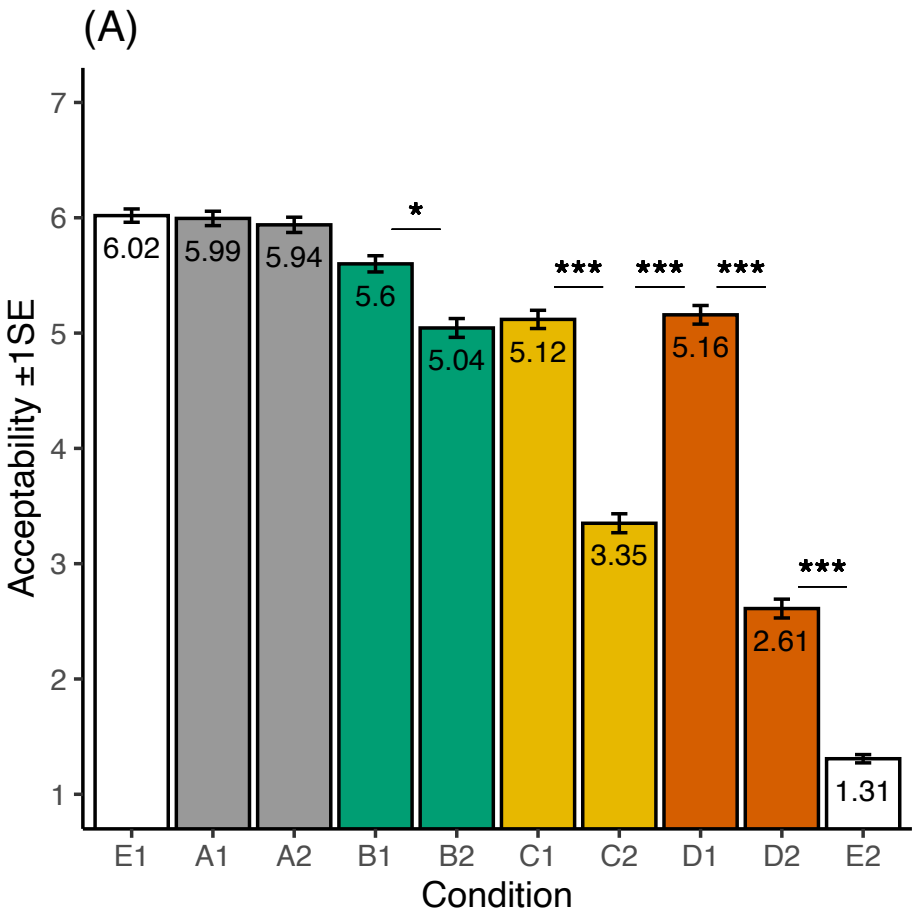
Setup

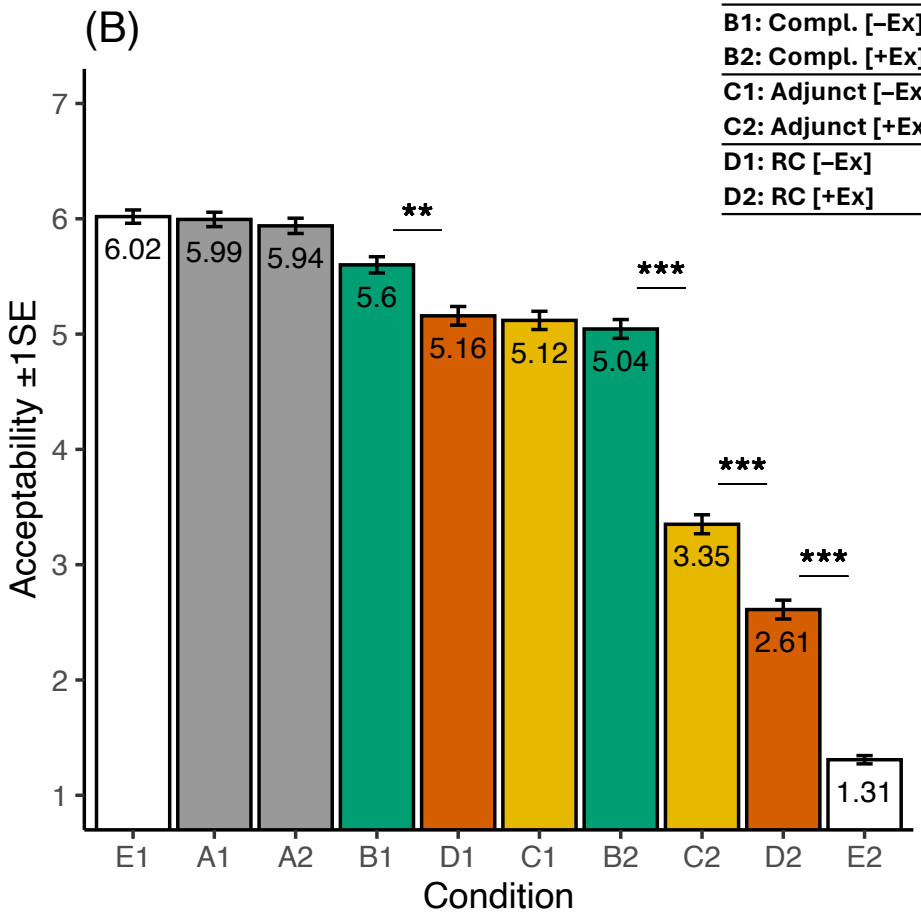
- Offline survey
  - Google Form
  - Latin-square design
  - Participants pseudo-randomly assigned to 1 of 10 lists based on their month of birth.
- Task:
  - Acceptability rating on a 7-point Likert scale
- Participants:
  - n=212 (Female: 196, Male: 14, Other: 1)
  - Age: 18-63 years, mean=28.8, SD=8.8



Results

Condition	Complexity (sum)
A1: Simplex [–Ex]	0.0
A2: Simplex [+Ex]	1.4
B1: Compl. [–Ex]	1.0
B2: Compl. [+Ex]	4.3
C1: Adjunct [–Ex]	2.0
C2: Adjunct [+Ex] (PG)	4.9
D1: RC [–Ex]	3.0
D2: RC [+Ex]	6.3





Mixed-effects models w. sliding contrasts
*** p < 0.001, ** p < 0.01, * p < 0.05

Complexity, acceptability, and frequency

(A): Significant positive correlation btw. construction frequency and acceptability (R^2 = 0.69, p < 0.004), describing 69% of the variation.

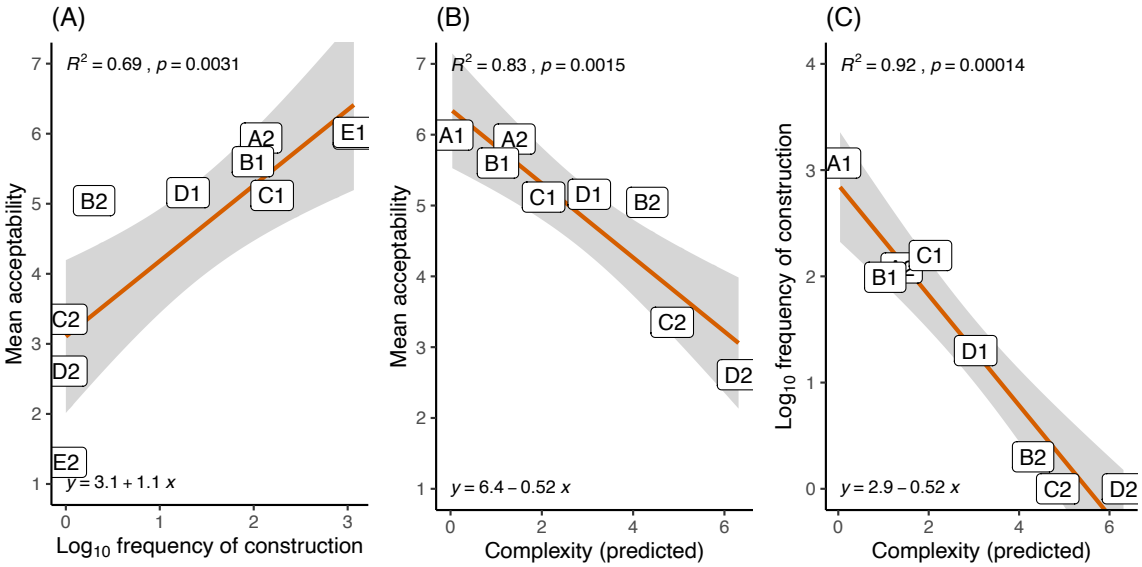
- However, the fit is not perfect (E2 and B2 are far from the linear trendline), which suggests that other factors are at play – processing factors, we argue.

(B): Significant negative correlation btw. complexity and acceptability, which has a larger effect size (R^2 = 0.83, p < 0.002) **describing 83% of the variation.**

- Excluding fillers. E2 = A1 (apart from the coordinate object), and E2 is ungrammatical (with scrambled, impossible word order) and has no meaningful complexity level.
- Excluding fillers in (A) only increases the effect from R^2 = 0.69 to R^2 = 0.70, p < 0.01).

(C): Complexity and (log10) frequency are also strongly negatively correlated (R^2 = 0.92, p < 0.001), describing 92% of the variation.

- The construction frequency patterns can be predicted from complexity.



Effects of the individual complexity factors

- With one exception, each factor shows significant negative correlations with acceptability.
  - The effect of Move-Out is not significant. However, it is probably covered by the effects of embedding, adjunction, and path.
- Interestingly, fillers are extra costly. The effect of adding one more filler is more than twice as big as the effect of embedding.
- In contrast, the cost in moving across a single (overt) XP is –0.073 points on the acceptability scale. So, the longer the movement, the worse, as expected.
  - The relatively small effect provides justification for z-transformation (Path.z)
  - Path affects the relative acceptability, but it does not determine the base level.

	Estimate	SE	df	t	p	
(Intercept)	7.283	0.281	117.051	25.890	0.000	***
Embedding	**–0.402**	0.167	86.561	–2.401	0.019	*
Adjunction	**–0.724**	0.250	121.150	–2.901	0.004	**
Move-Out	**–0.329**	0.260	169.283	–1.266	0.207	
Fillers	**–1.035**	0.266	90.585	–3.894	0.000	***
Path	**–0.073**	0.024	168.170	–2.999	0.003	**

Mixed effects model:

```
FREQ_FACTORS_MODEL <- lmer(ACCEPTABILITY ~ EMBED+ADJ+MOVEOUT+FILLERS+PATH +
                    (1+EMBED+ADJ+MOVEOUT+FILLERS+PATH+TRIAL|SUBJ) +
                    (1+ADJ+TRIAL|ITEM), data=DATA)
```

Fitted complexity of construction

We computed new complexity levels by multiplying each factor level (0, 1, or 2with the estimates (effect sizes) and adding them up.

- Complexity (fitted) = 0.402 × embedding + 0.742 × adjunction + 0.329 × Move-Out + 1.035 × fillers + 0.073 × (raw) path

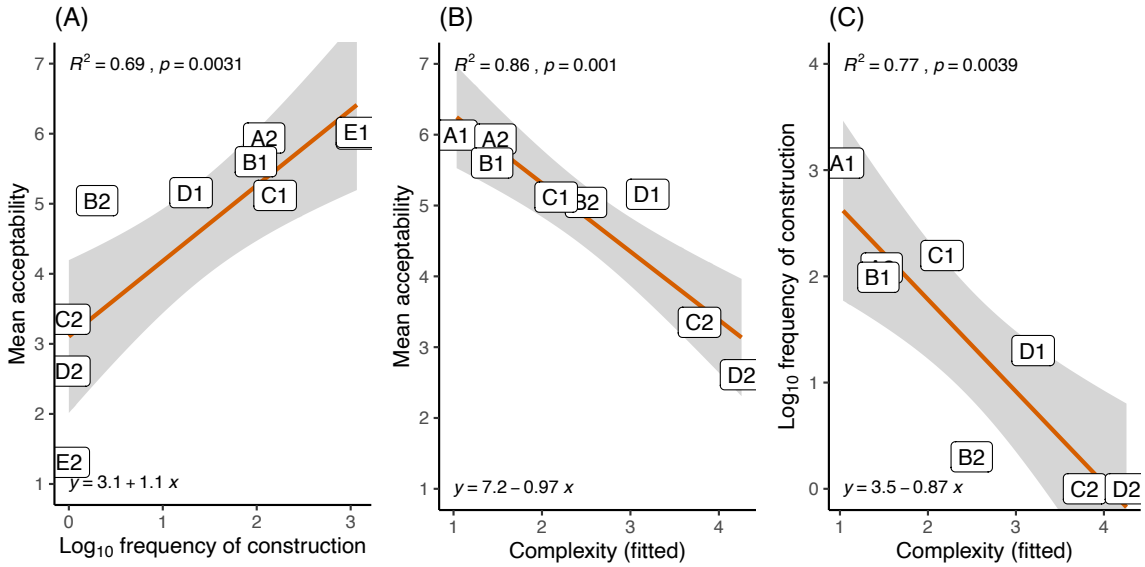
(B): Significant negative correlation btw. acceptability and complexity (R^2 = 0.86)

- (almost) the same as for the predicted complexity (R^2 = 0.83).

(C): Significant correlation btw. complexity and frequency (R^2 = 0.77)

These results show that our predicted complexity levels were neither completely arbitrary nor off target.

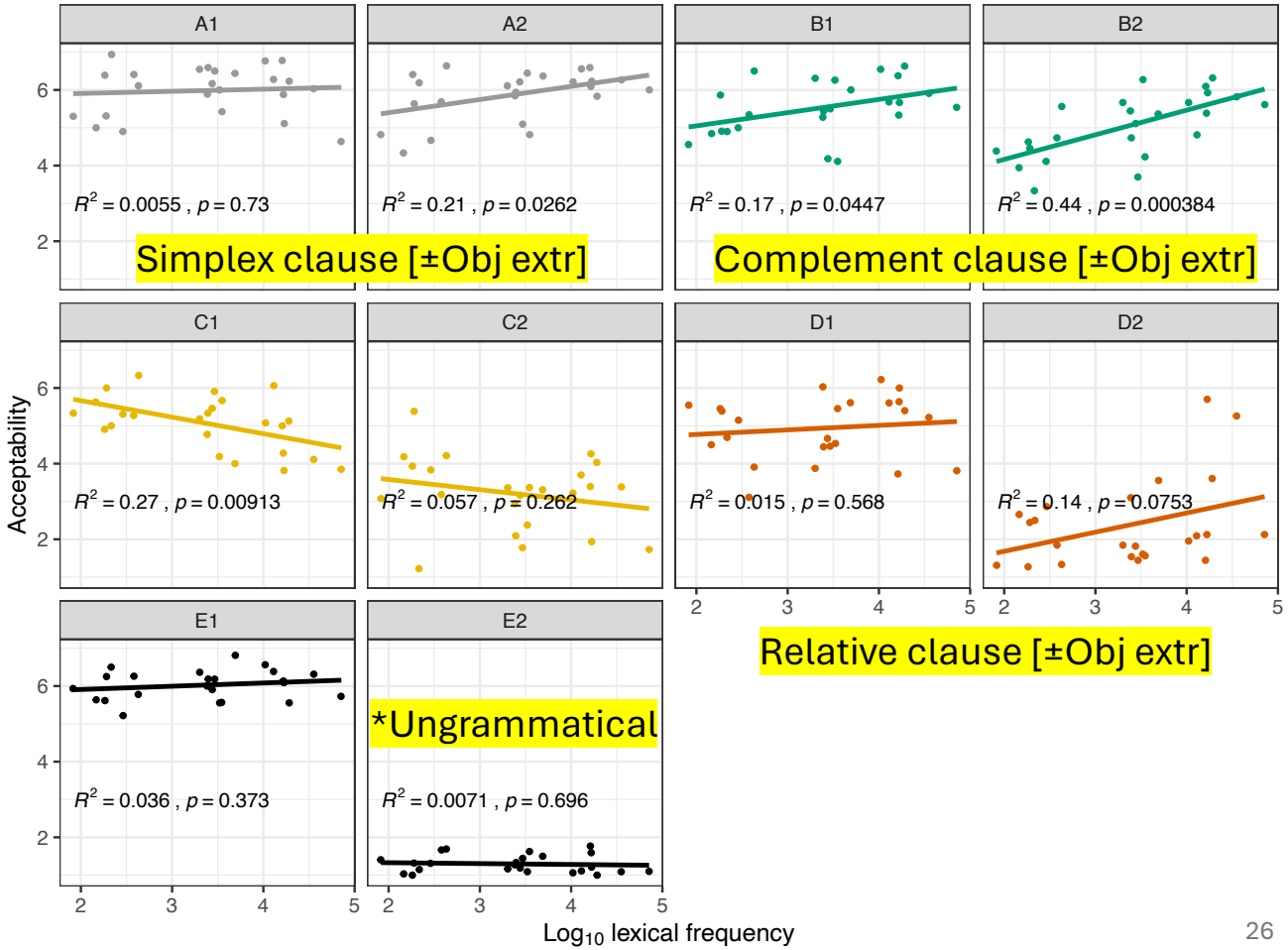
- The exact relative complexity levels of some the intermediate-level constructions (B2, C1, C2) are not important. The overall pattern is the same.
- Indeed, the predicted model is a stronger predictor of ($\log_{10}$) frequency (R^2 = 0.92).



Lexical (mean $\log_{10}$) frequency

- As expected, lexical frequency has **no significant effect** on clearly ungrammatical (impossible) fillers (E2).

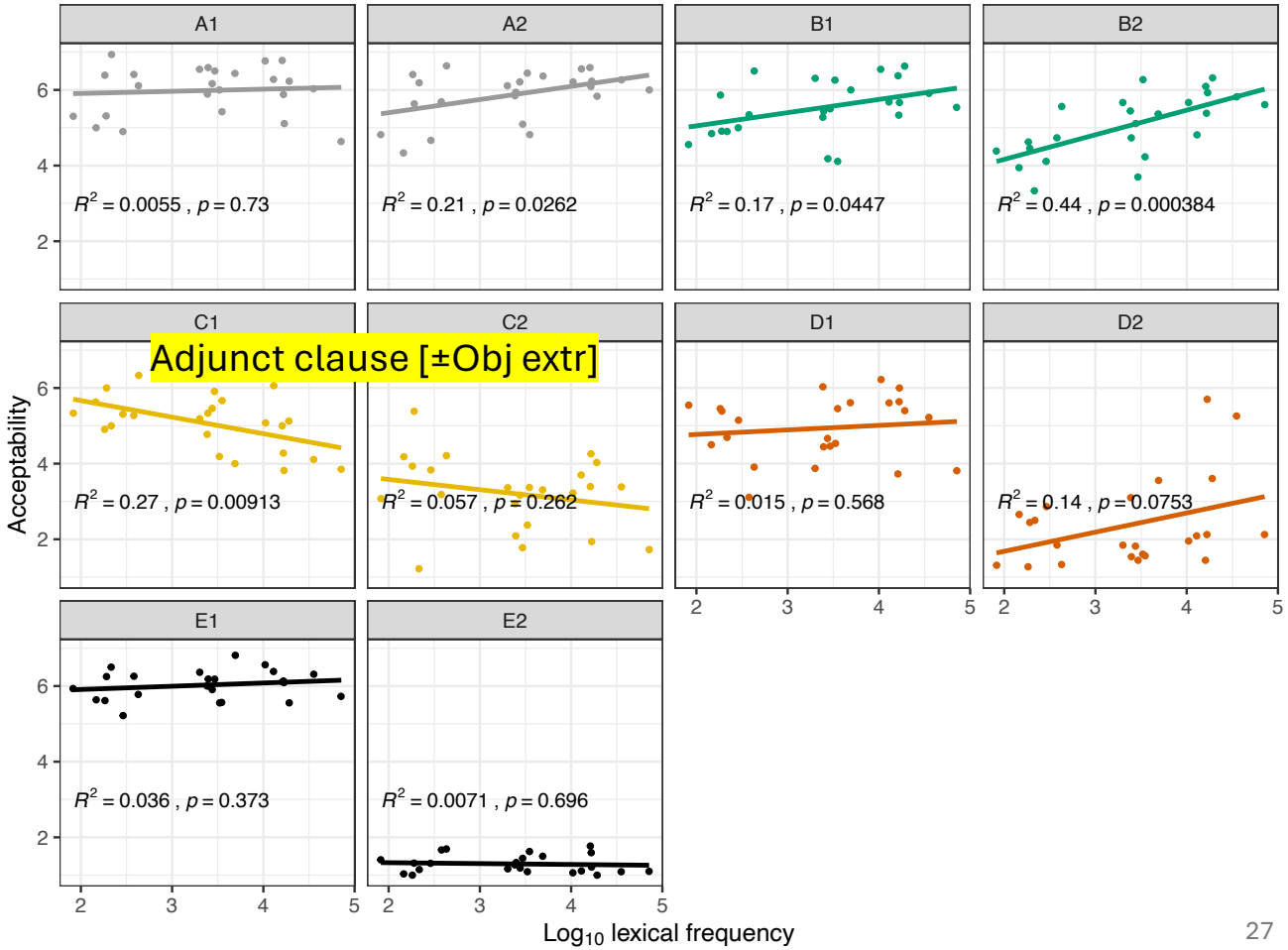
- Lexical frequency has a positive effect on the acceptability of extraction, especially with long extraction (B2):
  - Higher freq. ➔ faster lexical retrieval ➔ lower processing load ➔ higher acceptability



Lexical (mean $\log_{10}$) frequency

- An unexpected (and unexplained) inverse frequency effect:

- The cost of adjunction to VP (C1) increases with lexical frequency, but less so with extraction (C2)!?!
  - Why does lexical frequency have a negative effect on the acceptability of clausal adjunction, which is fairly common?
  - And why is the effect small and non-significant for parasitic gaps, which are very rare?



Summary and conclusions

- The predictions from the complexity model are all borne out

1. **Acceptability decreases as complexity increases.**
   Structural complexity (here) can be described as a function of (at least) five factors: (1) **embedding**, (2) **adjunction**, (3) **Move-Out**, (4) **path**, and (4) number of **fillers**.

2. **Acceptability is also predicted by construction frequency, but the correlation is weaker.**
   Indeed, it seems more likely that the construction frequency is a function of complexity, which in turn also predicts acceptability. Frequency of occurrence is the *explanandum* (what needs to be explained), not the *explicans* (the explanation).

3. **The level of acceptability is somewhat but not dramatically affected by lexical frequency.**
   The 'baseline' acceptability is determined by complexity.

Summary and conclusions

- **The results show a pattern that is consistent with grammatical principles and processing constraints.**

- **Mind the structure!**
  "The more processing involved, the rarer the structure" (Newmeyer 2005, 126).

Newmeyer, Frederick J. 2005. *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*. New York: Oxford University Press.

Thank you for you attention