



Christensen, Ken Ramshøj & Nyvad, Anne Mette. 2024. Complexity, frequency, and acceptability. *Glossa: a journal of general linguistics* 9(1). pp. 1–44. DOI: <https://doi.org/10.16995/glossa.10618>

# Open Library of Humanities

## Complexity, frequency, and acceptability

**Ken Ramshøj Christensen**, Aarhus University, Denmark, [krc@cc.au.dk](mailto:krc@cc.au.dk)

**Anne Mette Nyvad**, Aarhus University, Denmark, [amn@cc.au.dk](mailto:amn@cc.au.dk)

---

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.

---

*Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 OPEN ACCESS



## 1 Introduction

The frequency with which we are exposed to different linguistic expressions undoubtedly matters to whether or not, and how fast, we learn to use such expressions. However, it is still a matter of controversy whether such frequency is a determining factor in how easy sentences are to process and how acceptable speakers find them, or whether frequency of occurrence or exposure plays a much lesser role. In functional (usage-based) linguistics, it is sometimes argued that (at least certain types of) lexical frequency and constructional frequency play a central and crucial role in the ease and speed of language processing and in what is and is not learnt, in first as well as second language learning. In formal linguistics, including processing accounts in general, on the other hand, it is often argued that a number of structural properties interact with frequency, and indeed often take priority over stochastic information. However, so far, the relationship between construction frequency, lexical frequency, structural complexity, and native speaker acceptability has not been studied in one and the same experiment in a way that allows for generalizations about the influence of frequency. In this paper, we present the results of an experiment (a quantitative acceptability survey) on constructional and lexical frequency effects across a number of sentence types with varying degrees of structural complexity. Before we get to the experiment itself, we will first briefly review the literature on a number of factors that have been argued to determine or influence how acceptable native speakers find various types of sentences: lexical frequency, construction frequency, syntax vs. semantics, usage-based approaches, grammar-based approaches, (factors in) structural complexity, and animacy.

### 1.1 Lexical frequency

It is well-known that lexical frequency is positively correlated with word learning and processing speed. The more frequently children hear certain words, the faster they learn to use those words correctly. Higher frequency words (common words) are recognized and processed faster than lower frequency (or rare) words (see the discussion in Harley 2014: 143, 171–172). Furthermore, higher frequency words tend to be shorter than words with lower frequency, cf. ‘Zipf’s Law’ (Zipf 1935). In spite of this negative correlation between word frequency and word length, and though function words tend to be shorter than content words, the correlation between processing speed and frequency only applies to lexical content words, not to grammatical function words, e.g. auxiliary verbs, determiners, and complementizers (Harley 2014: 315; see also Bell et al. 2009; Piantadosi et al. 2011). Likewise, children’s acquisition of function words depends less on frequency of exposure than on the level of complexity (see section 1.5 below).

The question then is how this affects parsing, i.e. whether lexical frequency guides the online formation of the syntactic representation. It has been debated whether or not lexical frequency of e.g. verb-bias (selectional properties) is part of the lexical information which is directly available to inform online parsing strategies to resolve structural ambiguity (see e.g. Trueswell 1996; Bever

et al. 1998; Pickering et al. 2000). However, the picture that emerges is that lexical frequency does not seem to have a clear, uniform (across the board) effect on processing, but instead seems to interact with other factors, including word class and word length. (For a brief discussion, see Newmeyer 2003: 695–698.)

## 1.2 Construction frequency and acceptability

Just like lexical frequency, the connection between acceptability and construction frequency (how often certain types of syntactic structures occur in corpora) seems to be less clear than one might expect. On the one hand, a close connection between construction frequency and acceptability rating has been reported by e.g. Bybee & Eddington (2006), Featherston (2008), and Kempen & Harbusch (2008). However, the study by Kempen & Harbusch (2008) on scrambling in German also shows that structures that received high acceptability ratings showed considerable variation in frequency of occurrence. Both Featherston (2008) and Kempen & Harbusch (2008) also found that structures with the highest acceptability ratings were also the most frequent ones, whereas differences in lower acceptability ratings were not reflected in the frequencies. Similar results were found by Bader & Häussler (2010), who also found that acceptability cannot be predicted from low construction frequency: Sentences with high acceptability ratings span a frequency range from high to very low (close to zero). Furthermore, very low frequency was associated with a wide range of acceptability ratings, from low to high.<sup>1</sup> In a study on different types of *wh*-questions in French, Adli (2015) found that acceptability ratings and frequency of occurrence in spontaneous speech did not correlate. Though all types of *wh*-questions (nine in total, incl. *wh* in-situ and *wh*-fronting with and without subject-auxiliary inversion) were found to be fully acceptable, some of them were very common, others very rare. In addition, those with the highest acceptability scores were those which were considered more formal (or not colloquial), even though they had lower frequencies.

Overall, the empirical evidence shows that there seems to be a relationship between how acceptable people find a given sentence, and how frequent the individual lexical words (but not the function words) are (section 1.1), and how frequent the given construction is (at least if the construction is not very rare, and if the acceptability rating is not at ceiling, Bader & Häussler 2010).<sup>2</sup> This relationship, it seems, is more complicated than a simple positive correlation between frequency and acceptability.

---

<sup>1</sup> See in particular Bader & Häussler's (2010: 314) figure 8.

<sup>2</sup> Cf. Also that satiation effects (that is, increased acceptability as a function of repeated exposure) seem to only apply to sentences with intermediate acceptability ratings, not those at floor or ceiling (Sprouse 2008; Brown et al. 2021; Christensen et al. 2013a).

### 1.3 Syntax vs. semantics

In order to argue that different levels of structural complexity itself might be a determining factor on acceptability independently of frequency of construction, lexical frequency, or indeed meaning, the first step is to determine whether it is possible to disentangle the semantics (the meaning, the individual lexical item) from syntactic structure (the construction). Indeed, evidence from neurolinguistic and psycholinguistic research suggests that this is actually possible.

Brain imaging studies have shown repeatedly that syntactic processing, as well as (some) semantic processing, engages the left inferior frontal gyrus, LIFG (also called Broca's area) (e.g. Stromswold et al. 1996; Grodzinsky & Friederici 2006; Christensen & Wallentin 2011; Pallier et al. 2011; Grodzinsky et al. 2021). Furthermore, it has also been argued that the posterior superior part (pars opercularis, BA 44) is involved in syntactic processing, whereas the frontal superior part (pars triangularis, BA 45) is involved in thematic integration (e.g. Dapretto & Bookheimer 1999; Newman et al. 2003; Clos et al. 2013; Zaccarella et al. 2015; Schell et al. 2017). From a neurolinguistic point of view, there is ample evidence to support the hypothesis that semantic and syntactic information are processed differently and can be disentangled (see also Friederici 2002). In other words, in language comprehension, the structure itself also matters, not just the meanings of the words, and the associated brain activation is a function of complexity: The localization depends on the nature of the complexity, and the amount of activation is determined by the level of complexity.

Psycholinguistic studies also suggest that semantic information and syntactic information are processed differently. According to Bever et al. (1998) semantic information is processed before syntactic information. In addition, probabilistic information (i.e. frequency) is used to propose an initial semantic representation which forms the basis for the syntactic analysis. In particular, the frequency of the construction (the frequency with which the syntactic representation or construction occurs) constrains the parser such that at all times, it chooses the most likely interpretation. However, other studies have shown that syntactic information has priority over semantics, and that temporary (intermediate), implausible (semantically incongruous) interpretations are made during parsing, but only if such analyses are compatible with syntactic selectional properties of the verb (Kizach et al. 2013; Nyvad et al. 2014; cf. also Clifton et al. 2003; but see Trueswell 1996).<sup>3</sup> Furthermore, studies have also shown that (at least some kinds of) grammatical illusions, such as *More people have been to Paris than I have*, are parsed via direct syntactic analysis in accordance with syntactic properties, not via semantic coercion (Christensen 2016; but see also Wellwood et al. 2018). Similarly, “nonliteral interpretation” of implausible,

---

<sup>3</sup> Compare *Sally knew/presumed Tom was ill*, where *knew* is compatible with either a DP object or a clausal one, whereas *presumed* is only compatible with a clausal object. With *knew*, *Tom* is temporarily attached as the object of *knew*, and subsequently reanalyzed as the subject of the embedded clause. With *presumed*, such an intermediate attachment is blocked.

anomalous sentences, such as *The mother gave the candle the daughter*, have been shown to reduce structural priming compared to well-formed double object constructions, e.g. *The mother gave the daughter the candle* (Cai et al. 2022). Interestingly, the reduction in priming was larger for the corresponding anomalous DP-PP versions, such as *The mother gave the daughter to the candle*. The authors argue that their results (showing significant interaction between structure and plausibility) support the hypothesis that language comprehension involves constructing a syntactic representation which forms the basis for semantic decoding, and that their results are incompatible with a semantic approach, according to which meaning can be inferred from semantic relations among the words.

In sum, the studies discussed in this section suggest that meaning and structure (construction) can be teased apart, and that parsing is primarily, though not necessarily exclusively, guided or informed by syntactic properties, rather than by semantic ones. These findings favor a structure-based account and strongly suggest that if we want to understand what determines how and why speakers find some sentences more acceptable than others, and why some sentences are more difficult than others, we need to take into account their structural properties.

#### 1.4 Usage-based approaches

According to usage-based approaches, linguistic rules (i.e. the grammar of the ambient language) are structural regularities emerging from learners' lifetime analysis of the distributional characteristics of the language input. Moreover, it is assumed that acceptability judgements reflect the particular frequencies of the speaker's "accidental experience" (Ellis 2002; see also Real & 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). In other words, the idea is that the sentence types that people find most acceptable are the ones that they frequently hear or have frequently heard. From this, it seems to follow that acceptability patterns are expected to show a high degree of inter-individual variation, since a lifetime of "accidental experience" plays out at the level of the individual speaker. Ellis (2002: 161–162) also argues that a speaker's grammar is "a statistical ensemble of language experiences that changes slightly every time a new utterance is processed" and that intuitions about grammaticality are "unstable" and "fluid". However, there is ample evidence that intuitions are stable across as well as within speakers, and that while intuitions do change slightly with exposure, such repetition (priming, satiation, or trial) effects are constrained by the grammar. Not all constructions are affected to the same degree, quantitatively as well as qualitatively, if at all (Snyder 2000; Sprouse & Almeida 2017; Snyder 2022). Furthermore, as we saw in section 1.2, while high frequency is generally correlated with high acceptability, low frequency does not entail low acceptability. A construction or sentence

type may be rare or indeed nonexistent for a number of reasons, including high complexity, formality, or ungrammaticality, or indeed there might be an accessible simpler alternative (e.g. *I expect it to rain* instead of *I expect that it will rain*).

## 1.5 Grammar-based approaches

Contra the usage-based accounts in section 1.4, others have argued that grammatical principles interact with and often override frequency effects, e.g. in ambiguity resolution (Pickering et al. 2000; Bornkessel et al. 2002) and acceptability of subcategorization (White & Rawlins 2020). It has also been shown that sentences with derived (non-canonical) word order are difficult to comprehend and produce for people with agrammatism, independently of construction frequency and information structure (Bastiaanse et al. 2009). For example, in Dutch, sentences with the finite verb in the second position (main clauses with V2) are more frequent than sentences with the verb in final position, and scrambled order is as frequent as base order. However, sentences with V2 and scrambling (all highly frequent) are more impaired than sentences with base (canonical) orders (which have lower frequency). Indeed, agrammatism is also characterized by impaired or non-use of function words, such as determiners, pronouns, prepositions, complementizers – the most frequent types of words (Damasio 1992).

In language acquisition, children show production patterns that deviate robustly from the input they receive from their surroundings. Children tend to ignore the speech errors they hear from others (positive evidence), and they also ignore when the errors they make themselves are corrected (negative evidence). Furthermore, they produce overgeneralizations and syntactic structures that are neither in the target grammar nor in the actual input they get (Thornton & Crain 1994; Pinker 2004; Yang 2004). In their earliest productions, children also show a preference for structurally simpler, less frequent forms over more frequent, but more complex forms. For example, in Norwegian, the phrase corresponding to *my car* can be constructed either as *min bil* ('my car'), the simplex version, which has low frequency, or as *bil-en min* ('car-the my'), highly frequent, but complex, and derived by fronting *bil* (*bil<sub>1</sub>-en min t<sub>1</sub>*). Norwegian children show a production preference for the former (simplex structure, low frequency) over the latter (complex structure, high frequency) (Anderssen & Westergaard 2010).

The frequency of the input not only underdetermines what aphasics and children actually produce, but also what language users do *not* produce. As Yang (2015: 290) puts it: "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".

Based on the discussion above, we take the acceptability rating given by a participant to be a function of the total amount of difficulty experienced by the comprehender due to various types of complexity. We thus take acceptability judgments to reflect processing difficulties (Fanselow

& Frisch 2006; Hofmeister & Sag 2010; Christensen et al. 2013a), and therefore, we predict that acceptability ratings depend less on construction and/or word frequencies than on complexity factors (grammar-based processing factors), which we now turn to.

## 1.6 Five factors in structural complexity and graded acceptability.

We take it to be fairly uncontroversial that the list of factors that increase the ‘complexity’ of a sentence is long, but also that it includes morpho-syntactic structure and clause type, constituency, noncanonical word order (derived word-orders), number, type, and length of dependencies (e.g. extraction or movement, and coreference), argument structure, number of propositions, pragmatics, finiteness, etc. Here, we will focus on the following five factors which we will employ as our complexity metric: clausal embedding, adjunction, move-out (crossing a clause boundary), number of fillers (moved elements), and path (structural distance between filler and gap in movement).

**Clausal embedding** increases structural complexity by increasing the number of constituents, and it is well-known that structural expansion increases working-memory (WM) load and, hence, the processing load (Hawkins 1994; Pallier et al. 2011), in particular finite clauses, which are full propositions with full syntactic representations, and which arguably constitute derivational phases, cf. the *Phase-Impenetrability Condition* (PIC) (Chomsky 2001: 13). (Compare e.g., *Someone [wearing a tinfoil hat] was arrested* and *Someone [who was wearing a tinfoil hat] was arrested*.) People with agrammatism also have production problems with finiteness and with embedded clauses (Friedmann & Grodzinsky 1997; Friedmann 2001; 2003). In this paper, by embedding, we refer to the presence of two propositions, one embedded in the other, not simple clefting (e.g. *It was someone who was wearing a tinfoil hat*).

**Adjunction**, the process of modifying one phrase with another (e.g. *the* [<sub>NP</sub> [<sub>AdjP</sub> *completely untuned*] [<sub>NP</sub> *electric guitars*]]), always increases the number of XPs (i.e. phrasal constituents). In addition to the combined structure of the two clauses involved, adjunction creates another instantiation of the modified constituent (unlike complementation), cf. the trees in (7)–(10) below. As such, it also increases the syntactic WM load. Extraction, e.g. by *wh*-movement, from an adjoined position (typically adverbials such as *where*, *when*, or *why*) is more difficult to process than extraction from a position selected, e.g. by a lexical verb such as *see* or *read* or by a preposition such as (*talk*) *to* or (*point*) *at* (typically *what*, *which*, or *who*) (Nyvad et al. 2014). A gap (or trace of movement) in an adjoined position is more difficult to reconstruct than a gap in a selected position, cf. the *Empty Category Principle* (ECP) (Haegeman 1994: 442). Furthermore, adjunct clauses are often assumed to be islands, that is, structural configurations that (to varying degrees) block extraction, cf. the *Condition on Extraction Domain* (CED) (Huang 1982: 505; but see Müller 2019; Bondevik et al. 2020; Nyvad et al. 2022). In agrammatism, patients seem to have an adjunction deficit which means that they tend to avoid modifiers and prefer to produce



predicative adjectives over attributive ones (Lee & Thompson 2011; Meltzer-Asscher & Thompson 2014). In an EEG study on healthy speakers of Hebrew, Prior & Bentin (2006) investigated the electrophysiological effects of anomalous or incongruent nouns (arguments), verbs (predicates), and adjectives (adjuncts modifying a noun phrase). The results showed that incongruent adjectival adjuncts (inside NPs) induced much smaller (or nonsignificant) N400s than incongruent verbs and nouns. This, again, suggests that complements and adjuncts are processed differently, even when they are not clauses. All in all, adjunction (modification) increases complexity, and adjuncts have different structural and semantic properties than complements.

Extractions are in themselves costly, as they create long-distance dependencies and fillers that have to be kept in working memory until they can be integrated at the gap site. As mentioned above, sentences with derived word order also cause processing problems for people with agrammatism (Bastiaanse et al. 2009). Derived word orders tend to have lower acceptability, and the longer the **path of movement** (the structural distance between filler and gap), the lower the acceptability (Clifton & Frazier 1989). Here we follow Hawkins (1999: 248–249) and measure length of path or distance in terms of 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’) (see also Collins 1994: 56; O’Grady et al. 2003: 435; Christensen et al. 2013a). In short, only overt XPs count for length of path, while non-overt ones (gaps/traces, non-overt subjects, and non-overt operators) do not. In addition to path, the relative ease or difficulty of extraction also depends on **the number of fillers** (i.e. extracted elements), compare *They did the job* (no filler), *What did they do?* (1 filler: *what*), and *Why did they ask which job to finish?* (2 fillers: *why* and *which job*).

Extraction from an embedded clause, which we here refer to as ‘**Move-Out**’, following Christensen et al. (2013b), is particularly costly in terms of processing, because it creates a long-distance dependency spanning two clauses (i.e., two propositions). As mentioned, extraction from an adjunct clause is often difficult or perhaps even impossible, resulting in (various degrees of) reduced acceptability (Nyvad et al. 2022). However, extraction from complement clauses is also costly. In theoretical syntax, it is standardly assumed that extraction proceeds via a position at the left periphery in the embedded clause, CP-spec, cf. the PIC (Chomsky 2001: 13). When this position is occupied, e.g. by a *wh*-element, extraction is impeded, leading to reduced acceptability or ungrammaticality (Christensen et al. 2013a; Christensen & Nyvad 2019). In a brain imaging study with fMRI, Christensen et al. (2013b) found that the effect of increased syntactic processing typically found in and around Broca’s area only reached significance when the extraction crossed a clause boundary. In other words, moving out of the embedded clauses induced a massive increase in activation in Broca’s area and other frontal regions, compared to movement targeting a position within the embedded clause.



As the discussion of these five factors show, there is a close connection between structural complexity on the one hand, and processing load and acceptability on the other, and that the former causes the latter seems to be the standard assumption in psycho- and neurolinguistics, at least since Miller and Chomsky (1963) showed that unacceptability of a structure that does not violate any syntactic constraints may be the result of processing complexity. In theoretical syntax, the preference for structural simplicity is linked to computational economy and it is captured in the principle of *Economy of Derivation*: “make derivations as short as possible, with links as short as possible” (Chomsky 1995: 91; cf. also the processing accounts of Hawkins 1994; 2004; Gibson 1998). In turn, this seems to be the central tenet of the *Derivational Theory of Complexity* (Fodor & Garrett 1967). As noted by Christensen (2005: 308): “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)]” (see also Hornstein 2014).

## 1.7 Animacy

One of the non-structural factors that has been argued to interact with morpho-syntax and, hence, to affect constraints on word order and the level of acceptability, is animacy, i.e., whether an NP denotes a living being or not. Furthermore, there seems to be a preference to maximally differentiate subject and object, especially if both are animate, and even more so if the object is a topic. According to Aissen (2003), many languages have differential object marking, which means that the case marking of the object of a transitive verb depends on the animacy or definiteness of that object. The prototypical subject is Agent (hence, also animate) and Topical, and to maximally differentiate subject and object, the higher on the animacy scale (Human > Animate > Inanimate) and/or the definiteness scale (personal pronoun > proper name > definite NP > indefinite specific NP > non-specific NP) the object is, the more likely it is to be case-marked. The point here is that animacy, which is a semantic property, influences or interacts with the grammatical constraints on word order and case marking.

Animacy has also been shown to be relevant for superiority effects in *wh*-movement. In a multiple question, that is, a clause with two *wh*-elements, it is more acceptable to front the highest *wh*-element and leave the second, lower one in situ than vice versa, where extraction of the lower *wh*-element crosses the position of the higher one, compare *Who will buy what?* and *\*What will who buy* (Clifton et al. 2006). This superiority effect is stronger in English (where it leads to ungrammaticality) than in German (where it only lowers the acceptability) (Fanselow et al. 2011; Häussler et al. 2015). In German, however, there is a significant interaction between word order and animacy. Fronting a *wh*-object across a *wh*-subject is significantly less acceptable when both are animate (hence, not maximally distinct) than when they differ in animacy. In

short, fronting a *wh*-object across a *wh*-subject is significantly less acceptable when the two are both animate (*They knew who<sub>1</sub> who criticized t<sub>1</sub>*) than when the subject is animate and the object inanimate (*They knew what<sub>1</sub> who criticized t<sub>1</sub>*) (Fanselow et al. 2011; Häussler et al. 2015).<sup>4</sup>

Animacy has also been shown to interact with relativization (Gennari & MacDonald 2009). Relative clauses with an inanimate antecedent and an animate subject (*The movie that the director watched received a prize*) are easier to process than ones with an animate antecedent and an inanimate subject (*The director that the movie pleased received a prize*).

In an fMRI study on German simplex transitive clauses with scrambling, Grewe et al. (2006) found that sentences where an inanimate subject preceded an animate object or where an inanimate object preceded an animate subject (both inanimate > animate) increased activation in Broca's area in the brain (compared to their animate > inanimate counterparts), presumably reflecting an increase in processing load. However, though this fixed/main effect was significant, they did not find any significant interaction between extraction (or word order) and animacy, neither in the acceptability ratings, nor in the brain signal. The animacy effect in Broca's area was also only detectable using a so-called region-of-interest analysis. Furthermore, in the behavioral data, there were no fixed/main effects of order or animacy on acceptability either. At the very least, these complications suggest that the overall effect of animacy might be subtle.

In short, it is possible that animacy is relevant to the acceptability of extraction, such that at least part of a potential reduction in acceptability stems from having an inanimate argument before an animate one, as in Grewe et al. (2006). Alternatively, it could be that having both an animate subject and an extracted animate object leads to reduced acceptability, as in Fanselow et al. (2011). In either case, we would expect to see a significant (main or interaction) effect of animacy.

## 1.8 Predictions

Based on the review of frequency effects and complexity factors, we made a number of experimental predictions, all of which are based on processing factors:

- I. Acceptability decreases as complexity increases. Because structural complexity, which is a function of (at least) the five factors discussed above (embedding, adjunction, Move-Out, path/distance, and number of fillers), increases processing cost, it is negatively correlated with acceptability.
- II. Acceptability is also predicted by construction frequency, but the correlation is weaker. (Low acceptability is not correlated with zero frequency, or vice versa.) We predict that

---

<sup>4</sup> Fanselow et al. (2011) did not control for the ambiguity of the German word *was* (it can mean 'something' or 'what'), which may have contributed to the effect.

acceptability can best be described as a function of processing factors (see prediction I), not by construction frequency.

- III. The level of acceptability is 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 the ‘baseline’ acceptability is determined by complexity.
- IV. The acceptability of object extraction interacts with the animacy of the object.

To test **prediction I** and **prediction II** (acceptability is a function of complexity rather than construction frequency), we included sentences with different structural complexities, each with and without extraction/fronting of an object: simplex (mono-clausal) sentences, sentences with an embedded complement clause, sentences with a clausal adjunct, sentences with an embedded relative clause, and a set of ungrammatical sentences (see the Materials section below). Relative clauses were included because they are grammatical and highly complex, and because extraction from relative clauses significantly lowers acceptability (but to varying degrees). We also included so-called parasitic gaps (Engdahl 1983), sentences with a clausal adjunct and one extracted filler but two gaps (i.e. they are also highly complex), because they are notorious for being acceptable in spite of apparently being very rare (Mayo 1997; Newmeyer 2003; Phillips 2013; Boxell & Felser 2017; Momma & Yoshida 2023).<sup>5</sup> In this way, we had fully grammatical sentences with high, intermediate, and low frequency, plus ungrammatical ones with low (zero) frequency, which enabled us to test for the effects of complexity and construction frequency. Relative clauses are interesting because it has been shown that in Danish, the acceptability rating is affected by extraction as well as by lexical frequency of the matrix verb (Christensen & Nyvad 2014).

To test **prediction III** (lexical frequency can affect the average acceptability of grammatical structures), we constructed the sentences from a set of lexical items controlled for low, middle, and high frequency of occurrence in a Danish text corpus (KorpusDK).<sup>6</sup> This procedure ensures that we had examples with continuous frequency across the full frequency span from very low to very high.

To test **prediction IV** (animacy interacts with extraction), we constructed our stimuli such that while all the sentences had an animate subject, half of them had an animate object [+Anim], the other half an inanimate one [-Anim].

---

<sup>5</sup> For example: “I believe that these are rare in actual speech, though I do not know of any statistical studies to confirm that claim” (Newmeyer 2003: 691). “Linguists have sometimes described parasitic gaps as a marginal phenomenon, but controlled judgment experiments confirm that they are very real” (Phillips 2013: 72). (See also Potts 2004; Resnik 2004).

<sup>6</sup> KorpusDK contains a collection of electronic texts derived from a wide variety of different sources from 1983-2002, incl. newspapers, various types of popular magazines, and literary texts. In total, it contains 56 million words.

## 2 Experiment

### 2.1 Materials

The stimulus consisted of 24 sets of four conditions with increasing complexity plus one filler condition (A–E), each with and without object extraction, corresponding to (2)–(6) below, resulting in 240 items in total. The sentences were each preceded by a context to make them as natural as possible and to maximally facilitate acceptability (following the approach in Nyvad et al. 2022). The context as well as the target sentences were carefully constructed to be as coherent as possible.<sup>7</sup> Each set shared the same unique context, as in (1). Half of the sentences had an animate object [+Anim], the other half an inanimate one [–Anim], all with an animate, pronominal subject (*jeg* ‘I’ / *han* ‘he’ / *hun* ‘she’). The main verb in condition A (simplex, cf. (2)) was used as either a matrix or an embedded main verb in conditions B–D (complex, (3)–(5)). All object nouns were definite and non-compound. High-frequency discourse particles (PRT) were inserted to make the sentences more natural and colloquial (e.g. *jo*, *da*, and *bare*, which can be translated roughly as ‘after all / you know’, ‘at least’, and ‘just / simply’, respectively). (The entire stimulus set, including frequencies, is available on the OSF repository, see the ‘Data availability’ section below.)

(1) **Context:**

*Min bedste vens nye kæreste vil gerne have at jeg skal møde hans  
My best friend’s new girlfriend will preferably want that I shall meet his  
datter og hans hund i weekenden.  
daughter and his dog in weekend-the  
‘My best friend’s new girlfriend would like me to meet his daughter and his dog  
this weekend.’*

(2) **Simplex**

- |    |   |    |
|----|---|----|
| a. | <i>Men jeg kender da allerede barnet.</i>           | A1 |
|    | <i>But I know PRT already child-the</i>             |    |
|    | <i>‘But I already know the child.’</i>              |    |
| b. | <i>Men <u>barnet</u> kender jeg da allerede __.</i> | A2 |
|    | <i>But child-the know I PRT already</i>             |    |
|    | <i>‘But the child, I already know.’</i>             |    |

(3) **Complement clause**

- |    |   |    |
|----|---|----|
| a. | <i>Men jeg tror da allerede [at jeg kender barnet].</i> | B1 |
|    | <i>but I believe PRT already that I know child-the</i>  |    |
|    | <i>‘But I believe I already know the child.’</i>        |    |

---

<sup>7</sup> It is conceivable that the fit between context and target sentence was consistently less good for sentences with object extraction (A2, B2, C2, and D2) compared to sentences without it (A1, B1, C1, D1). However, we see no reason to assume so, given that the coherence between sentence and context was carefully checked by three native speakers. We argue here that the reduced acceptability is explained independently by complexity factors.

- b. Men barnet tror jeg da allerede [at jeg kender \_]. B2  
*but child-the believe I PRT already that I know*  
 ‘But the child, I believe I already know.’

(4) **Adjunct clause**<sup>8</sup>

- a. Men jeg kender da allerede barnet [uden at have mødt hende]. C1  
*But I know PRT already child-the without to have met her*  
 ‘But I already know the child without having met her.’
- b. Men barnet kender jeg da allerede \_ [uden at have mødt \_]. C2  
*But child-the know I PRT already without to have met*  
 ‘But the child, I already know without having met.’

(5) **Relative clause**

- a. Men jeg kender da allerede én [OP der \_ har mødt barnet]. D1  
*But I know PRT already someone that have met child-the*  
 ‘But I already know someone who has met the child.’
- b. Men barnet kender jeg da allerede én [OP der \_ har mødt \_]. D2  
*But child-the know I PRT already someone that have met*  
 ‘But the child, I already know someone who has met.’

(6) **Controls/fillers**

- a. Men jeg kender da allerede hunden og barnet. E1  
*But I know PRT already dog-the and child-the*  
 ‘But I already know the dog and the child.’
- b. \*Men barnet jeg kender hunden og da allerede. E2  
*But child-the I know dog-the and PRT already*

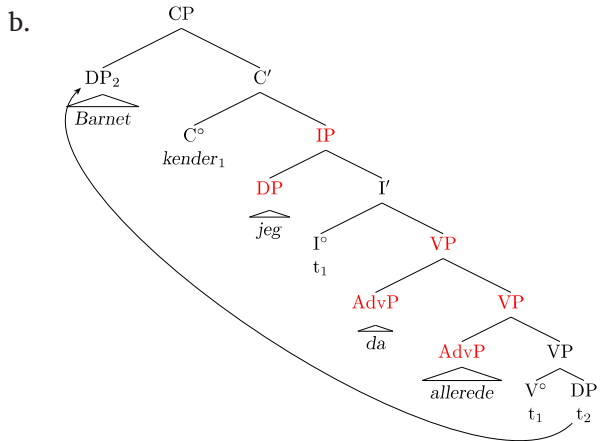
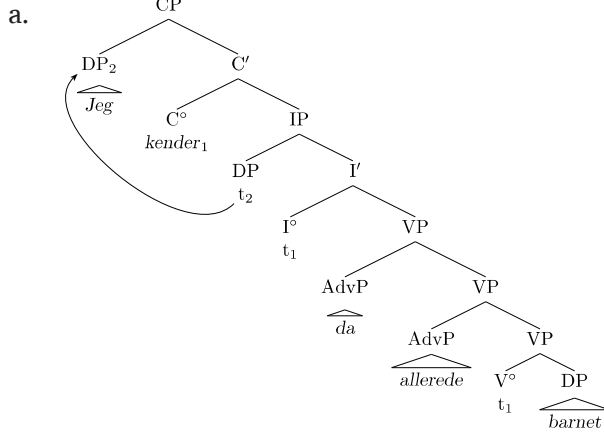
The ungrammatical controls/fillers (E2), as in (6)b, had scrambled, impossible word order and were ungrammatical for three reasons: (i) There is extraction from the second conjunct of a coordinated DP (cf. \**What did they like cookies and?*), (ii) The order violates the verb-second constraint on Danish main clauses, and (iii) the (remnant of) the object and the particle *da* and the adverb *allerede* are in the wrong order.

The syntactic structures of conditions A–D are given in (7)–(10). The XPs that count in the sum measurement of distance/path between filler and gap in extraction are marked with red. Recall that only overt XPs count (Hawkins 1999: 248–249).

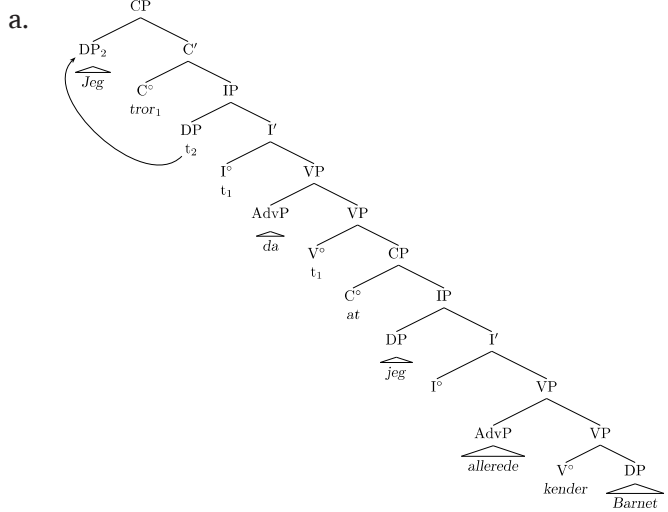
---

<sup>8</sup> Note that the facilitating context ensures that the object in the matrix clause (*barnet*) is coreferent with the object in the adjunct clause (*hende*), both in the non-extracted version C1 and in the extracted version C2.

(7) **Simplex (conditions A1–A2)**

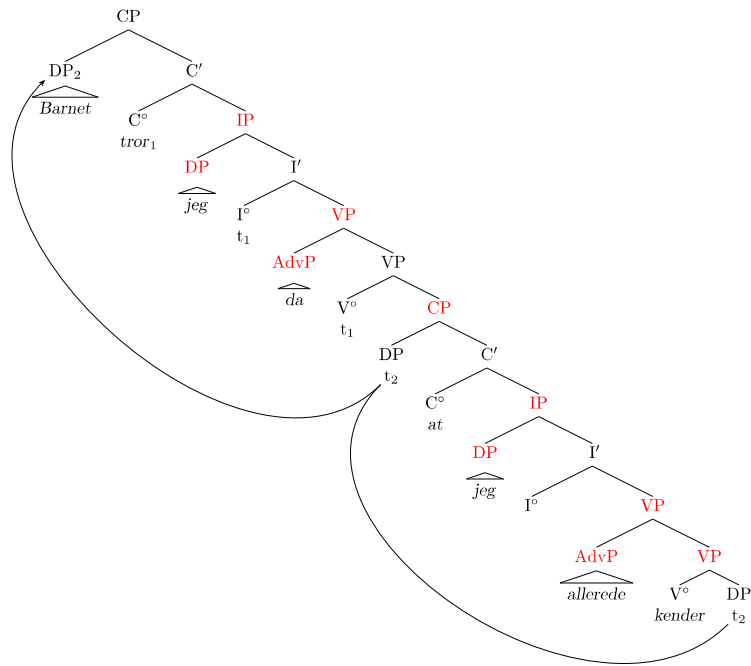


(8) **Complement clause (conditions B1–B2)**



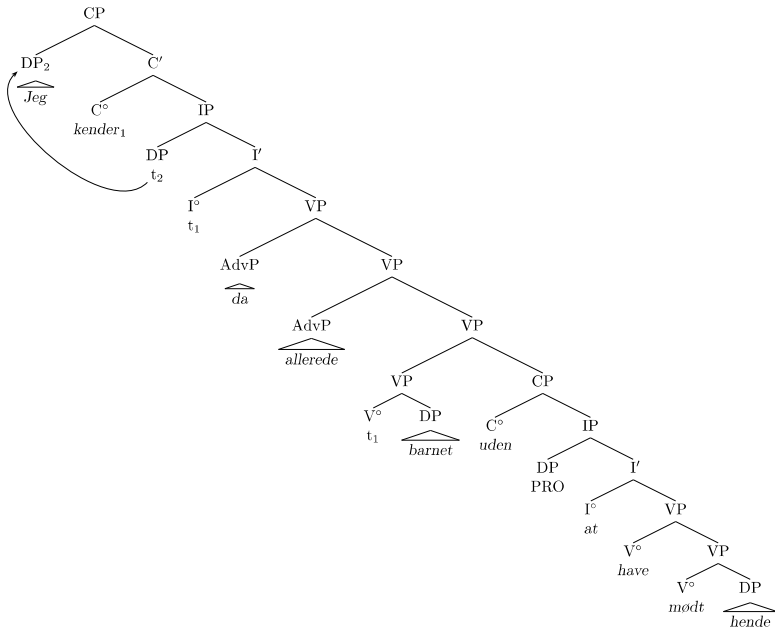


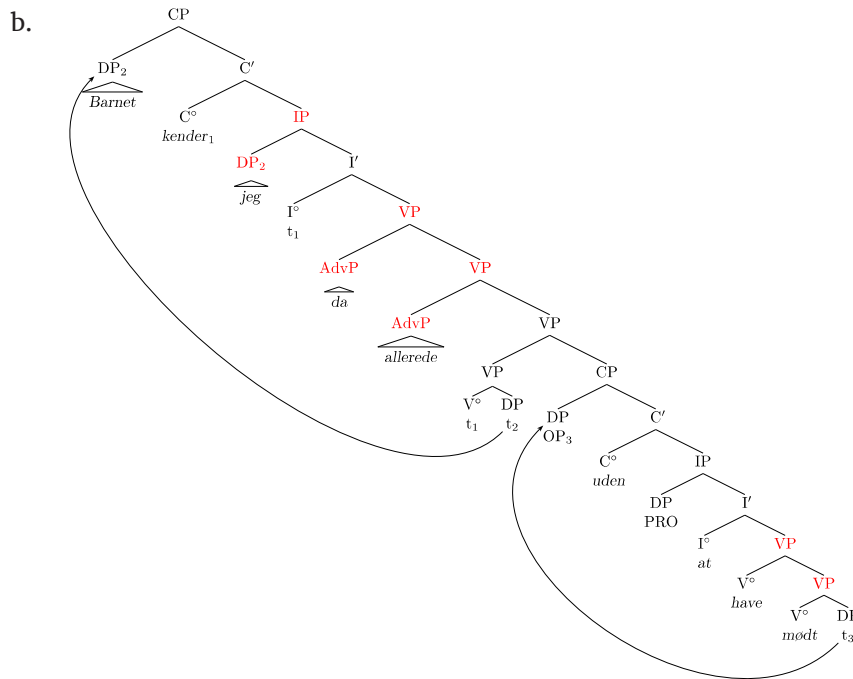
b.



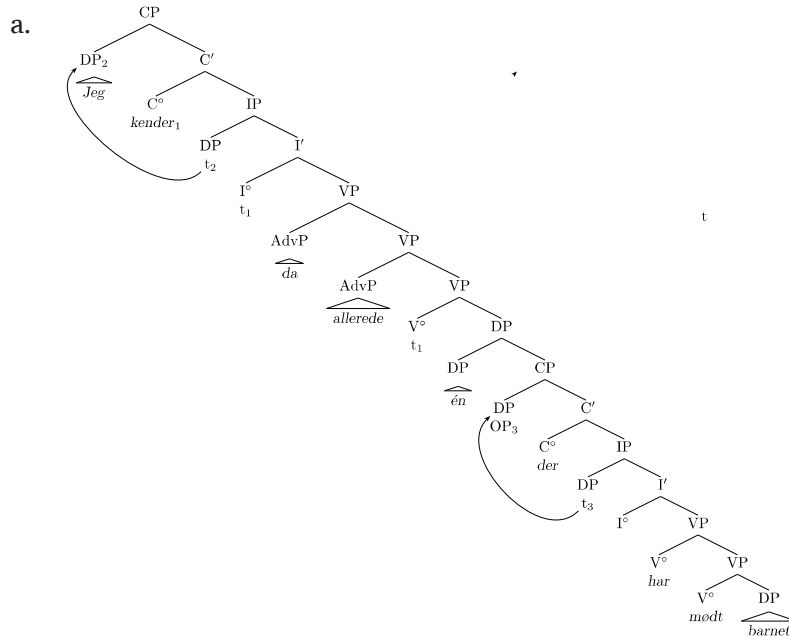
(9) **Adjunct clause (conditions C1-C2)**

a.

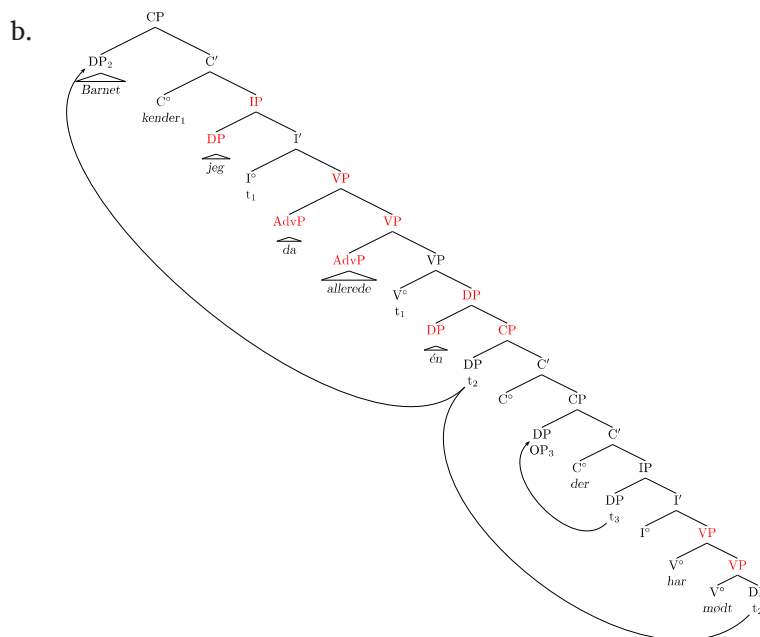




(10) Relative clause (conditions D1–D2)<sup>9</sup>



<sup>9</sup> In Danish, it is perfectly acceptable to attach a relative clause to a pronoun, e.g. *Jeg kender ham der bor her* (lit. ‘I know him who lives here’). This is unlike English, where it seems to be generally unacceptable, except in formal, poetic, or biblical texts.



The complexity of each of the four target conditions [ $\pm$ Ex] is summarized in **Table 1**. In order to avoid additional trivial movement steps in the derivation to skew the length of path, and hence, the sum complexity, the number of XPs in the path (the red ones in (2)–(6)) was normalized with z-score transformation (path.z). Otherwise, A2 would be six times as complex as A1, A2 would be 11 times more complex than A1, etc. The z-transformed Path also has the positive effect of (almost) neutralizing the complexity of fronting the subject (i.e. the filler effect), which results in canonical word order. In **Table 1**, the numbers of XPs used in the calculation are provided in parentheses.

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

**Table 1:** Under Path.z, the numbers in brackets refer to the number of XPs that are crossed in extraction (cf. the red XPs in tree structures in (7)–(10) above).

Based on **Table 1**, the conditions can be arranged in a complexity hierarchy from least to most complex:

(11) Complexity: A1 < B1 < A2 < C1 < D1 < B2 < C2 < D2

Note that in this paper, extraction is used as a short hand for object extraction unless explicitly specified otherwise. Note also that the difference between subject extraction and object extraction can be reduced to a difference in length of movement path (compare, for example, the two trees in (7)).

To investigate the effect of lexical frequency, the stimuli was balanced over three (relative) frequency zones within each condition, based on the frequency of occurrence of each lemma form of the main verb, medial adverbial, and object noun in a Danish corpus (KorpusDK). We used corpus frequency as a proxy for individual exposure (which is inaccessible to us in the absence of comprehensive longitudinal data). The same set of main verb, medial adverb, and object noun was used to construct a set in one of the frequency zones. For example, in (2)–(6) above, we used *kender* ‘know’ (freq. = 25,215), *barnet* ‘the child’ (freq. = 62,538), and *allerede* ‘already’ (freq. = 28,232) to construct a set in the high frequency zone. In the analysis, we used the mean of the  $\log_{10}$  transformed frequencies of verbs, nouns, and adverbs to get a single, composite measure of lexical frequency.

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

**Table 2:** Range of frequency of occurrence in KorpusDK (hits in brackets) for main verbs, object noun, and medial adverbial for each frequency zone (low, middle, high). Bottom rows contain the  $\log_{10}$  mean values (across verb, noun, and adverb) and range for the three zones.

To investigate the relationship between the acceptability of a construction type and how common it is, we looked up the frequency of the construction in each of the eight target conditions (A1–D2) in KorpusDK.<sup>10</sup> The search frame for each condition is given in **Table 3**. In A1–A2, the search frame begins with a coordinating conjunction (*og/men* ‘and/but’) in order to avoid error examples where the fronted object is the object of a fronted embedded clause, which would lead to an overrepresentation of OSV orders. The pronouns *den* ‘it (common gender)’, *det* ‘it (neuter gender)’, and *de* ‘they’ are avoided because of syncretism between nominative and accusative case and because they can be used as determiners, which would lead to an overrepresentation of SVO orders. The search frames only provide a subset of possible examples as the number of auxiliary verbs are fixed and adverbials are excluded, so the search results are very conservative (for a different approach, see Gries & Ellis 2015). However, the relevant feature is the *relative* ( $\log_{10}$  transformed) frequencies, not the actual number of hits.

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’) <sup>11</sup> REL (‘that <sub>Subj</sub> ’) 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 <sub>Subj</sub> ’)	1	0.00
E1	= A1	1,165	3.07
E2	–	1	0.00

**Table 3:** Search frame (translated and simplified), number of hits in KorpusDK, and  $\log_{10}$  transformed frequency for each condition (construction frequency). The frequency for C2 (parasitic gap) and E2 (ungrammatical filler) are estimated to be zero, and D2 (extraction from relative clause) resulted in zero hits. The values are set at 1 in order to avoid a  $\log_{10}$  calculation error.

<sup>10</sup> We use the term ‘construction’ to refer to the specific, global sentence types, just like clausal comparative constructions, passive constructions, cleft constructions, etc. This seems to us to be a fairly theory neutral approach. *Wh*-question formation is also typically referred to as a construction type, and so are main clauses with an adverbial clause or a complement clause. These are covered in our analysis with the factors: fillers,  $\pm$  adjunction, and  $\pm$  embedding.

<sup>11</sup> In Danish, the pronoun corresponding to ‘one’ can be realized as either *en*, *een*, or *én*; ‘anyone’ = *enhver*, ‘everybody’ = *alle*, ‘somebody’ = *nogen*.

The search results show that, all things being equal, simple SVO sentences are by far the most frequent type. Furthermore, there are 9.7 times as many simple SVO sentences as corresponding OVS sentences. They also show that there are 49 times as many complement clauses with an in-situ object (and with a simple matrix clause) as corresponding ones with object extraction. Again, these are conservative estimates as they only include sentences with transitive verbs, pronominal subjects and objects, and no adverbials, and where the matrix clause is headed by a complementizer.

The entire set of stimulus sentences was distributed over 10 lists using a Latin-square design, such that each list contained 24 tokens.

## 2.2 Procedure

The experiment was set up as a survey using Google Forms. Participants were recruited through social media platforms and pseudo-randomly assigned to 1 of 10 lists based on their birth month. Each stimulus sentence was preceded by a context, such as (1) above. Participants were instructed to rate the acceptability of each sentence on a 7-point scale ranging from 1 ('completely unacceptable') to 7 ('completely acceptable'), and they were instructed to base their rating solely on the target sentence following the context.

## 2.3 Results

The survey was posted on Facebook and Instagram. A total of 211 people (196 female, 14 male, 1 other) participated in the survey. The mean age was 28.8 years (range = 18–63, std. dev. = 8.8). The number of participants per list ranged from 9 to 33 (32, 30, 33, 11, 9, 13, 18, 27, 22, and 16).<sup>12</sup>

We treat the acceptability rating scale, for which we only provided the endpoints (1 = 'completely unacceptable', 7 = 'completely acceptable'), as an ordinal approximation of a continuous variable, and analyzed the data using parametric models (Carifio & Perla 2008; Norman 2010; Sullivan & Artino 2013). This makes it easier to interpret the effect sizes directly in terms of differences in acceptability ratings. The data was analyzed using *R* version 4.1.2 (R Core Team 2021) with the *lmerTest Package* (Kuznetsova et al. 2017) for mixed effects models and the *MASS Package* (Venables & Ripley 2002) for sliding, pair-wise contrasts. The models in the analysis are the largest models that converged without any errors. All plots were made using the *ggplot2* package (Wickham 2016).

Following Taylor et al. (2022), we also analyzed the data as ordinal with cumulative mixed effects models (*clmm*), using the *ordinal* package for *R* (Christensen 2015) and got the same results as we got with parametric models (linear mixed effects), i.e. the same significant contrasts. We also

---

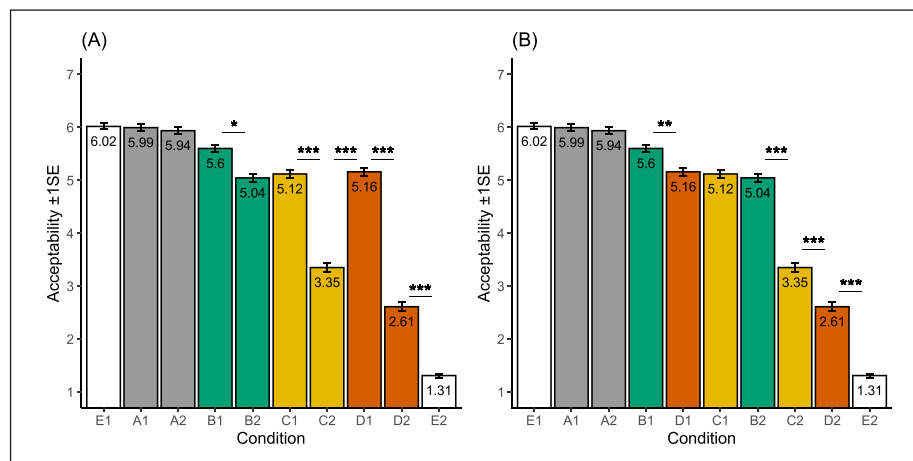
<sup>12</sup> One of the advantages of using linear mixed-effects models, as we did in our analysis, is that they allow unbalanced data, including uneven number of participants per list (Wu 2009; Sonderegger et al. 2018, section 7.1.2).



applied z-transformation to the ratings, which corrects for potential individual participant scale-bias (Sprouse et al. 2012; Nyvad et al. 2022), before analyzing the data with a parametric *lmer* test (Kuznetsova et al. 2017). Again, we got the same overall result. These models are summarized in the supplementary material. In short, the results we present below are highly robust across different types of analysis. Indeed, we had no a priori reason to assume that the results would not be robust, given the number of participants, the number of tokens per type, and the effect sizes. As we have sound reasons for treating the Likert scale as continuous (or as a continuous approximation of an ordinal variable), i.e. that effect sizes are easier to interpret, we have used parametric linear regression in the analysis. Furthermore, as Norman (2010: 631) points out, “parametric statistics can be used with Likert data, with small sample sizes, with unequal variances, and with non-normal distributions, with no fear of “coming to the wrong conclusion””.

### 2.3.1 Condition (as a proxy for complexity)

First of all, to explore the effects of complexity with pairwise contrasts within and between each type (A, B, C, and D), we used a linear mixed effects model with acceptability as outcome variable and condition (as a proxy for complexity), lexical frequency (mean  $\log_{10}$ -transformed), and object animacy as predictors (with sliding contrasts) and random intercepts for participant and item, and random slopes for lexical frequency, animacy, and trial by participant, and random slopes for trial by item. We ran the model twice, first with the condition sorted in the original order, as in **Figure 1 (A)**, then sorted by decreasing acceptability, as in **Figure 1 (B)**, both with sliding pairwise contrasts from left to right. The two statistical models are summarized in **Table 4** and **Table 5**, respectively.



**Figure 1:** Mean acceptability ratings sorted by condition (A) and in decreasing order (B). Conditions A1–A2: Simplex [ $\pm$  Extr], B1–B2: Complement clause [ $\pm$  Extr], C1–C2: Adjunct clause [ $\pm$  Extr], D1–D2: Relative clause [ $\pm$  Extr]; E1: Grammatical filler (= A1), E2: Ungrammatical filler. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

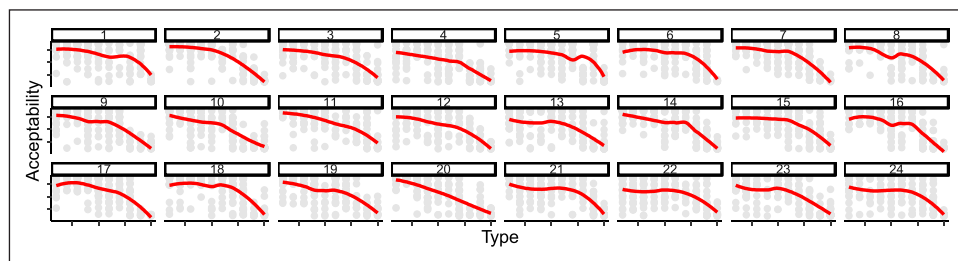
	Estimate	SE	df	t	p	
(Intercept)	4.128	0.200	233.143	20.671	0.000	***
A1-E1	-0.017	0.203	203.963	-0.083	0.934	
A2-A1	-0.094	0.202	194.013	-0.466	0.642	
B1-A2	-0.336	0.203	190.327	-1.657	0.099	
B2-B1	-0.501	0.205	211.811	-2.440	0.016	*
C1-B2	0.015	0.207	217.878	0.073	0.942	
C2-C1	-1.841	0.206	209.727	-8.938	0.000	***
D1-C2	1.730	0.203	203.644	8.504	0.000	***
D2-D1	-2.564	0.204	215.803	-12.539	0.000	***
E2-D2	-1.126	0.205	198.801	-5.480	0.000	***
Lexical Freq.	0.141	0.056	214.195	2.537	0.012	*
Animacy	-0.022	0.094	220.981	-0.239	0.812	

**Table 4:** Summary of the mixed-effects model (sorted by condition, sliding contrasts) with condition as proxy for complexity. A1–A2: Simplex [ $\pm$  Extr], B1–B2: Complement clause [ $\pm$  Extr], C1–C2: Adjunct clause [ $\pm$  Extr], D1–D2: Relative clause [ $\pm$  Extr]; E1: Grammatical filler (= A1), E2: Ungrammatical filler. Lexical Freq. = mean  $\log_{10}$  lexical frequency. Animacy = Object [ $\pm$  Animate]. \*\*\*  $p < 0.001$ , \*  $p < 0.05$ .

	Estimate	SE	df	t	p	
(Intercept)	4.128	0.200	233.464	20.662	0.000	***
A1-E1	-0.017	0.203	203.977	-0.083	0.934	
A2-A1	-0.094	0.202	194.025	-0.466	0.642	
B1-A2	-0.336	0.203	190.147	-1.657	0.099	
D1-B1	-0.597	0.203	185.309	-2.940	0.004	**
C1-D1	0.111	0.205	209.919	0.543	0.588	
B2-C1	-0.015	0.207	217.904	-0.073	0.942	
C2-B2	-1.826	0.206	217.549	-8.850	0.000	***
D2-C2	-0.834	0.206	212.050	-4.050	0.000	***
E2-D2	-1.126	0.205	198.633	-5.480	0.000	***
Lexical Freq.	0.141	0.056	214.367	2.537	0.012	*
Animacy	-0.022	0.094	221.097	-0.239	0.811	

**Table 5:** Summary of the mixed-effects model (sorted by acceptability in decreasing order, sliding contrasts) with condition as proxy for complexity. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

To see if the general acceptability pattern was consistent across the 24 sets of stimulus sentences, we plotted acceptability (sorted by decreasing acceptability, cf. **Figure 1 (B)**) by type (each grey dot represents the average for one participant) for each set. As can be seen in **Figure 2**, the pattern is fairly consistent. To investigate the level of inter-participant variation, we plotted the mean responses from each participant (sorted by decreasing acceptability as in **Figure 1 (B)**). As can be seen in **Figure 3**, the overall pattern is fairly stable across participants (with only very few exceptions). (Both the participant variation and the set variation are already taken into account in the mixed effects model with random factors for participant and item (each item occurs only in one set)).



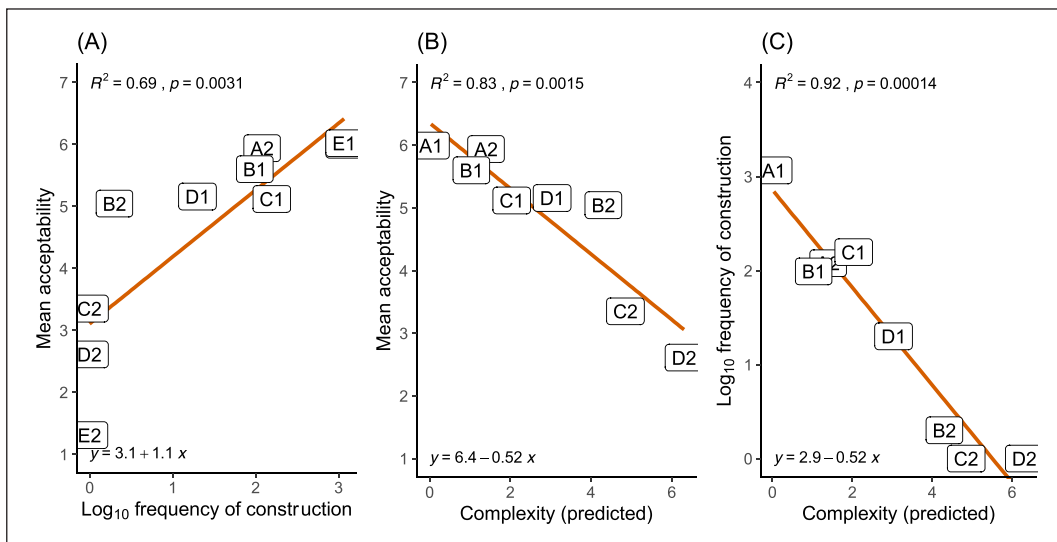
**Figure 2:** Mean acceptability ratings sorted in decreasing order as in Figure 1 (B) for each of the 24 sets (contexts).



**Figure 3:** Mean acceptability ratings sorted in decreasing order as in Figure 1 (B) for each of the 212 participants.

The plot in **Figure 4 (A)** shows acceptability as a function of ( $\log_{10}$ -transformed) frequency of construction. The correlation coefficient for the linear regression, calculated with the *ggpubr* package for R (Kassambara 2020) shows that the effect of construction frequency is large ( $R^2$

= 0.69, describing 69% of the variation) and highly significant ( $p < 0.004$ ). 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 shall argue). Indeed, as **Figure 4 (B)** shows, there was also a significant negative correlation between complexity and acceptability, which has a larger effect size,  $R^2 = 0.83$ , describing 83% of the variation ( $p < 0.002$ ). (In (B), the two fillers (E1 and E2) are excluded, because E1 has the same complexity as A1 (apart from the coordinate object) and E2 is ungrammatical (with scrambled, impossible word order) and therefore it has no meaningful complexity level.<sup>13</sup> Excluding E1 and E2 from the frequency plot in **Figure 4 (A)** only increases effect size from  $R^2 = 0.69$  to  $R^2 = 0.70$ ,  $p < 0.01$ ), still 0.14 lower than the effect size in (B). As **Figure 4 (C)** shows, complexity and ( $\log_{10}$ ) frequency are also strongly negatively correlated ( $R^2 = 0.92$ ,  $p < 0.0002$ ).<sup>14</sup>



**Figure 4:** Acceptability plotted as a function of  $\log_{10}$  transformed frequency of construction (A) and the predicted level of complexity (B), and  $\log_{10}$  transformed frequency as a function of predicted complexity (C). Note that, due to complete overlap, the label for A1 is hidden by E1 in the left panel (A). In (B) and (C), the filler conditions, E1 and E2, are excluded.

<sup>13</sup> It is of course possible for ungrammatical sentences to have measurable complexity levels, for example with resumptive pronouns, as in *That kind of TV, I really don't like it*, or with repeated prepositions, as in *They asked him several questions to which he had no clear answer to*. However, the ungrammatical type E2 in the current experiment is extremely unacceptable and ungrammatical for several reasons, so including a complexity measure here would not really be meaningful.

<sup>14</sup> Note also that this is why frequency of construction could not have been included together with condition in the model Table 4: Collinearity between two factors in the same model causes rank deficiency.

### 2.3.2 Individual complexity factors

In the previous analysis, section 2.3.1, we used condition as predictor, and the measure of complexity used in **Figure 4** was based on **Table 1** where we assigned the same numerical value to each of the complexity factors. Though this may seem a bit arbitrary, it is highly transparent, theoretically motivated, as well as empirically supported, and it is arguably also a conservative and careful approach since there were no a priori reasons for assigning different values to, e.g. embedding and adjunction, either. Furthermore, the fact that this way of tentatively computing relative (sum) complexity results in a very high correlation between complexity and acceptability, which is also in line with theoretical predictions, provides support for doing so. However, it would also be interesting to see if the conservative estimation also matched the relative effect size of each factor.

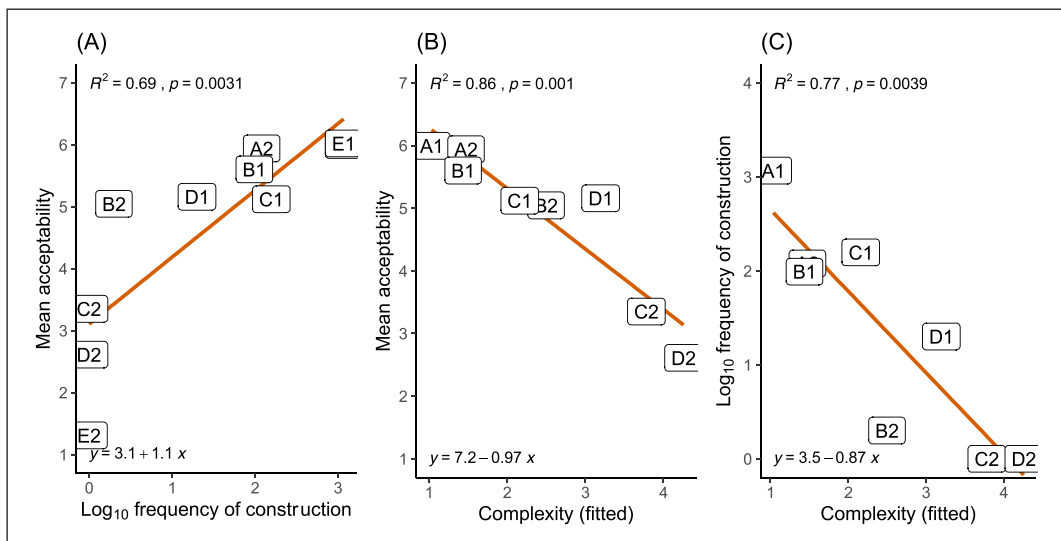
In order to analyze the effect of the individual complexity factors (instead of the sum complexity level), we analyzed the data with a linear mixed-effect model with acceptability as outcome and embedding, adjunction, move-out, fillers, and (raw, not z-transformed) path as predictors. The model included random intercepts for participant and item, and random slopes for the full set of factors plus trial by participant and random slopes for adjunction and trial by item. Again we treated acceptability as an ordinal approximation of a continuous variable and used parametric models (Carifio & Perla 2008; Norman 2010; Sullivan & Artino 2013). We did not know a priori whether the factors would be equally additive or whether they would be super-additive and interact. However, the assumption that they are simply additive makes sense: Adjunction (embedding of adverbial clause) and Move-Out both presuppose embedding in the first place, and both Move-Out and path presuppose fillers (movement creates fillers and gaps). The result is summarized in **Table 6**. (We also analyzed the data as ordinal with a cumulative mixed effects model, and again, we got the same results, i.e. same significant contrasts. See the supplementary material.)

	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	**

**Table 6:** Summary of the mixed-effects model with the five individual complexity factors as predictors of acceptability.

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 of 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 the  $z$ -transformation in **Table 1**: Path affects the relative acceptability, but it does not determine the base level.

We computed new complexity levels by multiplying each factor level (0, 1, or 2) in **Table 1** with the estimates (effect sizes) in **Table 6** and adding them up ( $0.402 \times \text{embedding} + 0.742 \times \text{adjunction} + 0.329 \times \text{Move-Out} + 1.035 \times \text{fillers} + 0.073 \times (\text{raw, not } z\text{-transformed}) \text{ path}$ ). **Figure 5** shows the correlations between this fitted complexity and acceptability (B) and the  $\log_{10}$  frequency of construction (C). ((A) is identical to **Figure 4 (A)**, repeated for convenience.)



**Figure 5:** Acceptability plotted as a function of  $\log_{10}$  transformed frequency of construction (A) and the fitted level of complexity (B), and  $\log_{10}$  transformed frequency as a function of fitted complexity (C). Note that, due to complete overlap, the label for A1 is hidden by E1 in the left panel (A). In (B) and (C), the filler conditions, E1 and E2, are excluded.

Note the correlation between acceptability and complexity is (almost) the same for the predicted complexity levels (from **Table 1**, **Figure 4**) and for the fitted complexity (**Table 6**, **Figure 5**). The correlation between complexity (fitted) and  $\log_{10}$  frequency of construction is still strong ( $R^2 = 0.77$ ,  $p < 0.004$ ), though less strong than predicted complexity ( $R^2 = 0.92$ ,  $p < 0.001$ ).



These results show that our predicted complexity levels were neither completely arbitrary nor off target. The exact relative complexity levels of some of 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.

### 2.3.3 Lexical frequency

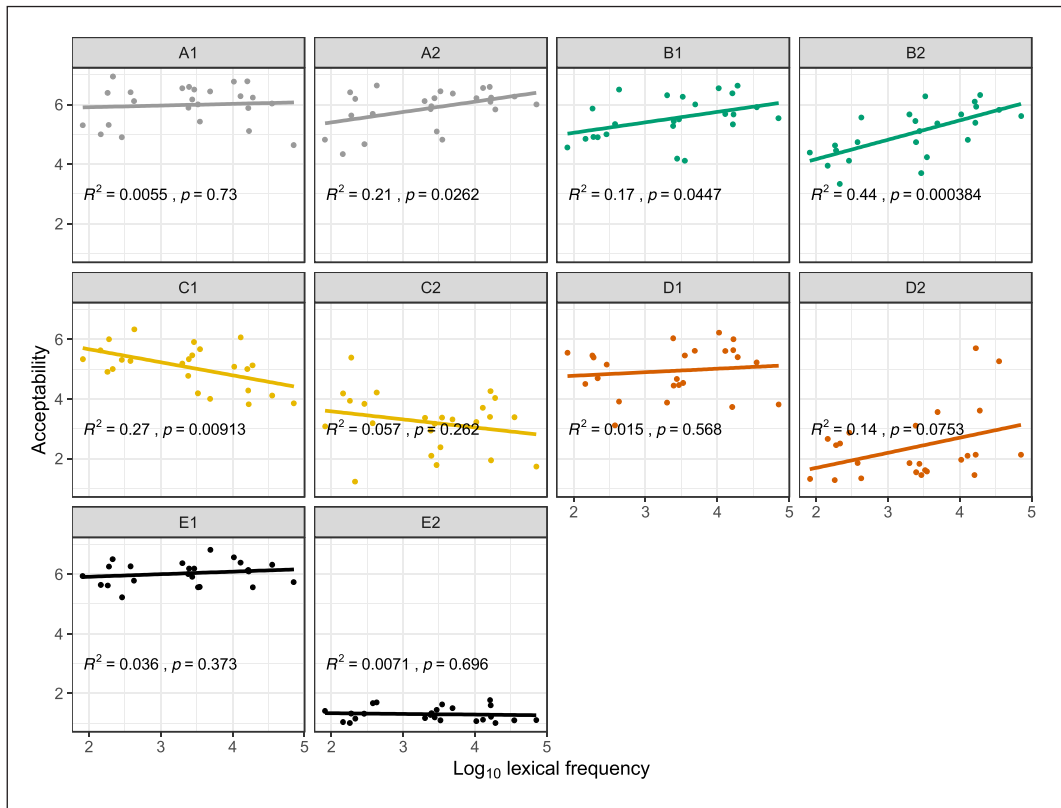
As shown in **Table 4** and **Table 5**, the overall fixed ('main') effect of (mean  $\log_{10}$ ) lexical frequency was significant ( $p = 0.012$ ). However, the effect size (which can easily be read off the estimate, 0.141, because we used a parametric model and treated acceptability as continuous) is very small. When mean  $\log_{10}$  lexical frequency increases by 1 (which means that raw frequency is multiplied by 10), acceptability increases by 0.141. For acceptability to increase by 1, mean  $\log_{10}$  lexical frequency would have to increase by  $1/0.141 = 7.092$ , which means an increase by  $10^{7.092} = 12,359,475$  tokens (i.e. >12 million hits in the corpus). All things being equal, an increase of 1 on the 7-point scale requires quite a massive increase in lexical frequency.

In **Figure 6**, the acceptability of each condition is plotted as a function of the mean ( $\log_{10}$  transformed) lexical frequency (see **Table 2**). For A1 (simplex clauses without extraction), the effect is very small and not significant (same for filler type E1, which is identical to A1, except that the object involves coordination of two nouns). The effects only become larger and reach significance as complexity increases. Comparing A1 and A2, the effect of fronting the object in A2 seems to interact with the lexical frequency ( $p < 0.03$ ), which explains 21% of the variance ( $R^2 = 0.21$ ). For B1 (complement clause without extraction), the effect is also significant ( $p < 0.05$ ), and in B2, the effect of extraction again seems to interact with lexical frequency ( $p < 0.001$ ), explaining a full 44% percent of the variance ( $R^2 = 0.44$ ). Surprisingly, the effect is negative and significant for C1 (adjunct clauses) ( $p < 0.01$ ,  $R^2 = 0.27$ ) and negative, but non-significant for C2 (parasitic gaps). For relative clauses (D1 and D2), the effect is non-significant (though it comes close with extraction in D2,  $R^2 = 0.14$ ,  $p < 0.08$ ). Crucially, however, there is no frequency effect for the ungrammatical filler type E2 ( $R^2 < 0.01$ ,  $p > 0.6$ ). We will discuss these effects further below.

### 2.3.4 Animacy

We also explored the effect of the animacy of the object noun. We ran a separate mixed effects model with acceptability as outcome variable and the interaction between object animacy and extraction as predictor, and random intercepts for participant and item, and random slopes for object animacy, extraction, and trial by participant and random slopes for trial by item. The statistical model is summarized in **Table 7**, and the results are shown in **Figure 7**. Overall, the animacy of the object noun did not have any significant effect on acceptability. The interaction

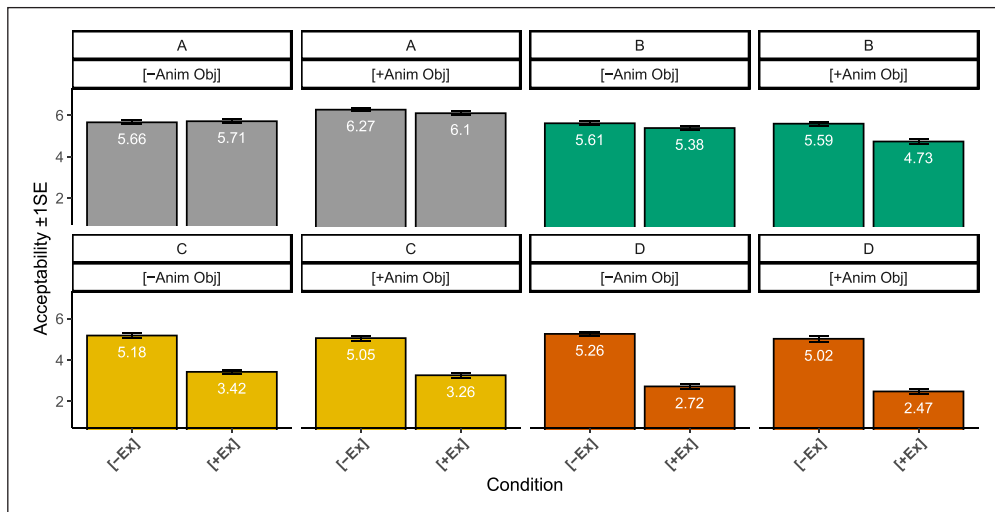
between animacy and extraction was not significant ( $p = 0.692$ ), and neither was the main effect of animacy ( $p = 0.979$ ). Extraction, however, was significant ( $p < 0.001$ ). In other words, the reduced acceptability for sentences with an inanimate object before an animate subject is explained by the effect of extraction; our data did not provide any evidence that animacy affects acceptability.



**Figure 6:** Lexical (mean  $\log_{10}$ ) frequency effects for each condition.

	Estimate	SE	df	t	p	
(Intercept)	5.390	0.196	205.704	27.462	0.000	***
Animacy	0.072	0.263	166.962	0.274	0.785	
Extraction	-1.187	0.269	185.571	-4.420	0.000	***
Animacy $\times$ Extraction	-0.100	0.374	175.953	-0.267	0.790	

**Table 7:** Summary of the mixed-effects model with object animacy and object extraction as predictors.



**Figure 7:** Acceptability as a function of  $\pm$  Extraction and  $\pm$  Animate of the object for each of the four conditions (A–D).

### 3 Discussion

We made five predictions about the factors that affect acceptability (see section 1.8 above). In the following sections, we address each prediction separately.

#### 3.1 Prediction I: Acceptability decreases as complexity increases

Prediction I was borne out. Overall, the levels of acceptability match the levels of complexity, such that the more complex a grammatical structure is, the less acceptable it is. Note that the order of conditions sorted by decreasing acceptability in **Figure 1 (B)** corresponds to the order of increasing complexity (see **Table 1** and (11)) (except the order of C1 and D1, which do not differ significantly in acceptability). First of all, this negative correlation between acceptability and complexity is clear for short versus long movement (types A and B). We predicted that the acceptability would decrease as a function of increased complexity, as measured in terms of embedding (complementation) and length of path (number of overt XPs between filler and gap). The overall effect of embedding a complement clause was significant ( $p = 0.019$ , see **Table 6**), though the difference in acceptability between A2 and B1 was non-significant ( $p = 0.099$ , see **Table 4**). (Recall that by embedding, we refer to the presence of two propositions, one embedded in the other, not simple clefting.) For simple, mono-clausal sentences, the effect of extraction was non-significant (compare A1 and A2), whereas the effect was significant for bi-clausal sentences where the embedded clause is the complement of the matrix verb (B1 vs. B2,  $p = 0.003$ ), similar to the results by Christensen et al. (2013a; 2013b). The major difference between extraction in A2 and B2 is that in B2, the filler is extracted across a clause boundary. That is, there is an effect

of Move-Out. Interestingly, the effects of lexical frequency also only become significant when the extraction crosses a clause boundary.

All things being equal, compared to an embedded complement clause, an adjunct clause is structurally more complex and costly in parsing, and hence less acceptable, cf. (8)a and (9)a, and the acceptability levels for B1 and C1 in **Figure 1**. The parasitic-gap version in C2 (see the structure in (9)b) is even more complex, because it involves two fillers instead of one, and longer movement paths. Consequently, the acceptability drops significantly, cf. C1 vs. C2 in **Figure 1**.

Admittedly, C2 and B2 cannot be compared directly, as B2 does not involve extraction from an embedded (adjunct) clause to the front of the matrix clause. For recent data and discussion on extraction from adjunct clauses in English, see Nyvad et al. (2022). Extraction from relative clauses, on the other hand, does involve Move-Out, (10)b, as well as two fillers. This long extraction also has a long movement path (it crosses many XP nodes in the structure). The result is, as predicted, a severe reduction in acceptability. Importantly, however, it is significantly higher in acceptability than ungrammatical fillers (1.3 points on the 7-point scale), compare D2 and E2 in **Figure 1**. Though the ungrammatical filler type E2 has a much simpler structure on all parameters, it is much less acceptable ( $p < 0.001$ ).

### 3.2 Prediction II: No evidence that construction frequency predicts acceptability

This prediction does not seem to be borne out. As shown in **Figure 4**, there is indeed a positive linear correlation between ( $\log_{10}$ ) frequency of construction and acceptability which describes 69% of the variation in the data ( $R^2 = 0.69$ ). However, the acceptability of a construction could also, and to a higher degree, be predicted from its level of complexity (prediction I), both from the predicted level of complexity (**Figure 4**,  $R^2 = 0.83$ ) and from the fitted level (**Figure 5**,  $R^2 = 0.86$ ).

According to the processing account presented in Culicover et al. (2022), probabilistic expectations concerning which structures we are likely to be exposed to next emerge on the basis of prior experience. They posit that higher complexity (defined as the result of many different factors, including memory constraints, parsing difficulty and dependency length) naturally leads to lower frequency, which in turn results in lower acceptability due to “surprisal” (i.e. degree of surprise triggered by a given structure). Frequency and surprisal should hence be inversely related, and we would predict that the more frequent a construction is, the more acceptable it is, and the more complex it is, the less frequent and the less acceptable it is. These predictions appear to be borne out in **Figure 4** and **Figure 5** (see also Featherston 2008, who found the same pattern; Kempen & Harbusch 2008).

However, it is important to note that zero-occurrence is associated with huge variation in acceptability, which is surprising from the perspective of a usage-based account. In other words,

frequency does not appear to drive the level of acceptability. This is evident from, in particular, C2 (parasitic gaps), D2 (extraction from relative clauses), and E2 (ungrammatical fillers), all of which have (close to) zero occurrence, but significantly different acceptability ratings, 3.35, 2.61, and 1.31, respectively ( $p < 0.001$ ). In fact, it seems that the (very) low frequency of C2, D2, and E2, follows naturally from the relatively high complexity of C2 and D2, and from the ungrammaticality of E2. Not the other way around. This is also evident in **Figure 4 (B)** and **Figure 5 (B)**, which show that there is a significant linear negative correlation between complexity and acceptability which describes 83–86% of the variation (predicted complexity:  $R^2 = 0.83$ , fitted complexity:  $R^2 = 0.86$ ;  $p < 0.002$ ), i.e., 14–17% more than the effect size for construction frequency.

In short, even though the level of acceptability can to some extent be predicted by construction frequency, the level of complexity has more predictive power and is hence able to explain more of the variation – as well as the relative levels of construction frequency. The question is whether what we are observing is merely a negative correlation between two factors (complexity and construction frequency) – indeed, as shown in **Figure 4 (C)** and **Figure 5 (C)**, that negative correlation is significant and strong,  $R^2 = 0.92$  and  $R^2 = 0.77$ , respectively) – or whether there is actually a causal link between the two, and if so, what the direction of causality is. It seems to us that it is more likely that increased complexity causes speakers to disprefer certain constructions over others, which leads to decreased frequency of occurrence, rather than rarity causing certain constructions to be more complex (Hawkins 1994; 2004). As Newmeyer (2005: 125) puts it: “The more processing involved, the rarer the structure”.

### **3.3 Prediction III: Acceptability is somewhat, but not dramatically affected by lexical frequency**

Overall, lexical frequency has a significant but very small effect: An increase of 1 on the 7-point acceptability scale requires an increase of more than 12 million in raw lexical frequency (i.e. tokens in the corpus). This variation also shows that it is very important to have a sufficiently large number of tokens per type in order to avoid accidental effects of selection bias. However, as shown in **Figure 6**, the effect size, polarity, and significance of frequency effects are highly dependent on the complexity of the syntactic construction. Crucially, there are no effects of lexical frequency on ungrammatical sentences. This is consistent with the hypothesis that in parsing, syntactic information takes priority over semantic and pragmatic information, i.e. ‘structure before meaning’ (Kizach et al. 2013). Lexical frequency effects seem to only apply to sentences with acceptability ratings in the intermediate zone; this is parallel to what has been found for satiation (repetition / trial) effects, namely, that it also applies to the intermediate zone (Sprouse 2008; Christensen et al. 2013a; but see also Brown et al. 2021). Ungrammatical sentences do not get better with more frequent words (a floor effect), and sentences with very

high acceptability, do not get better either (a ceiling effect). To a large extent, the relative rating within the intermediate zone depends on the syntactic complexity, cf. the factors in **Table 1** and the plots in **Figure 1** and **Figure 4 (B)**. The relative ranking can then be affected by the lexical frequencies of the words in the sentences, as shown in **Figure 6**. With complement clauses, the effect seems to depend on Move-Out. Comparing conditions A1 through B2 in **Figure 6**, top row, shows the general trend that as complexity increases, so does the effect of lexical frequency.

As also shown in **Figure 6**, there is a negative correlation between acceptability and lexical frequency in the adjunct clause condition C1 (i.e. an inverse or negative frequency effect). The effect is considerable, accounting for 27% of the variance ( $R^2 = 0.27$ ,  $p < 0.01$ ). In the parasitic gap condition (C2), which involves extraction, the effect is not significant ( $p > 0.2$ ). The question is what drives this negative effect. A possible answer might be that the negative effect of adjunction on acceptability (as complexity increases, acceptability decreases) has more ‘weight’ when the effect of lexical frequency has less ‘weight’ (i.e. with higher lexical frequency, adjunction is easier to process and therefore more acceptable). However, then we would expect relative clause extraction (D2) to show the same pattern, which is not the case.<sup>15</sup> Furthermore, it is also unclear why the negative effect of adjunction in C1 should be ameliorated by extraction in C2.

Another possibility could be that with high-frequency lexical material, the ‘unnaturalness’ of the syntactic structure itself becomes more apparent. In a sense, low frequency words might mask the ‘strangeness’ of the construction. Indeed, Christensen and Nyvad (2019) also found a negative correlation between the frequency of the matrix verb and long adjunct-extractions from *wh*-questions in English, which are argued to be ungrammatical. However, then we would also expect to see a negative effect especially with ungrammatical fillers (E2), but also with extraction from relative clauses (D2), which we do not (cf. **Figure 6**). Yet another possibility might be linked to finiteness: In C1 and C2, the embedded clauses is nonfinite, whereas in B1, B2, D1, and D2, it is finite, and the negative effect of lexical frequency is only found in C1 and (to a smaller degree) in C2. For speakers with Broca’s aphasia, finite verbs are more difficult to produce than infinitive verbs (Friedmann 2001; Bastiaanse & Edwards 2004)). Finite adjunct clauses have also been argued to be more resilient to extraction than nonfinite ones (e.g. Ross 1967; Truswell 2011; Michel & Goodall 2013; but see Müller 2019; Nyvad et al. 2022). In our study, the nonfinite clauses also do not have any auxiliary verbs, so the structure is simpler. However, there

---

<sup>15</sup> It also seems to predict an interaction between lexical frequency and complexity. However, that is exactly what **Figure 6** shows not to be the case, as lexical frequency only has a significant effect for some constructions, and the effect does not systematically increase (or decrease) with increased complexity. Indeed, a post-hoc linear mixed-effects analysis did not show a significant interaction between (mean  $\log_{10}$ ) lexical frequency and (predicted) complexity ( $p = 0.416$ ) or a fixed effect of lexical frequency ( $p = 0.65$ ), but as expected, it did show a significant effect of complexity ( $p < 0.001$ ).

is no negative frequency effects for the much simpler conditions A1 and A2. (It is also not clear how this could be explained with factivity, coherence, relevance, etc.; for a recent discussion, see Christensen & Nyvad 2022.) At present, we have no clear answer to why the acceptability of clausal adjunction, which is fairly common, is negatively correlated with lexical frequency in our data, or why the effect is small and non-significant for parasitic gaps, which are very rare.

In a previous study on extraction from relative clauses in Danish, the frequency of the matrix verb was positively correlated with acceptability (Christensen & Nyvad 2014), and this was particularly the case for sentences with extraction, where it described 78% of the variation ( $R^2 = 0.78$ ,  $p < 0.001$ ), see **Figure 8**. In the present study, there was no significant correlation between lexical frequency of the matrix verb and acceptability, neither with nor without extraction ( $R^2 < 0.04$ ,  $p > 0.5$ ). Though we emphasize that the two studies used different sets of verbs and that a direct comparison between the two is not possible, the difference in the size and significance of the frequency effects is striking. It seems plausible that the difference is due to the presence or absence of context. Without a facilitating context, the acceptability is more sensitive to lexical frequency. A facilitating context, on the other hand, helps to parse the complex structure, which involves a long-distance dependency into a relative clause. Easier and faster lexical access in turn reduces working memory load, which increases acceptability.

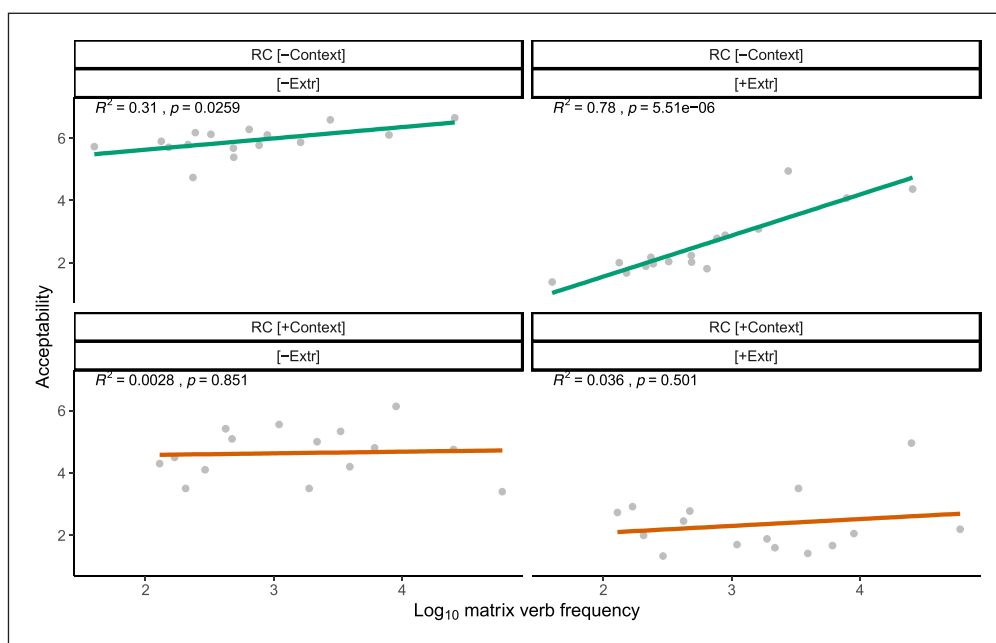


Figure 8: Acceptability as a function of the ( $\log_{10}$ ) lexical frequency of the matrix verb. Data in top row [-Context] from Christensen & Nyvad (2014).<sup>16</sup>

<sup>16</sup> The [+Context] plots exclude sentences from the present study where the verb in question is not the matrix verb, which is why the  $R^2$  and  $p$ -values are not identical to the ones in Figure 6.

Interestingly, in a study on extraction from relative clauses in English, Christensen & Nyvad (2022) found no effect of matrix verb frequency and acceptability of extraction. In this study, the sentences were also presented without a facilitating context, and it was argued that the absence of a frequency effect supports the assumption that English relative clauses are indeed strong syntactic islands. Like ungrammatical fillers, they are immune to any positive effect of high lexical frequency. The results from the present study together with the results from Christensen & Nyvad (2014), as shown in **Figure 8**, are compatible with the assumption that, unlike English, extraction from relative clauses is grammatical in Danish: Though the average rating for extraction is around 2.5 with context and 3 without context, cf. **Figure 8** (right panels), the range in acceptable ranges from very low to 4–5 depending on frequency; furthermore, they are easy to find in spoken as well as written language (Müller & Eggers 2022).

### **3.4 Prediction IV: Acceptability of extraction interacts with the animacy**

The prediction that acceptability of extraction interacts with the animacy of the object was not borne out. Neither the interaction between animacy and extraction nor the fixed effect of animacy were significant (cf. **Table 7**). The effect of extraction, on the other hand, was highly significant ( $p < 0.001$ ). As shown in **Figure 7**, the effects of animacy were also extremely small, while the effects of extraction were large, in particular when the syntactic complexity is already high (types C and D).

## **4 Conclusions**

We have presented the results from a study of a range of constructions with varying frequency, with lexical material with different degrees of frequency, and a substantial number of tokens and participants. In accordance with usage-based approaches, the results of our experiment show that frequency does have an effect on acceptability. There is a positive correlation between construction frequency and acceptability judgments. Likewise, acceptability is positively correlated with lexical frequency of the head noun of the filler, of the main verb, and of the sentential adverbial in some constructions. However, these effects are highly contingent on the syntax. In fact, the role of lexical frequency on acceptability is only significant when the construction in question involves extraction that crosses a clausal boundary. In addition, as expected under a structure-based account, the frequency level of the lexical items in ungrammatical strings does not affect acceptability. Similarly, the putative semantic effect of animacy is overshadowed by the effect of extraction. The primary factor that determines acceptability appears to be processing load, which in turn is reducible to a number of structural factors in our experiment, namely embedding, adjunction, extraction, and path, plus the extra processing cost of extraction out



of an embedded clause. Frequency of occurrence of the construction itself has less explanatory force than the structure itself. 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). This conclusion is fully compatible with a structure-based processing account.

---

## Data Accessibility Statement

The stimulus set used in this study, the dataset, and the R script are openly available at the Open Science Framework: [https://osf.io/rze9t/?view\\_only=60d711bb7ca249e29a8e4be511ba661d](https://osf.io/rze9t/?view_only=60d711bb7ca249e29a8e4be511ba661d)

## Ethics and consent

The study was granted exemption of approval by the Aarhus University Research Ethics Committee. According to §2.1 and §14.2 of the Danish Act on Research Ethics Review of Health Research Projects (the Committee Act. [https://danskelove.dk/komite\\_loven/](https://danskelove.dk/komite_loven/)), surveys that use questionnaires and do not involve human biological material are not subject to notification of and approval from the Research Ethics Committee (see also <http://en.nvk.dk>). The study was completely anonymous and voluntary, it did not involve sensitive personal data or offensive material, and it did not involve any physical or mental discomfort for the participants.

## Funding information

This study was partly funded by a grant from the Danish Council for Independent Research, DFF (grant ID: DFF – 9062-00047B).

## Acknowledgements

We thank the five anonymous reviewers for their constructive criticism and Christiane Müller and Clara Ulrich Eggers for their tremendous help on the construction the stimulus sentences. The authors would also like to thank Elisabet Engdahl, Sten Vikner, and Katrine Rosendal Ehlers, as well as the audiences at *Grammar in focus*, Lund University, Sweden, 2022, and *Grammatiknetværksmødet*, Odense, Denmark, 2023, for helpful comments.

## Competing interests

The authors have no competing interests to declare.

## Author contributions

Conceptualization, methodology, investigation: KRC & AMN. Data curation, formal analysis, and visualization: KRC. Writing draft, reviewing, and editing: KRC & AMN. Project administration and funding acquisition: AMN. Both authors have read and agreed to the published version of the manuscript.

---

## References

- Adli, Aria. 2015. What you like is not what you do: Acceptability and frequency in syntactic variation. *Variation in Language: System- and Usage-based Approaches*, 173–200. De Gruyter. DOI: <https://doi.org/10.1515/9783110346855-008>
- Aissen, Judith. 2003. Differential Object Marking: Iconicity vs. Economy. *Natural Language & Linguistic Theory* 21(3). 435–483. DOI: <https://doi.org/10.1023/A:1024109008573>
- Anderssen, Merete & Westergaard, Marit. 2010. Frequency and economy in the acquisition of variable word order. *Lingua* 120(11). 2569–2588. DOI: <https://doi.org/10.1016/j.lingua.2010.06.006>
- Bader, Markus & Häussler, Jana. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46(2). 273–330. DOI: <https://doi.org/10.1017/S0022226709990260>
- Bastiaanse, Roelien & Bouma, Gosse & Post, Wendy. 2009. Linguistic complexity and frequency in agrammatic speech production. *Brain and Language* 109(1). 18–28. DOI: <https://doi.org/10.1016/j.bandl.2008.12.004>
- Bastiaanse, Roelien & Edwards, Susan. 2004. Word order and finiteness in Dutch and English Broca's and Wernicke's aphasia. *Brain and Language* 89(1). 91–107. DOI: [https://doi.org/10.1016/S0093-934X\(03\)00306-7](https://doi.org/10.1016/S0093-934X(03)00306-7)
- Bell, Alan & Brenier, Jason M. & Gregory, Michelle & Girand, Cynthia & Jurafsky, Dan. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92–111. DOI: <https://doi.org/10.1016/j.jml.2008.06.003>
- Bever, Thomas G. & Sanz, Montserrat & Townsend, David J. 1998. The Emperor's Psycholinguistics. *Journal of Psycholinguistic Research* 27(2). 261–284. DOI: <https://doi.org/10.1023/A:1023206317518>
- Bondevik, Ingrid & Kush, Dave & Lohndal, Terje. 2020. Variation in adjunct islands: The case of Norwegian. *Nordic Journal of Linguistics* 44(3). 1–32. DOI: <https://doi.org/10.1017/S0332586520000207>
- Bornkessel, Ina & Schlesewsky, Matthias & Friederici, Angela D. 2002. Grammar overrides frequency: Evidence from the online processing of flexible word order. *Cognition* 85(2). B21–B30. DOI: [https://doi.org/10.1016/S0010-0277\(02\)00076-8](https://doi.org/10.1016/S0010-0277(02)00076-8)
- Boxell, Oliver & Felser, Claudia. 2017. Sensitivity to parasitic gaps inside subject islands in native and non-native sentence processing. *Bilingualism: Language and Cognition* 20(3). 494–511. DOI: <https://doi.org/10.1017/S1366728915000942>
- Brown, Jessica M. M. & Fanselow, Gisbert & Hall, Rebecca & Kliegl, Reinhold. 2021. Middle ratings rise regardless of grammatical construction: Testing syntactic variability in a repeated exposure paradigm. *PLOS ONE* 16(5). e0251280. DOI: <https://doi.org/10.1371/journal.pone.0251280>
- Bybee, Joan L. & Eddington, David. 2006. A Usage-based Approach to Spanish Verbs of “Becoming.” *Language* 82(2). 323–355. DOI: <https://doi.org/10.1353/lan.2006.0081>
- Cai, Zhenguang G. & Zhao, Nan & Pickering, Martin J. 2022. How do people interpret implausible sentences? *Cognition* 225. 105101. DOI: <https://doi.org/10.1016/j.cognition.2022.105101>

- Carifio, James & Perla, Rocco. 2008. Resolving the 50-year debate around using and misusing Likert scales. *Medical Education* 42(12). 1150–1152. DOI: <https://doi.org/10.1111/j.1365-2923.2008.03172.x>
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In Kenstowicz, Michael J. (ed.), *Ken Hale: A Life in Language*, 1–52. Cambridge, MA: MIT Press. DOI: <https://doi.org/10.7551/mitpress/4056.003.0004>
- Christensen, Ken Ramshøj. 2005. *Interfaces: Negation – Syntax – Brain*. Aarhus: Aarhus University. (PhD dissertation.) <http://tildeweb.au.dk/au572/Papers/krc-phd.pdf>.
- Christensen, Ken Ramshøj. 2016. The Dead Ends of Language: The (Mis)interpretation of a Grammatical Illusion. In Vikner, Sten & Jørgensen, Henrik & Van Gelderen, Elly (eds.), *Let us have articles betwixt us – Papers in Historical and Comparative Linguistics in Honour of Johanna L. Wood*, 129–160. Aarhus: Dept. of English, School of Communication & Culture, Aarhus University. <https://ebooks.au.dk/aul/catalog/view/119/107/466-1>.
- Christensen, Ken Ramshøj & Kizach, Johannes & Nyvad, Anne Mette. 2013a. Escape from the Island: Grammaticality and (Reduced) Acceptability of *wh*-island Violations in Danish. *Journal of Psycholinguistic Research* 42(1). 51–70. DOI: <https://doi.org/10.1007/s10936-012-9210-x>
- Christensen, Ken Ramshøj & Kizach, Johannes & Nyvad, Anne Mette. 2013b. The processing of syntactic islands – An fMRI study. *Journal of Neurolinguistics* 26(2). 239–251. DOI: <https://doi.org/10.1016/j.jneuroling.2012.08.002>
- Christensen, Ken Ramshøj & Nyvad, Anne Mette. 2014. On the nature of escapable relative islands. *Nordic Journal of Linguistics* 37(1). 29–45. DOI: <https://doi.org/10.1017/S0332586514000055>
- Christensen, Ken Ramshøj & Nyvad, Anne Mette. 2019. No escape from the island: On extraction from complement *wh*-clauses in English. In Christensen, Ken Ramshøj & Jørgensen, Henrik & Wood, Johanna L. (eds.), *The Sign of the V – Papers in Honour of Sten Vikner*, 95–112. Aarhus: Dept. of English, School of Communication & Culture, Aarhus University. DOI: <https://doi.org/10.7146/aul.348.91>
- Christensen, Ken Ramshøj & Nyvad, Anne Mette. 2022. The island is still there: Experimental evidence for the inescapability of relative clauses in English. *Studia Linguistica* 76(3). 1–25. DOI: <https://doi.org/10.1111/stul.12192>
- Christensen, Ken Ramshøj & Wallentin, Mikkel. 2011. The locative alternation: Distinguishing linguistic processing cost from error signals in Broca's region. *NeuroImage* 56(3). 1622–1631. DOI: <https://doi.org/10.1016/j.neuroimage.2011.02.081>
- Christensen, Rune Haubo Bojesen. 2015. ordinal—Regression Models for Ordinal Data. <http://www.cran.r-project.org/package=ordinal/>.
- Clifton, Charles & Fanselow, Gisbert & Frazier, Lyn. 2006. Amnestying superiority violations: Processing multiple questions. *Linguistic Inquiry* 37(1). 51–68. DOI: <https://doi.org/10.1162/002438906775321139>
- Clifton, Charles & Frazier, Lyn. 1989. Comprehending Sentences with Long-Distance Dependencies. In Carlson, Greg N. & Tanenhaus, Michael K. (eds.), *Linguistic Structure in Language Processing*,

- vol. 7, 273–317. Dordrecht: Springer Netherlands. DOI: [https://doi.org/10.1007/978-94-009-2729-2\\_8](https://doi.org/10.1007/978-94-009-2729-2_8)
- Clifton, Charles & Traxler, Matthew J. & Taha Mohamed, Mohamed & Williams, Rihana S. & Morris, Robin K. & Rayner, Keith. 2003. The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language* 49(3). 317–334. DOI: [https://doi.org/10.1016/S0749-596X\(03\)00070-6](https://doi.org/10.1016/S0749-596X(03)00070-6)
- Clos, Mareike & Amunts, Katrin & Laird, Angela R. & Fox, Peter T. & Eickhoff, Simon B. 2013. Tackling the multifunctional nature of Broca's region meta-analytically: Co-activation-based parcellation of area 44. *NeuroImage* 83. 174–188. DOI: <https://doi.org/10.1016/j.neuroimage.2013.06.041>
- Collins, Chris. 1994. Economy of Derivation and the Generalized Proper Binding Condition. *Linguistic Inquiry* 25(1). 45–61. <https://www.jstor.org/stable/4178848>.
- Culicover, Peter W. & Varaschin, Giuseppe & Winkler, Susanne. 2022. The Radical Unacceptability Hypothesis: Accounting for Unacceptability without Universal Constraints. *Languages* 7(2). 96. DOI: <https://doi.org/10.3390/languages7020096>
- Damasio, Antonio R. 1992. Aphasia. *The New England Journal of Medicine* 236(8). 531–539. DOI: <https://doi.org/10.1056/NEJM199202203260806>
- Dapretto, Mirella & Bookheimer, Susan Y. 1999. Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron* 24(2). 427–432. DOI: [https://doi.org/10.1016/S0896-6273\(00\)80855-7](https://doi.org/10.1016/S0896-6273(00)80855-7)
- Ellis, Nick C. 2002. Frequency effects in language processing. *Studies in second language acquisition* 24(02). 143–188. DOI: <https://doi.org/10.1017/S0272263102002024>
- Engdahl, Elisabet. 1983. Parasitic gaps. *Linguistics and Philosophy* 6(1). 5–34. DOI: <https://doi.org/10.1007/BF00868088>
- Fanselow, Gisbert & Frisch, Stefan. 2006. Effects of processing difficulty on judgments of acceptability. In Fanselow, Gisbert & Féry, Caroline & Schlesewsky, Matthias & Vogel, Ralf (eds.), *Gradience in Grammar: Generative Perspectives*, 291–316. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199274796.003.0015>
- Fanselow, Gisbert & Schlesewsky, Matthias & Vogel, Ralf & Weskott, Thomas. 2011. Animacy effects on crossing wh-movement in German. *Linguistics* 49(4). DOI: <https://doi.org/10.1515/ling.2011.021>
- Featherston, Sam. 2008. The Decathlon Model of Empirical Syntax. In Kepser, Stephan & Reis, Marga (eds.), *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*, 187–208. Berlin: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110197549.187>
- Fodor, J. A. & Garrett, M. 1967. Some syntactic determinants of sentential complexity. *Perception & Psychophysics* 2(7). 289–296. DOI: <https://doi.org/10.3758/BF03211044>
- Friederici, Angela D. 2002. Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences* 6(2). 78–84. DOI: [https://doi.org/10.1016/S1364-6613\(00\)01839-8](https://doi.org/10.1016/S1364-6613(00)01839-8)
- Friedmann, Na'ama. 2001. Agrammatism and the psychological reality of the syntactic tree. *Journal of Psycholinguistic Research* 30(1). 71–90. DOI: <https://doi.org/10.1023/A:1005256224207>

- Friedmann, Na'ama. 2003. The fragile nature of the left periphery: CP deficits in agrammatic aphasia. *Proceedings of the 18th IATL conference*. <http://linguistics.huji.ac.il/IATL/18/Friedmann.pdf> (Accessed December 13, 2013.)
- Friedmann, Na'ama & Grodzinsky, Yosef. 1997. Tense and Agreement in Agrammatic Production: Pruning the Syntactic Tree. *Brain and Language* 56(3). 397–425. DOI: <https://doi.org/10.1006/brln.1997.1795>
- Gennari, Silvia P. & MacDonald, Maryellen C. 2009. Linking production and comprehension processes: The case of relative clauses. *Cognition* 111(1). 1–23. DOI: <https://doi.org/10.1016/j.cognition.2008.12.006>
- Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1). 1–76. DOI: [https://doi.org/10.1016/S0010-0277\(98\)00034-1](https://doi.org/10.1016/S0010-0277(98)00034-1)
- Grewe, Tanja & Bornkessel, Ina & Zysset, Stefan & Wiese, Richard & Yves von Cramon, D. & Schleesky, Matthias. 2006. Linguistic prominence and Broca's area: The influence of animacy as a linearization principle. *NeuroImage* 32(3). 1395–1402. DOI: <https://doi.org/10.1016/j.neuroimage.2006.04.213>
- Gries, Stefan Th. & Ellis, Nick C. 2015. Statistical Measures for Usage-Based Linguistics. *Language Learning* 65(S1). 228–255. DOI: <https://doi.org/10.1111/lang.12119>
- Grodzinsky, Yosef & Friederici, Angela D. 2006. Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology* 16(2). 240–246. DOI: <https://doi.org/10.1016/j.conb.2006.03.007>
- Grodzinsky, Yosef & Pieperhoff, Peter & Thompson, Cynthia. 2021. Stable brain loci for the processing of complex syntax: A review of the current neuroimaging evidence. *Cortex* 142. 252–271. DOI: <https://doi.org/10.1016/j.cortex.2021.06.003>
- Haegeman, Liliane. 1994. *Introduction to government and binding theory*. 2nd edn. Oxford: Blackwell.
- Harley, Trevor A. 2014. *The psychology of language: from data to theory*. 4th edn. Hove, East Sussex: Psychology Press, Taylor & Francis Group.
- Häussler, Jana & Grant, Margaret & Fanselow, Gisbert & Frazier, Lyn. 2015. Superiority in English and German: Cross-Language Grammatical Differences? *Syntax* 18(3). 235–265. DOI: <https://doi.org/10.1111/synt.12030>
- Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511554285>
- Hawkins, John A. 1999. Processing complexity and filler-gap dependencies across grammars. *Language*. Linguistic Society of America 75(2). 244–285. DOI: <https://doi.org/10.2307/417261>
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199252695.001.0001>
- Hofmeister, Philip & Sag, Ivan A. 2010. Cognitive constraints and island effects. *Language* 86(2). 366–415. DOI: <https://doi.org/10.1353/lan.0.0223>



- Hornstein, Norbert. 2014. DTC redux. *Faculty of Language*. <http://facultyoflanguage.blogspot.com/2014/02/dtc-redux.html> (Accessed January 16, 2022.)
- Huang, Cheng-Teh James. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Cambridge, MA: MIT. (PhD dissertation.) [http://pubman.mpdl.mpg.de/pubman/item/escidoc:403106:7/component/escidoc:403105/chinese\\_huang1982\\_o.pdf](http://pubman.mpdl.mpg.de/pubman/item/escidoc:403106:7/component/escidoc:403105/chinese_huang1982_o.pdf) (Accessed November 13, 2015.)
- Kassambara, Alboukadel. 2020. *ggpubr: “ggplot2” Based Publication Ready Plots*. <https://CRAN.R-project.org/package=ggpubr>.
- Kempen, Gerard & Harbusch, Karin. 2008. The Relationship between Grammaticality Ratings and Corpus Frequencies: A Case Study into Word Order Variability in the Midfield of German Clauses. In Kepser, Stephan & Reis, Marga (eds.), *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*, 329–350. Berlin: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110197549.329>
- Kizach, Johannes & Nyvad, Anne Mette & Christensen, Ken Ramshøj. 2013. Structure before Meaning: Sentence Processing, Plausibility, and Subcategorization. *PLoS ONE* 8(10). e76326. DOI: <https://doi.org/10.1371/journal.pone.0076326>
- KorpusDK. <https://ordnet.dk/korpusdk>.
- Kuznetsova, Alexandra & Brockhoff, Per B. & Christensen, Rune H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13). DOI: <https://doi.org/10.18637/jss.v082.i13>
- Lee, Jiyeon & Thompson, Cynthia K. 2011. Real-time production of arguments and adjuncts in normal and agrammatic speakers. *Language and Cognitive Processes* 26(8). 985–1021. DOI: <https://doi.org/10.1080/01690965.2010.496237>
- Marantz, Alec. 2005. Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review* 22(2–4). 429–445. DOI: <https://doi.org/10.1515/tlir.2005.22.2-4.429>
- Mayo, Pilar García. 1997. Non-Occurrence of Subject and Adjunct Parasitic Gaps. *Atlantis. AEDEAN: Asociación española de estudios anglo-americanos* 19(2). 125–133. <https://www.jstor.org/stable/41055466>.
- Meltzer-Asscher, Aya & Thompson, Cynthia K. 2014. The forgotten grammatical category: Adjective use in agrammatic aphasia. *Journal of Neurolinguistics* 30. 48–68. DOI: <https://doi.org/10.1016/j.jneuroling.2014.04.001>
- Michel, Dan & Goodall, Grant. 2013. Finiteness and the nature of island constraints. In Goto, Nobu & Otaki, Koichi & Sato, Atsushi & Takita, Kensuke (eds.), *Proceedings of GLOW in Asia IX 2012: The Main Session*, 187–197. [https://faculty.human.mie-u.ac.jp/~glow\\_mie/IX\\_Proceedings\\_Oral/12Michel&Goodall.pdf](https://faculty.human.mie-u.ac.jp/~glow_mie/IX_Proceedings_Oral/12Michel&Goodall.pdf).
- Miller, George A. & Chomsky, Noam. 1963. Finitary models of language users. In Luce, Robert Duncan (ed.), *Handbook of Mathematical Psychology*, vol. 2, 419–491. New York: John Wiley & Sons.
- Momma, Shota & Yoshida, Masaya. 2023. Planning multiple dependencies in sentence production. *Language, Cognition and Neuroscience*. Routledge 38(9). 1183–1213. DOI: <https://doi.org/10.1080/023273798.2023.2212820>

- Müller, Christiane. 2019. *Permeable islands. A contrastive study of Swedish and English adjunct clause extractions*. Lund: Lund University. (PhD dissertation.) [https://lucris.lub.lu.se/ws/portalfiles/portal/63305372/Permeable\\_islands\\_Christiane\\_M\\_ller.pdf](https://lucris.lub.lu.se/ws/portalfiles/portal/63305372/Permeable_islands_Christiane_M_ller.pdf).
- Müller, Christiane & Eggers, Clara Ulrich. 2022. Island Extractions in the Wild: A Corpus Study of Adjunct and Relative Clause Islands in Danish and English. *Languages* 7(2). 125. DOI: <https://doi.org/10.3390/languages7020125>
- Newman, Sharlene D. & Just, Marcel Adam & Keller, Timothy A. & Roth, Jennifer & Carpenter, Patricia A. 2003. Differential effects of syntactic and semantic processing on the subregions of Broca's area. *Cognitive Brain Research* 16(2). 297–307. DOI: [https://doi.org/10.1016/S0926-6410\(02\)00285-9](https://doi.org/10.1016/S0926-6410(02)00285-9)
- Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language* 79(4). 682–707. <http://www.jstor.org/stable/4489522>. DOI: <https://doi.org/10.1353/lan.2003.0260>
- Newmeyer, Frederick J. 2005. *Possible and probable languages: a generative perspective on linguistic typology*. New York: Oxford University Press.
- Norman, Geoff. 2010. Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education* 15(5). 625–632. DOI: <https://doi.org/10.1007/s10459-010-9222-y>
- Nyvad, Anne Mette & Kizach, Johannes & Christensen, Ken Ramshøj. 2014. (Non-)Arguments in Long-Distance Extractions. *Journal of Psycholinguistic Research* 44. 1–13. DOI: <https://doi.org/10.1007/s10936-014-9300-z>
- Nyvad, Anne Mette & Müller, Christiane & Christensen, Ken Ramshøj. 2022. Too True to Be Good? The Non-Uniformity of Extraction from Adjunct Clauses in English. *Languages* 7(4). 244. DOI: <https://doi.org/10.3390/languages7040244>
- O'Grady, William & Lee, Miseon & Choo, Miho. 2003. A subject-object asymmetry in the acquisition of relative clauses in Korean as a second language. *Studies in Second Language Acquisition*. Cambridge University Press 25(3). 433–448. DOI: <https://doi.org/10.1017/S0272263103000172>
- Pallier, Christophe & Devauchelle, Anne-Dominique & Dehaene, Stanislas. 2011. Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences* 108(6). 2522–2527. DOI: <https://doi.org/10.1073/pnas.1018711108>
- Phillips, Colin. 2013. On the nature of island constraints I: Language processing and reductionist accounts. In Sprouse, Jon & Hornstein, Norbert (eds.), *Experimental Syntax and Island Effects*, 64–108. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139035309.005>
- Piantadosi, Steven T. & Tily, Harry & Gibson, Edward. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108(9). 3526–3529. DOI: <https://doi.org/10.1073/pnas.1012551108>
- Pickering, Martin J. & Traxler, Matthew J. & Crocker, Matthew W. 2000. Ambiguity Resolution in Sentence Processing: Evidence against Frequency-Based Accounts. *Journal of Memory and Language* 43(3). 447–475. DOI: <https://doi.org/10.1006/jmla.2000.2708>



- Pinker, Steven. 2004. Clarifying the logical problem of language acquisition. *Journal of Child Language* 31(4). 949–953. DOI: <https://doi.org/10.1017/S0305000904006439>
- Potts, Christopher. 2004. Parasitic gaps in the wild. *Language Log*. <http://itre.cis.upenn.edu/~myl/languageelog/archives/000561.html> (Accessed November 21, 2023.)
- Prior, Anat & Bentin, Shlomo. 2006. Differential integration efforts of mandatory and optional sentence constituents. *Psychophysiology* 43(5). 440–449. DOI: <https://doi.org/10.1111/j.1469-8986.2006.00426.x>
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Realí, Florencia & Christiansen, Morten H. 2007. Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* 57(1). 1–23. DOI: <https://doi.org/10.1016/j.jml.2006.08.014>
- Resnik, Philip. 2004. Wild parasitic gap construction escapes. *Language Log*: <http://itre.cis.upenn.edu/~myl/languageelog/archives/000571.html> (Accessed November 21, 2023.)
- Ross, John Robert. 1967. *Constraints on variables in syntax*. Cambridge, MA: Massachusetts Institute of Technology. (PhD dissertation.) <https://dspace.mit.edu/handle/1721.1/15166> (Accessed April 16, 2018)
- Schell, Marianne & Zaccarella, Emiliano & Friederici, Angela D. 2017. Differential cortical contribution of syntax and semantics: An fMRI study on two-word phrasal processing. *Cortex* 96. 105–120. DOI: <https://doi.org/10.1016/j.cortex.2017.09.002>
- Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31(3). 575–582. DOI: <https://doi.org/10.1162/002438900554479>
- Snyder, William. 2022. On the Nature of Syntactic Satiation. *Languages* 7(1). 38. DOI: <https://doi.org/10.3390/languages7010038>
- Sonderegger, Morgan & Wagner, Michael & Torreira, Francisco. 2018. *Quantitative Methods for Linguistic Data*. 1.0. <https://people.linguistics.mcgill.ca/~morgan/qmld-book/lmem.html> (Accessed May 23, 2023.)
- Sprouse, Jon. 2008. The differential sensitivity of acceptability judgments to processing effects. *Linguistic Inquiry* 39(4). 686–694. DOI: <https://doi.org/10.1162/ling.2008.39.4.686>
- Sprouse, Jon & Almeida, Diogo. 2017. Setting the empirical record straight: Acceptability judgments appear to be reliable, robust, and replicable. *Behavioral and Brain Sciences* 40. e311. DOI: <https://doi.org/10.1017/S0140525X17000590>
- Sprouse, Jon & Wagers, Matt & Phillips, Colin. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88(1). 82–123. DOI: <https://doi.org/10.1353/lan.2012.0004>
- Stromswold, Karin & Caplan, David & Alpert, Nathaniel & Rauch, Scott. 1996. Localization of syntactic comprehension by positron emission tomography. *Brain and language* 52(3). 452–473. DOI: <https://doi.org/10.1006/brln.1996.0024>

- Sullivan, Gail M. & Artino, Anthony R. 2013. Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education* 5(4). 541–542. DOI: <https://doi.org/10.4300/JGME-5-4-18>
- Taylor, Jack E. & Rousselet, Guillaume A. & Scheepers, Christoph & Sereno, Sara C. 2022. Rating norms should be calculated from cumulative link mixed effects models. *Behavior Research Methods*. DOI: <https://doi.org/10.31234/osf.io/3vgwk>
- Thornton, Rosalind & Crain, Stephen. 1994. Successful Cyclic Movement. In Hoekstra, Teun & Schwartz, Bonnie D. (eds.), *Language Acquisition Studies in Generative Grammar*, 215–252. Amsterdam: John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/lald.8.11tho>
- Tomasello, Michael. 2003. *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Trueswell, John C. 1996. The Role of Lexical Frequency in Syntactic Ambiguity Resolution. *Journal of Memory and Language* 35(4). 566–585. DOI: <https://doi.org/10.1006/jmla.1996.0030>
- Truswell, Robert. 2011. *Events, Phrases, and Questions*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199577774.001.0001>
- Venables, William N. & Ripley, Brian D. 2002. *Modern Applied Statistics with S* (Statistics and Computing). New York, NY: Springer New York. DOI: <https://doi.org/10.1007/978-0-387-21706-2>
- Wellwood, Alexis & Pancheva, Roumyana & Hacquard, Valentine & Phillips, Colin. 2018. The Anatomy of a Comparative Illusion. *Journal of Semantics* 35(3). 543–583. DOI: <https://doi.org/10.1093/jos/ffy014>
- White, Aaron Steven & Rawlins, Kyle. 2020. Frequency, acceptability, and selection: A case study of clause-embedding. *Glossa* 5(1). 105. DOI: <https://doi.org/10.5334/gjgl.1001>
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer. (<https://ggplot2.tidyverse.org>). DOI: <https://doi.org/10.1007/978-3-319-24277-4>
- Wu, Lang. 2009. *Mixed Effects Models for Complex Data*. New York: Chapman and Hall/CRC. DOI: <https://doi.org/10.1201/9781420074086>
- Yang, Charles D. 2004. Universal Grammar, statistics or both? *Trends in Cognitive Sciences* 8(10). 451–456. DOI: <https://doi.org/10.1016/j.tics.2004.08.006>
- Yang, Charles D. 2015. For and against frequencies. *Journal of Child Language* 42(02). 287–293. DOI: <https://doi.org/10.1017/S0305000914000683>
- Zaccarella, Emiliano & Meyer, Lars & Makuuchi, Michiru & Friederici, Angela D. 2015. Building by syntax: The neural basis of minimal linguistic structures. *Cerebral Cortex* bhv234. DOI: <https://doi.org/10.1093/cercor/bhv234>
- Zipf, George Kingsley. 1935. *The psycho-biology of language: An introduction to dynamic philology*. Boston: Houghton Mifflin Company.

