Language (2026), **102**:2, 329–371
doi:10.1017/S009785072600010X

Linguistic Society of America

GENERAL RESEARCH ARTICLE

Linear asymmetries and incremental parsing

Klaus Abels and Ad Neeleman

Department of Linguistics, University College London, UK

Corresponding author: Ad Neeleman; Email: a.neeleman@ucl.ac.uk

Received: 02 December 2023; **Revised:** 26 February 2026; **Accepted:** 28 March 2026

Keywords: left-right asymmetries; universal 20; heavy-XP shift; parsing; typological frequency

Abstract

Linear asymmetries pose a challenge for syntactic theory but may yield to an explanation in terms of sentence parsing, given the inherent left-to-right nature of this process. We therefore explore the option of a parsing-based account of Greenberg's (1963) universal 20. Universal 20 describes a word-order asymmetry in noun phrases that can be captured if movement of the noun is exclusively leftward (Cinque 2005 and Abels & Neeleman 2012). Our account is based on the hypothesis that rightward movement of obligatory elements leads to a specific processing difficulty that we call the missing-filler effect. Independent evidence for this effect comes from self-paced reading and maze paradigm experiments that replicate and extend earlier findings by Staub et al. (2006) regarding the parsing of heavy-XP shift in the context of optionally and obligatorily transitive verbs. We conclude that the absence of rightward movement in the noun phrase need not be attributed to a syntactic constraint but can be traced back to the missing-filler effect if we assume that parsing effort affects language acquisition, which in turn affects typological frequency (as argued in Hawkins 1990, 2009 and Kirby 1999).

1. Introduction

Sentence grammar shows a number of pervasive left-right asymmetries. These include asymmetries in interpretive dependencies (for example, binders tend to precede bound categories; Barker 2012, Bruening 2014), word order (for example, verb-object adjacency is absent in SOV languages, but characteristic of a subset of SVO languages; Haider 2005, 2014), and movement (for example, long-distance movement is predominantly leftward; Ross 1967, Bach 1971, Perlmutter 1983, Kayne 1994). Rules that refer to structural notions like c-command and dominance cannot capture such left-right asymmetries unless they are accompanied by a constraint like Kayne's (1994) linear correspondence axiom, which guarantees a rigid mapping between syntactic structure and linear order. However, a number of theoretical and empirical problems have been identified that militate against such a rigid mapping (Abels & Neeleman 2012, Abels 2016, Neeleman 2017, Neeleman & Payne 2020). The question of how to account for left-right asymmetries therefore remains open, despite its vital importance for syntactic theory.

We aim to contribute to the solution of this problem by examining one particularly striking left-right asymmetry: the ban on rightward movement of the noun in the derivation of neutral noun phrase orders. This ban is integral to the explanation of a set of observations known as **Universal 20** (Greenberg 1963, Hawkins 1983, Rijkhoff 1990, 2002, Cinque 2005, 2025, Dryer 2018), yet remains—in syntactic terms—a stipulation.

Greenberg's original formulation of universal 20 and some of the data patterns supporting it are given in 1 and 2 below. In 2 we have added the number of attestations (languages, genera) according to Dryer

© The Author(s), 2026. Published by Cambridge University Press on behalf of the Linguistic Society of America. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

2018 and Cinque 2025, respectively.[1] We should note that universal 20 is a generalization about neutral orders in the extended nominal projection (as opposed to marked orders).[2] While the definition of neutral order is a matter of debate, the three-step procedure in 3 summarizes common practice in the field (see Dryer 2007 and Abels 2016 for discussion).

(1) Universal 20: When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite. (Greenberg 1963:87)

(2) a. Dem Num A N (attested (113, 57) (457, 116))
b. N A Num Dem (attested (182, 85) (652, 141))
c. N Dem Num A (attested (8, 6) (93, 27))
d. A Num Dem N (unattested (0,0) (0, 0))

(3) A neutral order is either (i) the only grammatical order, (ii) the only grammatical order that can be used in an out-of-the-blue context, or (iii) the most frequent grammatical order that can be used in an out-of-the-blue context.

It will be clear that universal 20 describes a left-right asymmetry: the order of modifying elements is constant across languages when they appear to the left of the noun, but variable when they appear to its right. Abels & Neeleman 2012 shows that this pattern in the data can be explained under the simultaneous assumption of symmetric base generation and asymmetric, leftward-only, movement (see also Ackema & Neeleman 2002). The analysis is a reinterpretation of Cinque's (1996, 2005) antisymmetric account.

By symmetric base generation, we mean that there is no universal linearization of a projecting category and its sister. The hierarchy in the noun phrase is presumably constant crosslinguistically, with the demonstratives attached higher than numerals and numerals attached higher than adjectives.

(4) Dem > Num > A > N

However, demonstratives, numerals, and adjectives can each be linearized to the left or to the right of the noun depending on the grammar of the particular language, yielding 5a and 5b as the two cases in which all modifiers appear on the same side of the noun.

(5) Symmetric base generation
a. [Dem [Num [A N]]] (attested)
b. [[[N A] Num] Dem] (attested)

By asymmetric movement, we mean that further neutral word orders can be generated through leftward but never rightward movement of the noun.[3] Thus, 6a can be derived from 5a, but the rightward movement that connects 5b to 6b is not available.

(6) Asymmetric movement
a. [N [Dem [Num [A ~~N~~]]]] (attested)
b. [[[~~N~~ A] Num] Dem] N] (unattested)

Of course, the number of possible word orders of four elements (demonstrative, numeral, adjective, and noun) is not four but twenty-four. We therefore need to show how the proposed account scales up.

[1] Cinque's sample is accessible at https://iris.unive.it/handle/10278/3750226.

[2] For discussion of the notion of extended projection, see Grimshaw 1991, 2005.

[3] It is generally accepted that movements that derive neutral word order must be restricted to the noun or constituents containing the noun (so that movement of an adjective or a numeral cannot give rise to a neutral order). It is an open question why this should be so. For discussion, see Cinque 2005, Georgi & Müller 2010, and Steddy & Samek-Lodovici 2011.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

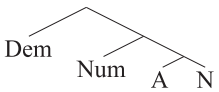
	I	II	III	IV
a.	Dem Num A N	N A Num Dem	N Dem Num A	A Num Dem N
b.	Dem Num N A	A N Num Dem	Dem N Num A	A Num N Dem
c.	Dem A N Num	Num N A Dem	A N Dem Num	Num Dem N A
d.	Dem N A Num	Num A N Dem	N Num A Dem	Dem A Num N
e.	A Dem Num N	N Num Dem A	N Dem A Num	Num A Dem N
f.	A Dem N Num	Num N Dem A	N A Dem Num	Num Dem A N

Table 1. *Attested and unattested orders in the noun phrase according to Cinque 2005, 2025 (unattested orders are grayed out).*

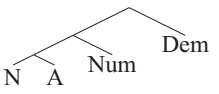
The complete paradigm of all attested and unattested neutral orders of demonstrative, numeral, adjective, and noun was described in Cinque 2005 and supported by additional typological evidence in Cinque 2025 (see Table 1, where unattested orders are shaded and all unshaded orders are attested). The data are characterized by the same basic asymmetry originally observed by Greenberg: the order of elements before the noun is constant, but there is variation following the noun.

Symmetric base generation allows for the following representations, which all yield orders that are attested—they correspond with the unshaded orders in columns I and II in Table 1 (N.B.: 7.Ia and 7.IIa are the arboreal counterparts of the bracketings in 5a,b).

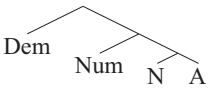
(7) Ia.



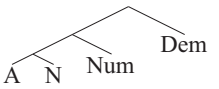
IIa.



Ib.



IIb.



Ic.



IIc.



Id.

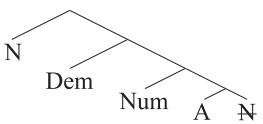


IId.

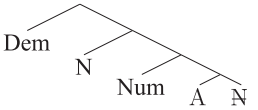


The remaining attested orders can be derived by leftward movement of the noun, as shown in 8 (N.B.: 8. IIIa is the arboreal counterpart of 6a). We should point out that such movement must be allowed to pied-pipe the adjective, given that the orders in IIIc and IIIf are attested.

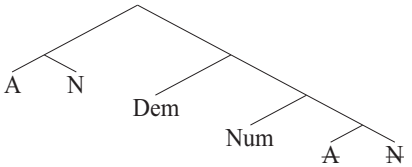
(8) IIIa.



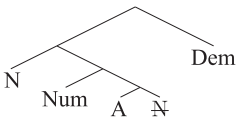
IIIb.



IIIc.

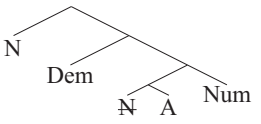


IIId.

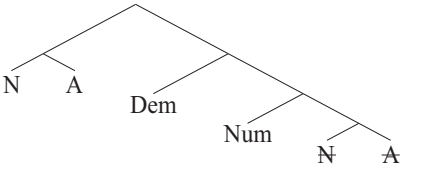


https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

IIIe.



IIIf.



The analysis just sketched renders the unattested orders in Table 1 underivable (as neutral orders). The constraint that movements deriving neutral word order must be leftward is crucial, since it implies that none of the trees in 8, and therefore none of the orders in column III of Table 1, can have a mirror image. This is in line with observations. In Table 1, the relevant mirrored orders, which are given in column IV, are all shaded.

A more precise way of explaining the effect of the ban on rightward movement is as follows. Given the hierarchy in 4, the noun and the adjective must be adjacent in any base-generated structure. Consequently, if they are separated, movement must have taken place. Since neutral orders cannot be derived by movement of adjectives, and since the noun cannot move rightward, it is impossible to separate noun and adjective if they come in A-N order, as in 9a. Similarly, given that numerals must be adjacent in any base-generated structure to the substring comprising the adjective and the noun, the order in 9b is excluded. As a consequence, all of the orders in IVa–f are ruled out, as are the orders in Ie,f.

(9) a. *A … X … N
b. *Num … X … A+N

Two unattested orders remain, namely IIe,f. In these orders, the adjective and the noun are separated, suggesting that N has moved (leftward, as required). But if so, the underlying orders for IIe and IIf must have been either Num-Dem-A-N or Num-Dem-N-A. Neither of these underlying orders is grammatical, as we have already shown in our discussion of IVc,f.

In sum, a ban on rightward movement of the noun (or constituents containing the noun) is instrumental in explaining the full pattern of attested and unattested neutral orders in the extended nominal projection.[4] The crucial question that presents itself at this point is **why** such movement must be leftward, especially in view of the fact that base generation is symmetric. The answer, we suggest, lies not in syntactic theory, but rather in the way movement dependencies are processed (as also argued in Ackema & Neeleman 2002).

We assume that a hearer/reader constructs a syntactic representation of the input string that adheres to universal principles of grammar, as well as to language-specific constraints. Although this representation is syntactic in nature, its construction is aided by parsing strategies that are independent of the grammar and psychologically motivated. This perspective, first advocated in Kimball 1973, Frazier 1978, and Frazier & Fodor 1978, allows us to argue that there is no grammatical ban on rightward movement of the noun. Instead, grammars that employ rightward movement of the noun are subject to typological attrition because such movement systematically causes processing difficulties that can be traced back to an independently motivated parsing strategy (namely de Vincenzi's (1991) minimal chain principle).

Our argument is presented as follows. In Section 2, we discuss how phrasal movement is parsed, arguing that rightward movement of obligatory material causes particular parsing difficulties that we refer to as the **missing-filler effect** (the counterpart of the **filled-gap effect** found with leftward phrasal movement). Section 3 reports on two experiments that confirm that the missing-filler effect is real: it can be detected at the onset of each constituent that separates the launching and landing sites of obligatory material that has moved rightward. In Section 4, we revisit universal 20 and demonstrate that rightward

[4] Steedman (2020) proposes an alternative analysis that rules out only two of the twenty-four logically possible noun phrase orders and allows the equivalent of rightward movement of the noun. However, a comparison with other work in this domain is far from trivial, as his analysis is not restricted to neutral orders. Note that our account could in principle permit rightward movement of the noun if it derives a marked word order. Such movement might in fact be required for the Shupamem data from Nchare 2012 that Steedman bases himself on.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

movement of the noun systematically leads to missing-filler effects. This outcome is expected, given that the lexical head of an extended projection is an obligatory element. We also show that **leftward** movement of the noun, by contrast, is not expected to give rise to a filled-gap effect. This may be surprising at first sight, but it follows from the parsing principles we discuss, and evidence for the absence of filled-gap effects in related structures can be found in processing studies that examine the effects of verb fronting in German main clauses. Finally, we discuss why missing-filler effects, if given enough time, lead to the typological attrition of grammars that make use of rightward movement of the noun, thereby explaining the data in Table 1. Section 5 concludes.

2. Parsing movement

The first step in our argument is to discuss how movement dependencies are parsed and to identify the specific difficulties that the parser faces when dealing with rightward movement of obligatory material.

We assume that the output of the parser is a tree and that the theory of parsing must hence describe how that tree is constructed. For expository purposes, we treat parsing as a one-track structure-building process, that is, as a procedure that develops a single partial tree structure with the aim of covering the symbols in the input string. Nothing hinges on this. Our proposal can equally well be implemented in parsers that generate multiple ranked representations of the input string.

The speed of human parsing cannot be understood unless the process of tree construction is incremental (Frazier & Rayner 1982, Frazier 1987, Just & Carpenter 1992, Gorrell 1995, among others). Incrementality is generally understood to cover two things: immediacy of analysis and faithfulness to commitments. In other words, the parser assigns each input symbol a structural position as soon as it comes in, and once the parser commits to an analysis any divergence from it in further development of the parse tree comes at a cost (see also Marcus et al. 1983). This combination of properties militates against overcommitment: a parser that commits to more structure than necessary will frequently have to retract part of that structure.

An important principle regulating commitment is Frazier and Rayner's (1982:180) strategy of **minimal attachment**, according to which a hearer/reader is to '[a]ttach incoming material into the phrase-marker being constructed using the fewest nodes consistent with the well-formedness rules of the language' under analysis. We give an alternative formulation in 10, which preserves the effects of the original (assuming that the amount of structure is measured in nodes), but makes explicit that additional structure can be posited when required.

(10) Minimal attachment: Do not posit more structure than required by the input.

Given that it would facilitate speedy parsing if firm predictions were structurally reified, we assume that the postulation of unambiguously predicted structure is part of what is required by the input. As an example, consider the sentence *I was reading your letter*. Minimal attachment implies that when the parser encounters *reading*, no object position is postulated, as the verb *read* is optionally transitive and hence there is no unambiguously predicted internal argument. However, when it reaches *your*, a position must be built that can host the nominal category that obligatorily follows this possessive pronoun.

With these basics in place, consider how movement dependencies are parsed. Much current syntactic research in the generative tradition characterizes traces as unpronounced copies (or partial copies) of moved categories (Chomsky 1993, 1995). This characterization has important processing consequences. If traces depend for their content on the moved category, it follows as a matter of logic that the parser cannot insert a trace before having hypothesized that a given category has undergone movement. After all, for there to be a copy, there must be an original. The best the parser can do in the absence of a category identified as filler is to generate a position where a copy can be inserted later on. But this is not the same as inserting the copy itself. Thus, the copy theory of movement implies a filler-driven approach to the parsing of movement dependencies, in the specific sense that fillers must be identified before traces can be inserted (see Phillips & Wagers 2007 for discussion of filler-driven parsing and references).

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

On a filler-driven strategy, incremental parsing of a string X-A-B-C-D-E-F, in which a constituent X has moved leftward from a position following C, may be conceptualized as follows. (i) Upon encountering X, the parser assumes that this category has moved and hence initiates the search for a position in which it can insert a trace (i.e. a copy of X). We indicate that X is an active filler by placing it in a box, as in 11a. (ii) If the substring A-B-C is such that it can be followed by X, then an empty syntactic position is created at that point in the string, giving rise to the transition from 11b to 11c (the symbol *e* represents the empty position). (iii) Subsequently, a copy of X is inserted and X is deactivated as a filler, as in 11d. (iv) Finally, the rest of the input string is parsed, as in 11e.

(11) Leftward movement (single movement routine with forward search)

a. $\boxed{\text{X}}$

b. $\boxed{\text{X}}$ A B C

c. $\boxed{\text{X}}$ A B C *e*

d. X A B C ~~X~~

e. X A B C ~~X~~ D E F

Thus, we may assume the following routine for establishing movement dependencies in parsing.

(12) Movement routine

a. Identify the displaced category.
b. Suspend work on it and store a copy.
c. Insert the copy.
d. Continue work on the resulting chain members (if need be).

Steps 12a–c and the order in which they come are dictated by the copy theory. Step 12d takes place after the movement dependency has been resolved and therefore is, strictly speaking, not part of the movement routine. We include it here, alongside the instruction in step 12b that work on the constituent be suspended, because moved categories may contain elements that need to interact with material located in the parser's right context when the movement routine is initiated (for example, in structures that involve remnant movement or binding under reconstruction).

Given that traces are often called gaps in the psycholinguistic literature, we emphasize that there is a fundamental distinction between a trace and an empty position of the type that temporarily follows *your* in *I am reading your letter.* As explained, an empty position is a reified prediction: it is created to host material that the parser predicts will be present. A trace, however, is a copy of information encountered in the input string. Thus, an empty position, once created, can be filled by a trace, but a trace cannot be filled by anything. This is true even if the empty position is associated with predictions about the material expected to fill it (such as the prediction that what will follow *your* is a nominal constituent).

The distinction between traces and empty positions is specific to parsers that work with a Minimalist grammar. Notice, however, that equivalent distinctions must be made for parsers that work with grammars rooted in different analytical traditions. For example, a Head-Driven Phrase Structure Grammar (HPSG) parser must distinguish between an 'unfilled' object slot (and its associated properties) as introduced by a transitive verb and an object slot whose content has been licensed through a slash feature. The former is like an empty position; the information shared through a chain of slash features is (for current purposes) analogous to a copy.[5]

The process of creating and storing copies required by the movement regime in 12 is arguably costly. It therefore stands to reason (i) that the parser will avoid movement analyses when it can, and (ii) that when a movement analysis is assumed (that is, when a filler has been identified), a trace will be inserted at the

[5] The fact that other frameworks make equivalent distinctions makes it likely that our account of the way leftward and rightward movement are parsed permits a translation to these other frameworks. We cannot explore details here.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

earliest opportunity. After all, insertion of a trace terminates the movement routine. A parsing principle that has exactly these two effects—the **minimal chain principle**—was proposed by de Vincenzi (1991). We give it here in a paraphrase due to Lyn Frazier (from the foreword to de Vincenzi's monograph).

(13) Minimal chain principle: Postulate required chain members at the earliest point grammatically possible, but postulate no potentially unnecessary chain members.

As de Vicenzi notes, the first part of 13 is closely related to the active-filler hypothesis (Stowe 1986, Frazier & Flores d'Arcais 1989, Traxler & Pickering 1996, among many others). For our current purposes, it can be read as 'Insert stored copies at the earliest point grammatically possible'. The second part of 13 can be read as 'Do not create and store copies that are potentially unnecessary'. This implies that the parser avoids movement analyses when it can.[6]

The assumptions outlined above predict both filled-gap and missing-filler effects. A filled-gap effect comes about as follows. Suppose that a category (for instance, a clause-initial WH-expression in English) is hypothesized to be a filler. Then, the procedure in 12 is triggered and a trace is stored for later insertion. The minimal chain principle drives early attempts to insert the trace. Thus, upon encountering a transitive verb, an attempt is made to integrate the trace as the object of that verb, leading to a garden-path effect if the object position is in fact filled. This explains the well-known increase in reading times for the word *us* in 14a as compared to 14b (Stowe 1986).

(14) a. My brother wanted to know who Ruth will bring **us** home to at Christmas.
b. My brother wanted to know if Ruth will bring **us** home to Mom at Christmas.

The missing-filler effect is central to our article and requires detailed discussion. We believe that the following constraint—the **leading edge corollary**—can be derived from the minimal chain principle.

(15) Leading edge corollary: A missing-filler effect is triggered at the onset of every constituent that separates a gap from the leading edge of the parse tree until a filler is found.

The gap referred to in 15 is an empty position built to host material that the parser has not yet encountered, but that it has predicted based on the input. As such a position reifies a prediction, it is generated at the leading edge of the parse tree. As long as it remains there (with incoming material integrated to its left), the minimal chain principle militates against treating the empty position as a launching site of movement. If the parser did treat the empty position in this way, it would have to start looking for a category that can fill the empty position. This, of course, is a search that the gap already required, but one that must now be carried out with the added cost associated with the movement routine (see below). However, if the empty position is removed from the leading edge (that is, if incoming material is attached to its right), it can no longer be filled in the normal way (due to the **no-tangling condition**: In any well-formed constituent structure tree, for any nodes *x* and *y*, if *x* precedes *y*, then all nodes dominated by *x* precede all nodes dominated by *y*; Partee et al. 1993:440). Hence, the prediction it expresses is defeated unless the developing parse tree is treated as a rightward movement structure. This is what we call the missing-filler effect. Once it occurs, the parser must initiate a search for a suitable displaced category as it continues processing the input string.

Consider, then, the parsing steps involved in the processing of a string A-B-C-D-E-F-X, in which an obligatory constituent X that would normally immediately follow C has moved rightward and in which elements D, E, and F are separate constituents. (i) Once the substring A-B-C has been parsed, it is predicted that X will follow—because X is obligatory—and hence a position for it is created, giving rise to the transition from 16a to 16b. (ii) The parser initially attempts to fit D into or to the left of the empty element. By assumption, neither is possible. (iii) Therefore, *e* must remain empty, and a search for a filler must be initiated while the substring D-E-F is parsed (a first instance of the missing-filler effect). The procedure for this search is more complicated than the one for leftward movement, because a position, by

[6] These interpretations are narrower than intended by de Vicenzi (1991), who adopts a definition of chain members that includes (certain) null pronouns. This more inclusive definition does not affect the argumentation in this article, and hence we do not explore its implications here.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

its very nature, cannot be buffered in the way that a copy of a moved category can be. Rather, the search must involve an instruction to apply the movement routine in 12 to every incoming constituent, checking whether its trace can be inserted into the empty position, as in 16d and 16e. (iv) It is only when X presents itself that the parser can activate a filler that fits the empty position and establish the movement dependency, as in 16f,g. Thus, each constituent between the gap and the filler (D, E, and F in 16) triggers a new missing-filler effect.

(16) Rightward movement of obligatory material (repeated movement routine with backward search)

a. A B C
b. A B C *e*
c. A B C *e* D
d. A B C *e* D E
e. A B C *e* D E F
f. A B C *e* D E F X
g. A B C ~~X~~ D E F X

The derivation in 16 treats elements D–F as single-word units, but it is of course possible that these elements are complex constituents. In that case, we expect to see a sawtooth pattern in processing times, as in Figure 1 (where D–F are each assumed to consist of three words). The greatest effort is required at the onset of each constituent, where a failed attempt at integration is made. Beyond the onset and until the process reaches the next constituent, the only additional costs the parser faces are associated with maintenance (for example, the parser needs to retain the instruction to search for a filler and cannot close any constituent containing the empty position).

The sawtooth pattern predicted for missing-filler effects has been observed for filled-gap effects. Forster et al. (2009) report on two experiments that replicated the well-known finding that object relative clauses (as in 17a) are harder to parse than subject relative clauses (as in 17b). Detailed analysis revealed that the difficulties found with object relatives were confined almost entirely to a filled-gap effect triggered by the article introducing the subject. In experiment 1, there was no significant effect for either the subsequent noun (here: *lawyer*) or the verb (here: *irritated*). In experiment 2, there was a significant effect for the article, no significant effect for the verb, and a minor effect for the noun, which was significant only in the analysis by participant.

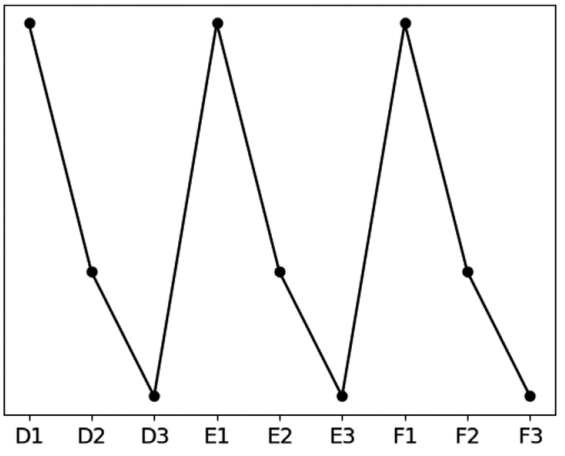


Figure 1. *The sawtooth pattern predicted for complex categories crossed by rightward movement of obligatory material.*

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

(17) a. The banker $\varnothing_1$ that the lawyer irritated t_1 played tennis every weekend.
b. The banker $\varnothing_1$ that t_1 irritated the lawyer played tennis every weekend.

There is an important interaction between minimal attachment and the leading edge corollary. The latter comes into play only when an empty position has been generated, but minimal attachment implies that such an empty position will not be generated for elements that are optional. Consequently, rightward movement of optional material will be parsed differently from rightward movement of obligatory material.

Consider a string A-B-C-D-E-F-X in which a constituent X that can optionally follow C has moved rightward. (i) In that case no empty position for X is created after the substring A-B-C (because that substring does not reliably predict that X follows). (ii) Rather, the parser simply continues to process the substring D-E-F, as in 18a,b. (iii) When the parser then encounters X, it can activate this element as a filler, as in 18c, and initiate a backward search for a position in which a copy of X can be inserted. (iv) Such a position can be created after C, as in 18d. (v) Subsequently, a copy of X is inserted, so that the movement dependency is resolved, as in 18e. Thus, this procedure resembles the procedure for leftward movement more closely than the procedure for rightward movement of obligatory material: like in 11 and unlike in 16, the leading edge corollary does not come into play, and the movement routine is applied only once.

(18) Rightward movement of optional material (single movement routine with backward search)

a. A B C
b. A B C D E F
c. A B C D E F X
d. A B C *e* D E F X
e. A B C ~~X~~ D E F X

Thus, rightward movement of obligatory material is qualitatively different from both leftward movement and rightward movement of optional material, at least on the assumptions made here. This qualitative difference predicts difficulties in the parsing of rightward movement of obligatory material that are absent otherwise.

We can test these predictions using heavy-XP shift (Ross 1967, Rochemont & Culicover 1990, and Overfelt 2015). This operation can shift an object away from a selecting verb that is either optionally or obligatorily transitive, as shown in 19 and 20.

(19) a. Ava ate (the food that her brother prepared).
b. Ava ate t_1 quickly yesterday [the food that her brother prepared]$_1$.
(20) a. Ava devoured *(the food that her brother prepared).
b. Ava devoured t_1 quickly yesterday [the food that her brother prepared]$_1$.

While the representations of 19b and 20b are indistinguishable from a syntactic point of view, the assumptions made above imply that they are parsed quite differently. Heavy-XP shift of the object of an optionally transitive verb will instantiate 18 and hence it will not give rise to missing-filler effects (see 21).

(21) Heavy-XP shift of optional object

a. Ava ate
b. Ava ate quickly yesterday
c. Ava ate quickly yesterday the food that her brother prepared
d. Ava ate *e* quickly yesterday the food that her brother prepared
e. Ava ate ~~the food that her brother prepared~~ quickly yesterday the food that her brother prepared

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

Heavy-XP shift of the object of an obligatorily transitive verb, however, will require early postulation of an empty position and hence trigger a search for the missing filler. In particular, when the parser encounters the verb *devour* it will generate an empty object position (see 22a,b). The next element in the input string is the adverb *quickly*, which the parser will attempt to insert into or to the left of the empty position. Given the grammar of English, neither is possible, triggering a missing-filler effect. Consequently, at stage 22c, an instruction is given to treat every subsequent category as a potential filler, which implies that *yesterday* (the next element in the input string) will be subjected to the movement routine (see 22d). This attempt at integration will fail, triggering a second missing-filler effect, after which the parser finally encounters the shifted object and can resolve the movement dependency by inserting a trace into the empty object position (see 22e,f).

(22) Heavy-XP shift of obligatory object

a. Ava devoured
b. Ava devoured *e*
c. Ava devoured *e* quickly
d. Ava devoured *e* quickly yesterday
e. Ava devoured *e* quickly yesterday the food that her brother prepared
f. Ava devoured ~~the food that her brother prepared~~ quickly yesterday the food that her brother prepared

Hence, the leading edge corollary makes the following prediction.[7]

(23) Heavy-XP shift of obligatory objects should cause parsing difficulties at the onset of each constituent crossed by the movement; heavy-XP shift of optional objects should not.

There is a further, more subtle, prediction. In both 21 and 22 the parser identifies the shifted object as a filler and starts a backward search to find a place to insert a copy. In 22 there is one subsequent step, namely the placement of a copy in the empty object position; in 21, however, there are two subsequent steps, as an empty object position must first be created before it can be filled by a copy of the shifted object. Thus, we expect that once the parser reaches the heavy NP, there should be a reversal in parsing effort.

(24) The processing of an obligatory object that has undergone heavy-XP shift should be easier than the processing of an optional object that has undergone this movement.

The effect described in 24 is expected to be smaller than the effect in 23, given the many steps involved in generation and execution of the instruction that each constituent following the obligatory empty object position should be treated as potential filler.

In sum, fairly standard assumptions about parsing predict that rightward movement of obligatory material is particularly costly. That prediction can be tested by looking at the way heavy-XP shift is processed in structures with optionally versus obligatorily transitive verbs.

[7] An anonymous referee disputes that optionally transitive verbs are parsed as intransitives by default. Their concern is based on psycholinguistic work by Townsend and Bever (2001) and Ferreira (2003), which suggests a preference for a transitive analysis of DP-V-DP strings where one is potentially available. Thus, in a sentence like *When Mary bathed the baby played with a ball*, the main-clause subject is initially parsed as the object of the verb that heads the adverbial clause (see Christianson et al. 2001 and references mentioned there). The apparent contradiction between this finding and the preceding discussion dissolves once it is acknowledged that two logically independent questions are in play. The first is whether the parser by default analyzes a DP that immediately follows an optionally transitive verb as the object of that verb; the second is whether the parser by default generates an object position when it encounters an optionally transitive verb (irrespective of whether a DP follows). The literature indeed shows that the first question should receive an affirmative answer (in our view because of late closure and/or minimal attachment: see Frazier 1978; see also Grillo & Costa 2014). However, as explained in the main text, we believe that the **second** question should be answered in the negative. This is not in conflict with the research reported in Townsend & Bever 2001 or Ferreira 2003, as neither presents evidence that bears on the second of our two questions.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

The predictions in 23 and 24 can be contrasted with those made by a uniform parsing strategy for heavy-XP shift. On that alternative, all instances of heavy-XP shift are parsed in the same way, based on the hypothesis that no gap is created until a filler is identified. Thus, 25a is followed by 25b, rather than 22b. A uniform account still predicts parsing difficulties at the first adverbial following an obligatorily transitive verb like *devour* (as marked by the dashed box in 25b). This is because the expectation that an object will follow is defeated (we refer to this effect as **category-based surprisal**). However, no parsing difficulties are predicted at the second adverbial, nor would one expect an advantage at the moved object with obligatorily transitive verbs. As in 21, a position must be created after identification of the filler (step 25e) and a copy inserted (step 25f).

(25) Heavy-XP shift of obligatory object (alternative uniform analysis)

a. Ava devoured
b. Ava devoured quickly
c. Ava devoured quickly yesterday
d. Ava devoured quickly yesterday the food that her brother prepared
e. Ava devoured *e* quickly yesterday the food that her brother prepared
f. Ava devoured ~~the food that her brother prepared~~ quickly yesterday the food that her brother prepared

The literature already contains work that suggests the predictions in 23 and 24 are correct. Staub et al. (2006) carried out two eye-tracking experiments intended to explore how the selectional requirements of the verb influence parsing of heavy-XP shift. In both experiments, evidence of processing difficulty appeared on the material that intervened between the verb and its object, but only when the verb was obligatorily transitive. In addition, Staub et al. found signs of increased effort in the processing of a shifted direct object when the verb was optionally transitive.

In Section 3, we report on two web-based experiments intended to replicate and extend the findings of Staub et al.: a self-paced reading task (Just et al. 1982, Mitchell 1984) and a maze task (Freedman & Forster 1985, Forster et al. 2009, Boyce et al. 2020). The self-paced reading task confirms the prediction in 23, while the more sensitive maze task confirms both predictions. In other words, the effects Staub et al. found are stable across experimental paradigms. The extension of the work by Staub et al. concerns sentences in which heavy-XP shift crosses not one, but two modifiers. As explained, the leading edge corollary predicts that parsing difficulties are found at the onset of each modifier, a prediction that is borne out.

3. Testing the predictions: heavy-XP shift

3.1. Literature review and experimental logic

As well as direct support for the predictions in 23 and 24 from the above-mentioned eye-tracking experiments by Staub et al. (2006), the literature provides corpus-linguistic evidence that is compatible with our hypothesis of increased parsing effort associated with rightward movement of obligatory material.

Wasow (1997a,b) carried out two corpus studies that both reveal that the verb's selectional requirements influence the frequency of heavy-XP shift of the object. Wasow distinguished obligatorily transitive verbs (V*t*) from verbs that optionally occur with NP objects but also have uses with an immediately following PP and no NP object (V*p*). Heavy-XP shift of the object occurred about twice as much in the V*p* condition as in the V*t* condition. The effect is found in both written and spoken corpora, as Table 2 shows.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

Brown corpus (written) $\chi^2(1) = 6.49, p < 0.02$		Switchboard corpus (spoken) $\chi^2(1) = 10.65, p < 0.01$	
V*t*	V*p*	V*t*	V*p*
5.6%	9.3%	1.45%	3.82%

Table 2. *Percentage of heavy-XP-shifted DP-objects in two corpora as determined by verb transitivity.*

These findings seem relevant, as they are suggestive of some kind of difficulty with heavy-XP shift of obligatory objects. Wasow locates this difficulty in the speaker, but as we now explain this is compatible with our proposal that it has its source in the parser.

Hawkins argues in several publications (see in particular Hawkins 1990, 2009) that parsing effort is inversely correlated with corpus frequency. If so, Wasow's findings are as predicted by the hypothesis that rightward movement of obligatory objects requires increased parsing effort (as compared to rightward movement of optional objects). The weight of this argument depends on the strength of the theory linking parsing difficulties to reduced production, so let us briefly explore what such a theory might look like. It stands to reason that speakers learn to avoid structures that are hard to produce, with the consequence that such structures will be less frequent in corpora. Production difficulties can of course be specific to this modality. However, there is no doubt that speakers monitor their production by parsing their own speech. This was argued explicitly by Levelt (1989) and is supported by experiments that show a breakdown in production when subjects hear their own speech slightly delayed through headphones (see Jackendoff 1987 and references given there). Self-monitoring therefore implies that parsing difficulties will affect not only perception, but also production.

As an example, consider the parsing and production of relative clauses. Santi et al. (2019) carried out a self-paced reading experiment and two planned production experiments to determine how production and perception compare in this domain. Items consisted of sentences containing subject and object relatives that were either center-embedded or located at the right edge of the matrix clause. While the production experiments uncovered effects specific to this modality, they also showed many of the same effects found in the self-paced reading experiment. For example, the production results replicated the self-paced reading finding that a relative clause has a significantly longer duration with an object gap than with a subject one, an effect localized on the embedded verb and attributed to the resolution of the movement dependency between the relative pronoun and the object position in which it originates (any movement in subject relatives is string-vacuous and hence resolved more easily).

The fact that parsing effort affects production in this way confirms the importance of self-monitoring. This, in turn, provides a theoretical grounding for Hawkins's claim that parsing difficulty suppresses corpus frequency.

As mentioned, more direct confirmation of the predictions in 23 and 24 comes from two eye-tracking experiments reported by Staub et al. (2006). The experiments were designed to determine how the selectional requirements of the verb influence parsing of heavy-XP shift. On the assumption that parsing effort gives rise to slower average reading times, both experiments showed evidence of processing difficulty on modifiers intervening between a verb and an obligatory object. Such modifiers showed a considerable slowdown on all measures as compared to intervening modifiers in the optionally transitive condition. This finding of course conforms well with our expectations.

We should note that the effect was categorical rather than gradient. The verbs that Staub et al. used in their second experiment showed considerable variation in their statistical preference for use in a transitive frame. However, the reported effect was not correlated with this preference. Instead, it was found only with obligatorily transitive verbs (i.e. verbs that required use in a transitive frame). This suggests that an alternative analysis in terms of gradient category-based surprisal (surprisal at finding an unexpected category) is insufficient. After all, the degree of surprisal upon finding an adverbial where an object is expected should be correlated with the strength of the expectation that an object should be there in the first place.

The findings reported by Staub et al. (2006) are summarized in Table 3. As the table shows, Staub et al. also found a difference in the processing of the shifted object. They suggest that this difference is

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

	Adverbial	Object
V_{OblTr}	slower on all measures	faster on some measures
V_{OptTr}	faster on all measures	slower on some measures

Table 3. *Comparison of reading times in two regions of sentences with a heavy-XP-shifted object.*

indicative of increased parsing effort in the processing of a shifted direct object when the verb is optionally transitive. With such verbs, there was a slowdown on some measures compared to the processing of shifted obligatory objects. The interpretation that Staub et al. give of these data is based on an assumption about the relative reliability of the measures involved. The optionally transitive condition showed a faster first-pass reading time on the shifted object, while the obligatorily transitive condition showed fewer regressions. Staub et al. (2006:396) write that 'the overall results support the notion that the noun phrase region was easier to process when the verb was obligatorily transitive, since the percent regressions data reflect a very small proportion of trials, and the first pass data, by contrast, reflect processing on every trial'. These findings thus confirm the prediction in 24.

Given the importance of the predictions in 23 and 24 for our overall argumentation, we ran two additional experiments. These were in the first instance intended to replicate the findings of Staub et al. using different experimental paradigms (self-paced reading and a maze task). However, we added an additional factor to the experimental set-up: a manipulation of the number of adverbials crossed by heavy-XP shift of the object. This is shown in Table 4.

	One adverbial	Two adverbials
V_{OblTr}	Subject V_{OblTr} *t* **Adv** Object	Subject V_{OblTr} *t* Adv_{delay} **Adv** Object
V_{OptTr}	Subject V_{OptTr} *t* **Adv** Object	Subject V_{OptTr} *t* Adv_{delay} **Adv** Object

Table 4. *Structures tested in the self-paced reading and maze tasks.*

The point of this extension was to determine whether any slowdown on the material separating verb and object persists beyond the first adverbial or is a temporary effect of encountering an obligatorily transitive verb without an object (that is, an instance of category-based surprisal). As explained in the previous section, category-based surprisal is the only effect predicted by a uniform account of the parsing of heavy-XP shift. By contrast, the leading edge corollary predicts that with obligatorily transitive verbs we should find a missing-filler effect at the onset of each constituent separating the gap from the filler (induced by a failed attempt at integration). Therefore, reaction times over the adverbial of interest (boldfaced in the table) should remain longer in the obligatorily transitive condition, whether a 'delay adverbial' (Adv_{delay}) is present or not. By contrast, if the slowdown reported by Staub et al. is merely the effect of the parser not finding an object where one is expected, reaction times should return to normal as the point of this category-based surprisal recedes.

In sum, in both experiments we varied the transitivity of the verb (obligatory versus optional) and the presence or absence of a delay adverbial. We measured reaction times, with two regions where predictions are tested: the adverbial of interest and the first few words of the shifted object. The predictions under scrutiny are the following.

(26) a. Reaction times at the onset of the adverbial of interest will be slower in the obligatorily transitive condition than in the optionally transitive condition, irrespective of the presence or absence of a delay adverbial.
b. Reaction times over the first few words of the shifted object will be faster in the obligatorily transitive condition than in the optionally transitive condition.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

These predictions are operationalizations of predictions 23 and 24 in terms of reaction times.

We report on the two experiments in the following subsections.

3.2. Experiment 1: self-paced reading

The first experiment we ran was a moving-window self-paced reading study.[8] Self-paced reading (Just et al. 1982, Mitchell 1984) is an experimental technique in which participants read a sentence on a computer screen. The sentence starts off masked, and the participant presses a button to reveal each successive word and mask the previous word (with the view window shifting rightward). The time between button presses is what is being measured; this response time corresponds to the reading time of the unmasked word. Although self-paced reading is less precise than eye tracking, especially because spillover effects can affect its temporal resolution, it has been widely and successfully employed in psycholinguistics.

We ran our experiment online rather than in the lab, using Gorilla as the experimental platform (https://app.gorilla.sc) and Prolific to recruit participants (https://www.prolific.co/). This method was pioneered by Enochson and Culbertson (2015), who showed that three classic psycholinguistic effects could be replicated using web-based data gathering on comparable platforms.

Our materials consisted of ninety-six test sentences and forty-eight fillers. The test sentences were either identical to or adapted from those used by Staub et al. (2006). They were grouped into sets of four sentences. Each set made up one experimental item and was shaped as in Table 3. Across all items, the subject, the adverbial of interest, and the shifted object were held constant. The verb varied and was either optionally or obligatorily transitive. The items also varied in the presence or absence of a delay adverbial, which we added to the items that Staub et al. used and which was held constant across test sentences that had one. A sample item made up of four test sentences is given in 27. We have underlined relevant regions. Region 1 is the verb. Regions 3 and 4 are the adverbial of interest and the onset of the shifted object, respectively. Region 2 is the delay adverbial. In the example at hand, *attack* is optionally and *bother* obligatorily transitive.

(27) a. Sara attacked$_1$ with no mercy$_3$ the red ants$_4$ living on the windowsill.
b. Sara attacked$_1$ for ten minutes$_2$ with no mercy$_3$ the red ants$_4$ living on the windowsill.
c. Sara bothered$_1$ with no mercy$_3$ the red ants$_4$ living on the windowsill.
d. Sara bothered$_1$ for ten minutes$_2$ with no mercy$_3$ the red ants$_4$ living on the windowsill.

The fillers were grammatical sentences which were comparable in complexity to the test items and in which some category had been fronted or extraposed.

Thus, the experiment had a 2 × 2 design, with the nature of the verb (obligatorily versus optionally transitive) and the presence or absence of a delay adverbial as the variables we manipulated, and reading times over the adverbial of interest and the onset of the heavy NP as dependent variables. This allowed us to test the predictions in 26. Prediction 26a is confirmed if (i) reading times at the onset of the adverbial of interest are slower in the obligatorily transitive conditions and (ii) there is no interaction with the presence or absence of a delay adverbial. In other words, the effect remains in place whether a delay adverbial is present or not. Prediction 26b is confirmed if reading times at the onset of the heavy NP are faster in the obligatorily transitive conditions.

We ran the experiment in Latin-square fashion. The test items were distributed over four lists in such a way that each list contained exactly one sentence from each item and equal numbers of sentences per condition. In addition, all fillers were added to all lists, so that each list contained twenty-four test sentences and forty-eight fillers. The fillers and test sentences were presented to the participants in pseudo-randomized order (randomization was restricted by the constraint that any two

[8] Stimuli and tasks are available at Gorilla (https://app.gorilla.sc/openmaterials/1263243). Raw data and analysis scripts are available at Trolling (https://doi.org/10.18710/HWWGL5).

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

test sentences were separated by at least one filler). All experimental and filler sentences were followed by a comprehension question in the form of a statement that participants needed to mark as true or false.

We recruited 116 participants, who were paid £4.50 for about thirty minutes of work. Of these, thirty-four participants were excluded because they reported a language other than English as their native language (nine), because they failed to answer at least 80% of the comprehension questions correctly (fifteen), because they reported reading some or all of the sentences out loud (three), or because they were reading at a speed more than 2.5 standard deviations slower than the overall mean (seven). This left eighty-two participants whose data were subjected to analysis.

In data cleaning, all 307 trials with a response time below 250 ms or above 2500 ms were discarded, as were all 30,470 responses for sentences for which the participant responded incorrectly to the comprehension question. The remaining 101,695 response times were log-transformed and residualized (as described in Jaeger 2008). Filler trials were subsequently stripped out, and the resulting 16,961 responses were subjected to analysis.

In Figures 2 and 3, residual reading times are plotted separately for the conditions with one and with two intervening adverbials. Error bars represent the standard error of the mean. As the graphs show, residual reading times are longer for obligatorily transitive verbs across both the adverbial of interest and the delay adverbial, with spikes at the onset of each. The difference between obligatorily and optionally transitive verbs disappears at the heavy NP. The four conditions are plotted together in Figure 4.

We fitted seven linear mixed-effects models to the residual reading times with contrast-coded conditions for the number of adverbials and the optional/obligatory transitivity of the verb. To do so,

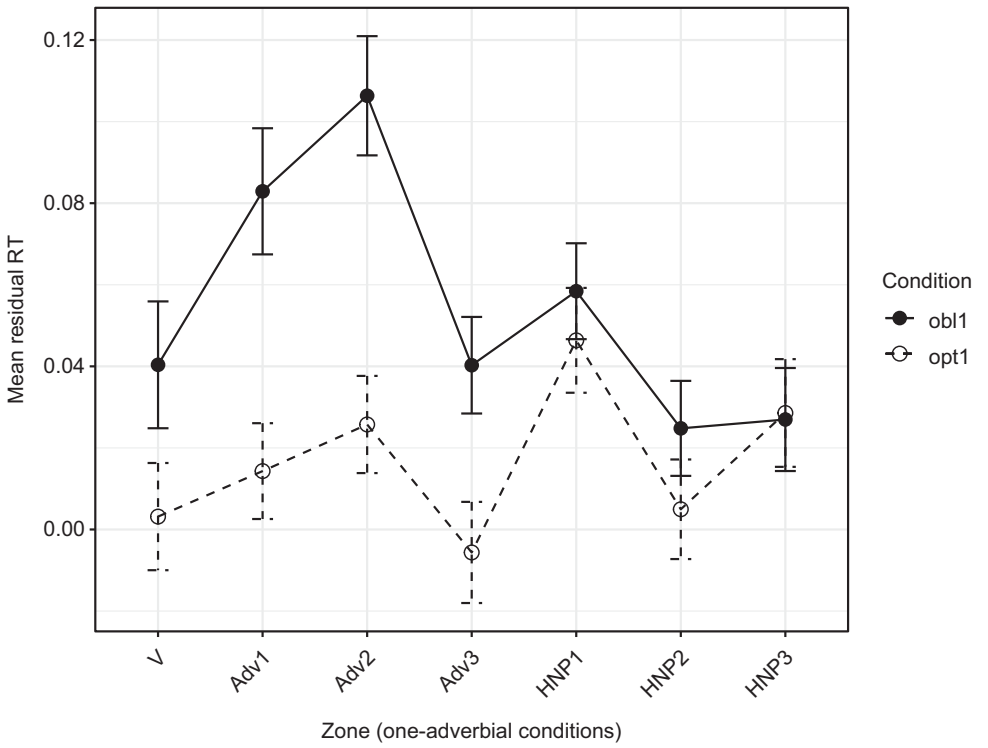


Figure 2. *Plot for residual reading times in the one-adverbial condition.* Note*: obl: obligatorily transitive verb, opt: optionally transitive verb.*

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

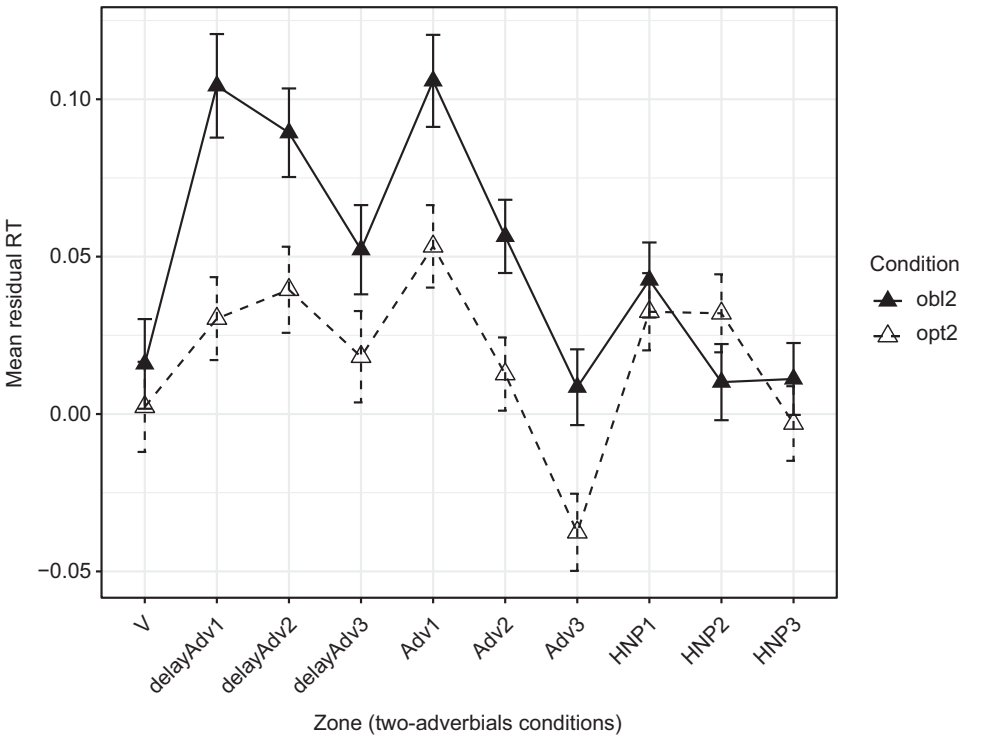


Figure 3. *Plot for residual reading times in the two-adverbials condition.* Note*: obl: obligatorily transitive verb, opt: optionally transitive verb.*

we used the lmer function of the lmerTest package (Kuznetsova et al. 2017) in R (R Core Team 2023): logRTresidual ~ OBL*PPS + (1 | itemnumber) + (1 | participant).

The random-effects structure was limited in order to ensure that all models would converge non-defectively. *P*-values were computed using the Satterthwaite approximation. Significance levels were calculated using Bonferroni correction to take into account the fact that seven models were fitted. Fixed effects for these models are summarized in Table 5.

As can be seen, our results show a significant main effect of the optional/obligatory transitivity of the verb on all three words of the adverbial of interest (Adv1, Adv2, and Adv3), but no main effect of the presence of a delay adverbial and no interaction. In other words, the slowdown over the adverbial of interest in the obligatorily transitive conditions occurs independently of whether this adverbial immediately follows the verb or appears further downstream, after the delay adverbial. No other effects were significant. In particular, obligatory transitivity did not affect reading times over the onset of the heavy NP.

This means that the data generated in our first experiment confirm prediction 26a, but not prediction 26b. In other words, we were able to replicate and extend the stronger of the two effects that Staub et al. found in their eye-tracking study, namely the evidence for parsing difficulties on the material separating an obligatorily transitive verb from an object that has undergone heavy-XP shift, and the absence of such difficulties when the verb is optionally transitive.

In the analysis reported above, the delay adverbial was excluded. However, the leading edge corollary predicts that the slowdown observed over the adverbial of interest when the verb is obligatorily transitive should also affect reaction times on the delay adverbial. We therefore ran an additional analysis that included the delay adverbial (with ten models, the three models for the delay adverbial missing all one-

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

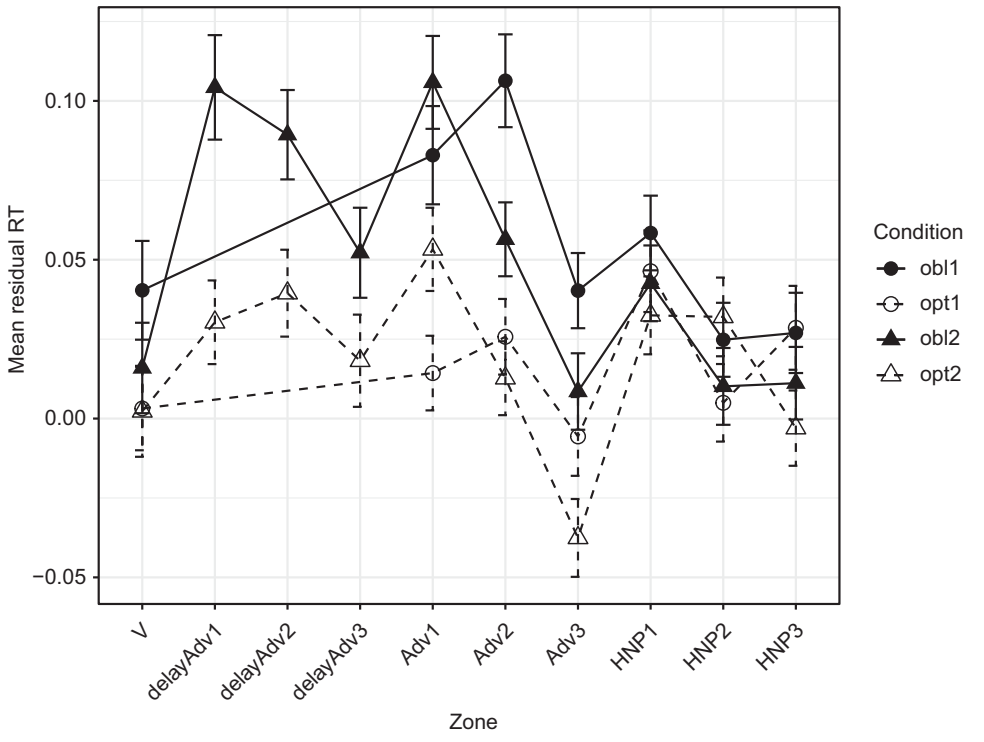


Figure 4. *Combined reading time plots.* Note*: obl: obligatorily transitive verb, opt: optionally transitive verb; 1: one intervening adverbial, 2: two intervening adverbials.*

adverbial conditions, of course). This analysis found a significant effect of obligatory transitivity on the three words of the adverbial of interest. In addition, it found a significant slowdown in the obligatorily transitive condition on the first word of the delay adverbial, in line with expectations. No other effects were significant (after Bonferroni correction).

Thus, our data seem to display the expected sawtooth effect, rather than the transitory effect of the parser encountering an obligatorily transitive verb without an object. To make this clearer, we plot the differences in mean residual reading times between the obligatory and the optionally transitive conditions in Figure 5. As can be seen, there is substantial slowdown right after the obligatorily transitive verb in both conditions and a renewed substantial slowdown at the beginning of the second adverbial in the two-adverbials condition. The advantage for the optionally transitive condition persists across both adverbials in the two-adverbials conditions.

It seems, then, that parsing difficulties persist, rather than subside, after the moment at which the parser's initial expectation is defeated. The leading edge corollary predicts such an effect at the onset of the adverbials. We would attribute the fact that evidence of increased parsing effort is found beyond these onsets at least partly to maintenance costs (the instruction to look for a filler must be maintained, and the VP with the empty position cannot be closed). However, given the poor temporal resolution of self-paced reading tasks, we cannot exclude that spillover effects play a role as well. We return to this in the discussion of our second experiment.

The fact that the experiment failed to confirm that obligatory transitivity of the verb facilitates parsing of the shifted object is not all that surprising. After all, self-paced reading is less sensitive than eye tracking, and the effect on the shifted object was relatively weak in the original study by Staub et al. (But note the negative value of the dashed line in Figure 5 indicating a relative speed-up in the obligatorily transitive versus the optionally transitive two-adverbials condition on the second word of the heavy NP.)

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

Zone	Term	Est.	*SE*	*t*-value	*df*	uncorrected *p*	corr. sig.
V	(intercept)	0.016	0.0149	1.07	50.2	0.29090	ns
V	OBL	0.025	0.0136	1.87	1890.2	0.06218	ns
V	ADV	−0.013	0.0137	−0.92	1891.2	0.35516	ns
V	OBL:ADV	−0.021	0.0273	−0.78	1889.1	0.43286	ns
Adv1	(intercept)	0.065	0.0127	5.12	31.7	1e–5	***
Adv1	OBL	0.060	0.0135	4.46	1895.5	1e–5	****
Adv1	ADV	0.030	0.0135	2.24	1897.1	0.02518	ns
Adv1	OBL:ADV	−0.016	0.0269	−0.59	1894.8	0.55806	ns
Adv2	(intercept)	0.051	0.0097	5.19	30.3	1e–5	****
Adv2	OBL	0.062	0.0123	5.04	1895.2	0.00000	****
Adv2	ADV	−0.031	0.0123	−2.52	1897.0	0.01168	ns
Adv2	OBL:ADV	−0.036	0.0246	−1.46	1894.2	0.14448	ns
Adv3	(intercept)	0.002	0.0102	0.15	32.2	0.88051	ns
Adv3	OBL	0.046	0.0119	3.84	1896.2	1.3e–4	***
Adv3	ADV	−0.032	0.0119	−2.68	1898.0	0.00747	ns
Adv3	OBL:ADV	0.000	0.0238	0.00	1895.2	0.99910	ns
HNP1	(intercept)	0.045	0.0099	4.54	30.0	9e–5	***
HNP1	OBL	0.011	0.0120	0.90	1897.6	0.37019	ns
HNP1	ADV	−0.015	0.0120	−1.23	1898.4	0.21945	ns
HNP1	OBL:ADV	−0.003	0.0240	−0.12	1897.1	0.90491	ns
HNP2	(intercept)	0.018	0.0122	1.49	35.5	0.14478	ns
HNP2	OBL	−0.001	0.0117	−0.11	1892.0	0.91110	ns
HNP2	ADV	0.006	0.0117	0.55	1892.5	0.58239	ns
HNP2	OBL:ADV	−0.042	0.0234	−1.82	1891.0	0.06914	ns
HNP3	(intercept)	0.017	0.0151	1.16	30.8	0.25626	ns
HNP3	OBL	0.006	0.0118	0.50	1895.3	0.61840	ns
HNP3	ADV	−0.022	0.0118	−1.85	1895.9	0.06516	ns
HNP3	OBL:ADV	0.014	0.0236	0.61	1894.9	0.53938	ns

Table 5. *Model parameters for fixed effects of obligatoriness and number of adverbials on residual log RTs at the verb, the adverbial of interest, and the heavy object with Bonferroni corrected significance (self-paced reading task).* Note*:* *** p < *0.001,* **** p < *0.0001.*

This lack of replication is one reason why we carried out a second experiment in a different and arguably more sensitive paradigm.

3.3. Experiment 2: A maze task

The second experiment that we ran to determine the effects of obligatory transitivity on the processing of heavy-XP shift made use of a methodology known as a **maze task** (Forster et al. 2009).[9] Like self-paced reading, a maze task has participants read a sentence word by word. However, whereas in self-paced reading only a single word is presented to the participant at a time, in a maze task participants are presented with a forced choice at each word position. The choice is between the target word, which allows a legitimate continuation of the sentence, and a distractor that does not fit with the preceding string. Participants are instructed to press a button corresponding to what they think is the correct word, with

[9] Stimuli and tasks are available at Gorilla (https://app.gorilla.sc/openmaterials/1288197). Raw data and analysis scripts are available at Trolling (https://doi.org/10.18710/HWWGL5).

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

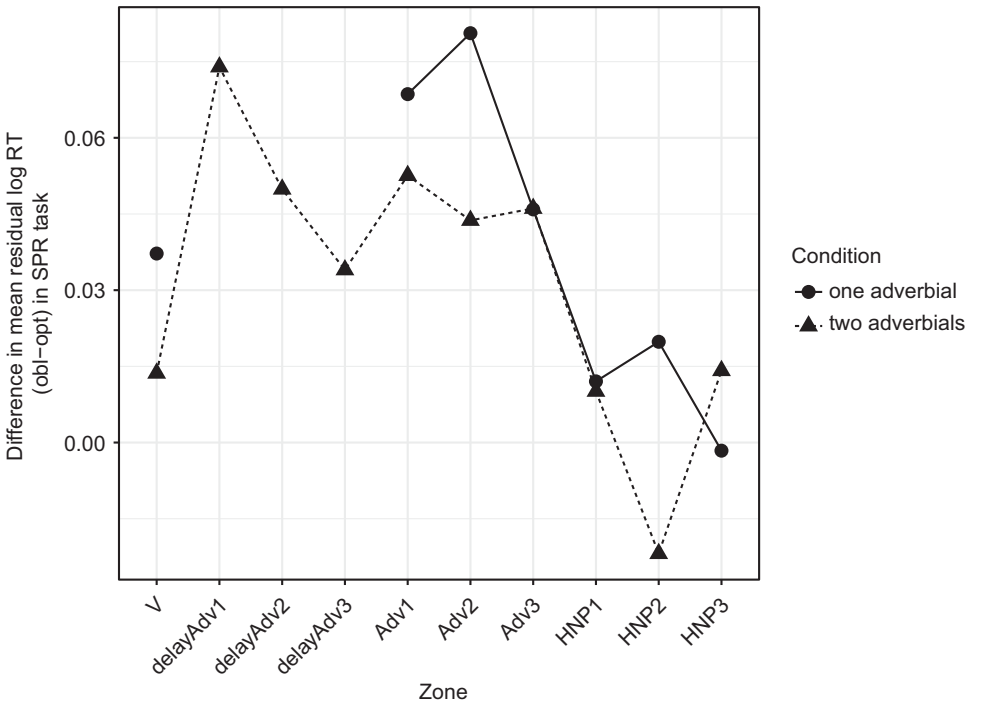


Figure 5. *Plot of mean differences in residual reading times in obligatorily versus optionally transitive conditions in the self-paced reading experiment.*

reaction time used as the dependent measure. If the participant chooses the correct word, a new pair of words is presented on the screen; if the participant chooses the wrong word, a message appears on the screen telling them that they made the wrong choice and instructing them to try again. (Other versions of the maze task move on to the next sentence at this point, but this may be frustrating for participants or even incentivize them to answer incorrectly in an online version of the experiment.)

The variant of the maze task that we use is known as G(rammatical)-maze. In this variant, the distractors are existing words (rather than nonce words, which are used in a variant known as L(exical)-maze). One of the challenges for researchers designing a G-maze task is the effort required to identify suitable distractor words, that is, words that are comparable to their correct counterparts in frequency and length but that are a poor fit with the preceding string. This problem has been addressed by Boyce et al. (2020), who leveraged natural language processing technology to automate the generation of distractors. In particular, the automation procedure made use of a neural net model to select words that have a high surprisal value given the context. From these, a distractor is chosen that matches the target word in length (as counted in characters) and frequency (as determined on the basis of the Google books Ngrams corpus; Michel et al. 2011). Of course, high surprisal does not necessarily mean that distractor items are ungrammatical in the context; they can simply be semantically incongruous. This is not a problem, however, as participants should still choose the target word over an incongruous distractor if they have parsed the input successfully. Boyce et al. (2020) validated their A(uto)-maze method by showing that it has dramatically superior statistical power and localization for well-established syntactic ambiguity resolution phenomena compared to self-paced reading.

As before, we ran our experiment online, using Gorilla as the experimental platform and Prolific to recruit participants. The target sentences were largely identical to those used in the earlier self-paced reading study. However, we increased the number of words preceding the verb in order to minimize errors that can occur in maze tasks at the beginning of sentences. In addition, each target word was paired with a suitable distractor. These were generated by A-maze, but subsequently checked by hand, which in some

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

cases led to adjustments. All in all, there were twenty-four sets of four sentences shaped as in Table 4. Each set made up an item crossing the two factors of optional versus obligatory transitivity with the number of intervening adverbials (one or two). In each item, the subjects, the adverbial of interest, the shifted object, and the delay adverbial (if present) were held constant. The verb varied and was either optionally or obligatorily transitive. The delay adverbial, which we added to original items that Staub et al. used, was the same in the test sentences that had one.

A sample quartet of test sentences is given in 28. In this sample, the target words appear leftmost in the columns labeled a–d, with distractors to their immediate right. Words in regions relevant to the experiment are underlined (region 1 is the verb, region 2 is the delay adverbial, region 3 is the adverbial of interest, and region 4 is the onset of the shifted object). In this sample, *attack* is optionally and *bother* obligatorily transitive.

(28)

a.		b.		c.		d.	
The	x-x-x	The	x-x-x	The	x-x-x	The	x-x-x
woman	yours	woman	yours	woman	yours	woman	yours
in	mid	in	mid	in	mid	in	mid
the	than	the	than	the	than	the	than
old	nor	old	nor	old	nor	old	nor
house	rates	house	rates	house	rates	house	rates
attacked$_1$	patience	attacked$_1$	patience	bothered$_1$	patience	bothered$_1$	patience
with$_3$	blog	for$_2$	tool	with$_3$	blog	for$_2$	tool
no$_3$	ha	ten$_2$	app	no$_3$	ha	ten$_2$	app
mercy$_3$	tends	minutes$_2$	improve	mercy$_3$	tends	minutes$_2$	improve
the$_4$	cent	with$_3$	blog	the$_4$	cent	with$_3$	blog
red$_4$	ago	no$_3$	ha	red$_4$	ago	no$_3$	ha
ants$_4$	vivo	mercy$_3$	tends	ants$_4$	vivo	mercy$_3$	tends
living	videos	the$_4$	cent	living	videos	the$_4$	cent
on	gain	red$_4$	ago	on	gain	red$_4$	ago
the	glad	ants$_4$	vivo	the	glad	ants$_4$	vivo
windowsill	stipulates	living	videos	windowsill	stipulates	living	videos
		on	gain			on	gain
		the	glad			the	glad
		windowsill	stipulates			windowsill	stipulates

The fillers of course also had distractors paired with target words. They were grammatical sentences with some category fronted or extraposed and had a complexity comparable to that of the test items.

Thus, the maze experiment had the same 2 × 2 design as our earlier self-paced reading study, with the verb (obligatorily or optionally transitive) and the presence or absence of a delay adverbial as the variables we manipulated, and reading times over the adverbial of interest and onset of the heavy NP as dependent variables. As before, this allowed us to test the predictions in 26. Confirmation of prediction 26a requires (i) that reading times over the adverbial of interest are slower in the obligatorily transitive conditions and (ii) that there is no interaction with the presence or absence of a delay adverbial. Confirmation of prediction 26b requires that reading times at the onset of the heavy NP are faster in the obligatorily transitive conditions.

In line with the experiment's 2 × 2 design, the test items were distributed over four lists in Latin-square fashion (so that each list contained exactly one sentence from the twenty-four items and equal numbers of sentences in each condition). In addition, the lists contained the full set of forty-eight fillers. Fillers and test sentences were presented to the participants in pseudo-randomized order (randomization was restricted by the constraint that any two test sentences were separated by at least one filler). Before starting the task, participants had to complete five practice items in order to get used to the paradigm.

We recruited 119 participants who were paid £6.00 for completing the task (which took about forty minutes). One participant was excluded because they failed to respond correctly at least 80% of the time. All remaining participants responded correctly at least 91.01% of the time (mean 97.93). A further five

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

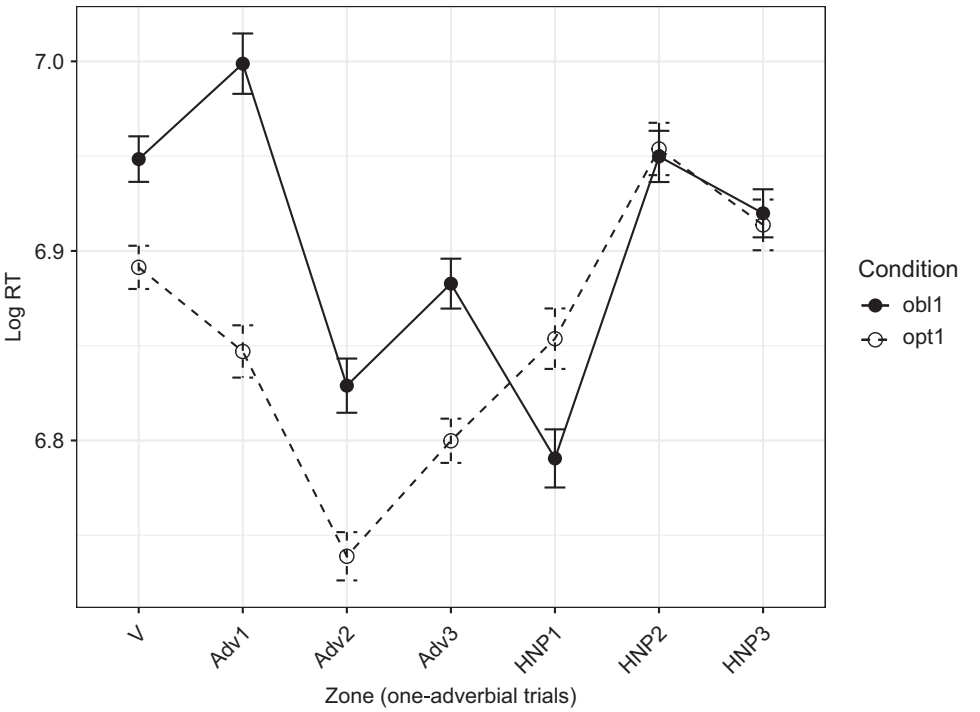


Figure 6. *Plot for choice times in the one-adverbial condition.* Note*: obl: obligatorily transitive verb, opt: optionally transitive verb.*

participants were excluded because they indicated that their native language was not English. The data generated by the remaining 113 participants were submitted to further analysis.

The experiment generated a large number of data points (reaction times per target word as measured from appearance on the screen to selection). Before the analysis of the reaction times in the areas of interest, data were excluded for all sentences in which a participant responded incorrectly at any word (before, during, or after the areas of interest). This led to the exclusion of 6,992 data points, leaving 15,017 data points in place for analysis (we should note, however, that this particular way of excluding data did not materially influence the results). The data from sentences parsed without incorrect selections are plotted in Figures 6 and 7. These figures show plots for the conditions with one and two intervening adverbials, respectively. Figure 8 shows the combined plot. In all plots, the error bars indicate the standard error of the mean.

Two aspects of the plot for the one-adverbial condition (Figure 6) stand out. In the first area of interest (Adv1–Adv3), the solid line, which represents data for items with obligatorily transitive verbs, is systematically above the dashed line, which represents data for items with optionally transitive verbs. In the second area of interest, the lines cross on HNP1 and then effectively merge. In other words, reaction times in the obligatorily transitive condition are slower while the intervening adverbial is parsed, faster at the onset of shifted object, and similar after that. The plot for the two-adverbials condition (Figure 7) shows a similar pattern. Over the delay adverbial and the adverbial of interest (delayAdv1–delayAdv3 and Adv1–Adv3), the solid line, which represents data for items with obligatorily transitive verbs, is systematically above the dashed line, which represents data for items with optionally transitive verbs. There is a reversal on the first word of the heavy NP, after which the lines come together.

For analysis, conditions were contrast coded. The log-transformed reaction times of these data were analyzed at each word (excluding the delay adverbial) in seven separate linear mixed-effects models using the lmerTest Package (Kuznetsova et al. 2017) in R (R Core Team 2023): log(RT) ~ OBL*Adv+ (1 | Item) + (1 | ParticipantID). The random-effects structure was kept simple to ensure nondegenerate convergence

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

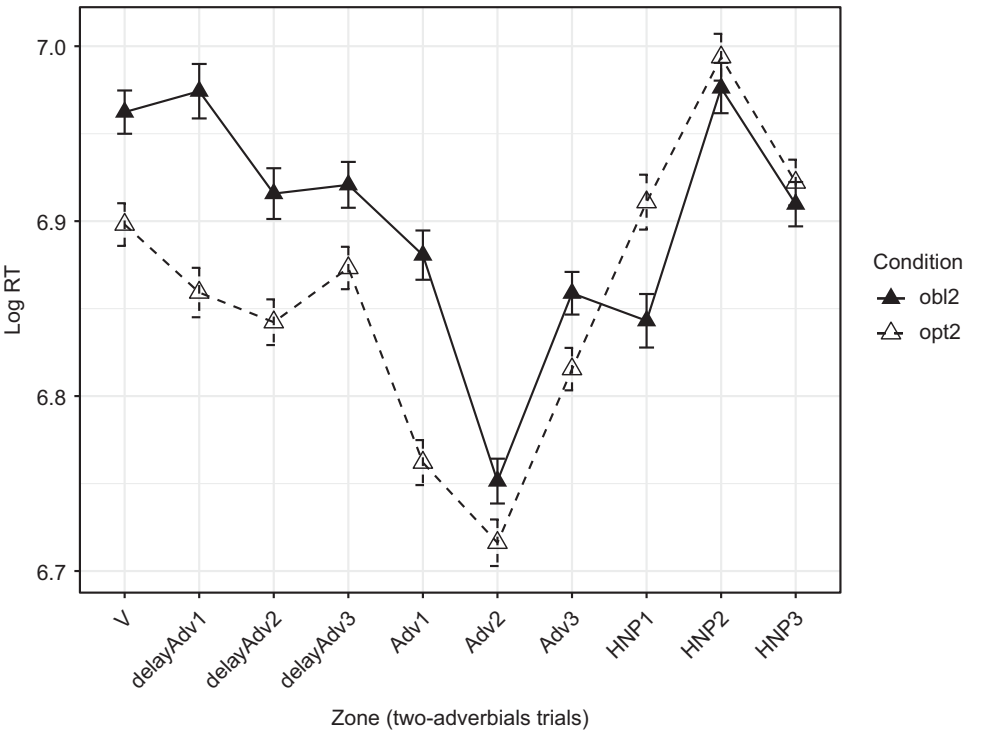


Figure 7. *Plot for choice times in the two-adverbials condition.* Note*: obl: obligatorily transitive verb, opt: optionally transitive verb.*

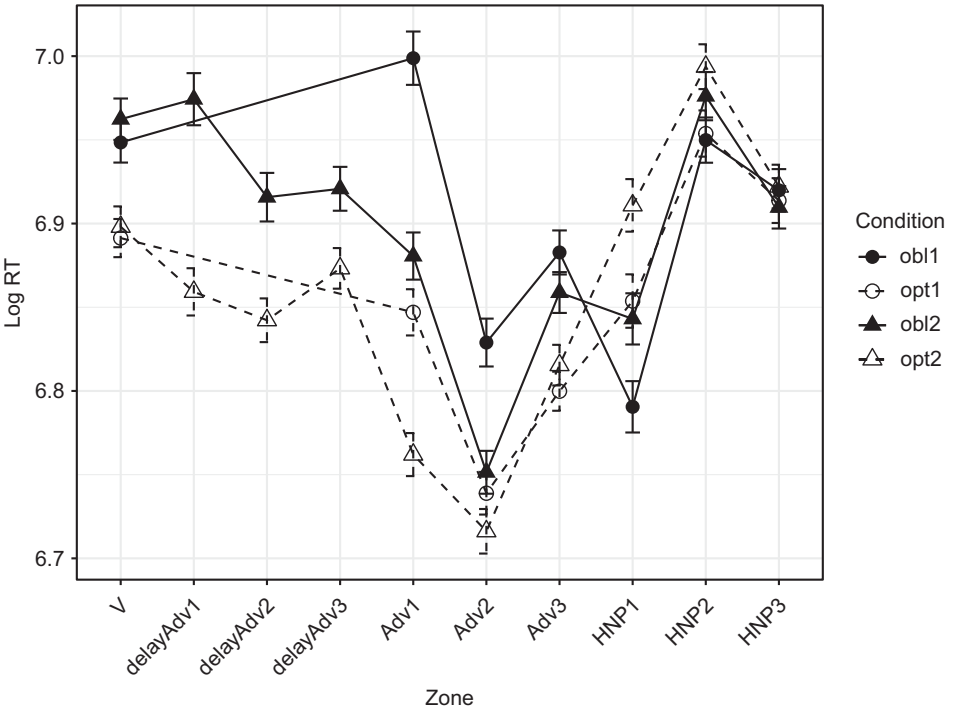


Figure 8. *Combined choice time plots.* Note*: obl: obligatorily transitive verb, opt: optionally transitive verb; 1: one intervening adverbial, 2: two intervening adverbials.*

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

Zone	Term	Est.	*SE*	*t*-value	*df*	uncorrected *p*	corr. sig.
V	(intercept)	6.929	0.023	304.22	54.9	0.00000	****
V	OBL	−0.078	0.012	−6.51	1677.1	0.00000	****
V	ADV	0.016	0.012	1.29	1675.9	0.19623	ns
V	OBL:ADV	−0.026	0.024	−1.07	1668.7	0.28274	ns
Adv1	(intercept)	6.888	0.026	263.43	55.3	0.00000	****
Adv1	OBL	−0.149	0.015	−10.25	1649.9	0.00000	****
Adv1	ADV	−0.107	0.015	−7.37	1649.7	0.00000	****
Adv1	OBL:ADV	0.059	0.029	2.05	1641.9	0.04093	ns
Adv2	(intercept)	6.769	0.041	165.99	30.3	0.00000	****
Adv2	OBL	−0.065	0.012	−5.46	1687.2	0.00000	****
Adv2	ADV	−0.049	0.012	−4.12	1684.8	4.00E–05	***
Adv2	OBL:ADV	0.060	0.024	2.55	1680.0	0.01098	ns
Adv3	(intercept)	6.844	0.026	265.27	55.6	0.00000	****
Adv3	OBL	−0.069	0.012	−5.72	1680.8	0.00000	****
Adv3	ADV	0.001	0.012	0.07	1680.5	0.94370	ns
Adv3	OBL:ADV	0.027	0.024	1.15	1674.1	0.24977	ns
HNP1	(intercept)	6.862	0.031	218.00	46.1	0.00000	****
HNP1	OBL	0.079	0.015	5.15	1620.0	0.00000	****
HNP1	ADV	0.058	0.015	3.76	1616.0	1.80E–04	**
HNP1	OBL:ADV	−0.024	0.031	−0.79	1610.9	0.43234	ns
HNP2	(intercept)	6.986	0.034	207.36	36.1	0.00000	****
HNP2	OBL	0.016	0.013	1.22	1619.4	0.22164	ns
HNP2	ADV	0.038	0.013	2.85	1616.8	0.00446	*
HNP2	OBL:ADV	0.031	0.027	1.18	1611.0	0.23846	ns
HNP3	(intercept)	6.930	0.030	231.98	38.1	0.00000	****
HNP3	OBL	0.013	0.013	0.99	1636.4	0.32397	ns
HNP3	ADV	−0.022	0.013	−1.72	1632.2	0.08591	ns
HNP3	OBL:ADV	0.037	0.026	1.42	1626.6	0.15530	ns

Table 6. *Model parameters for fixed effects of obligatoriness and number of adverbials on log reaction times at the verb, the adverbial of interest, and the heavy object, with Bonferroni corrected significance (maze task).* Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

of all models. *P*-values were computed using the Satterthwaite approximation. As before, significance was calculated using Bonferroni correction, given that seven separate analyses were conducted.

As Table 6 shows, we found a significant main effect of obligatory transitivity in reaction times on the verb and over the adverbial of interest (V, Adv1, Adv2, and Adv3). In addition, there was a significant main effect of obligatory transitivity on the onset of the heavy NP (HNP1). Obligatory transitivity did not have significant effects elsewhere. We also found a significant main effect of the presence of a delay adverbial on Adv1, Adv2, HNP1, and HNP2, but crucially there were no significant interactions with obligatory transitivity. The estimated coefficient for all effects of obligatory transitivity on the intervening adverbial is negative. In other words, reaction times are longer when the verb is obligatorily transitive. On the heavy NP, this coefficient turns positive, corresponding to shorter reaction times when the verb is obligatorily transitive.

At a conceptual level, our findings fully replicate the results reported by Staub et al. (2006). The obligatorily transitive conditions show a slowdown over the adverbial of interest relative to the optionally transitive conditions. Subsequently, the obligatorily transitive conditions show a speeding up at the beginning of the heavy NP relative to the optionally transitive conditions. There is also a main effect of the

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

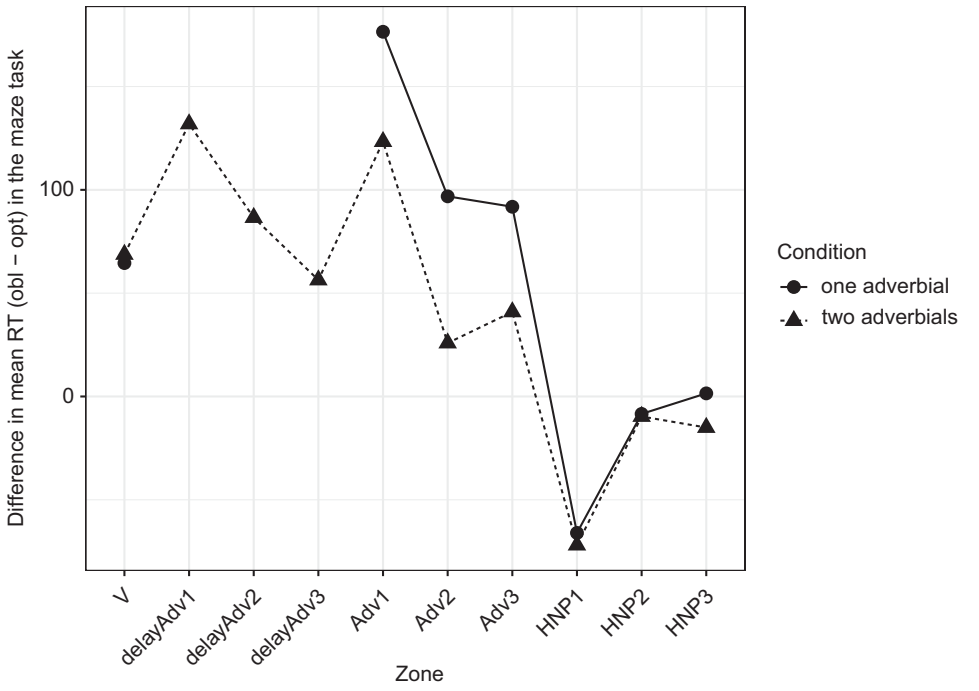


Figure 9. *Plots of raw reaction time differences in the obligatorily and optionally transitive conditions (maze task).*

number of intervening adverbials but, crucially, no significant interaction, suggesting that the slowdown in the obligatorily transitive conditions should not be attributed to transitory category-based surprisal at not finding an object where one was expected; the fact that the slowdown is also found after a delay adverbial is rather more compatible with an explanation based on the leading edge corollary. In sum, the maze task confirmed both prediction 23 and prediction 24 (as operationalized in 26).

We should note that the maze task does appear to be a more sensitive paradigm than self-paced reading: the reading time reversal on the heavy NP reported by Staub et al. (2006) did not appear as a significant effect in our self-paced reading experiment but does show up in the maze task.

The leading edge corollary also predicts that the verb being obligatorily transitive will affect reading times at the onset of the delay adverbial. Our main analysis (which reflected the logic of the experiment) excluded the delay adverbial, but we ran an additional analysis that included the delay adverbial, so that we could check this prediction. This analysis involved the comparison of ten models, with the three models for the delay adverbial missing data for the one-adverbial condition. As expected, we found a significant main effect of obligatory transitivity on delayAdv1, delayAdv2, Adv1, Adv2, and HNP1, and nowhere else. In addition, we found a significant main effect of presence of the delay adverbial on Adv1, Adv2, HNP1, and HNP2, but crucially no significant interaction with obligatory transitivity. The estimated coefficient for all effects of obligatory transitivity on the adverbials is positive, corresponding to longer reaction times. On the heavy NP, this coefficient turns negative, corresponding to the shorter reaction times in the obligatory condition.

As before, plotting the difference between mean reaction times makes the overall pattern very easy to see. Figure 9 shows a strong slowdown right after the obligatorily transitive verb. There is a second spike at the first word of the second adverbial in the two-adverbials condition. Finally, there is a reversal (lines drop below zero) at the first word of the heavy NP. All of this is as predicted by the leading edge corollary.

Note that, despite the obvious sawtooth pattern, the slowdown depicted in Figure 9 persists across the entire adverbial of interest. This is not predicted by the leading edge corollary but can plausibly be

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

attributed to maintenance costs—the instruction to find a filler needs to be kept active, and the VP containing the empty position cannot be closed. We found a similar effect in our first experiment, where, given the poor temporal resolution of self-paced reading tasks, we could not exclude an alternative account in terms of spillover effects. As noted, however, the maze task has a much better temporal resolution (see Boyce et al. 2020:12), so we can be more confident that the continued slowdown is a consequence of maintenance. That conclusion is strengthened by an observation regarding the heavy NP in our second experiment: the predicted advantage in the obligatorily transitive condition is found on the first word of that constituent and disappears immediately without any apparent spillover.

As explained in Section 2, leading edge corollary effects are more persistent than category-based surprisal (the effect of not finding an object where one is expected). The strong expectation of a DP following the obligatorily transitive verb is thwarted only once: on the word right after the verb. This is important, because it distinguishes our analysis of the processing of heavy-XP shift from a uniform alternative in which the same parsing steps are assumed to take place with optionally and obligatorily transitive verbs. In our analysis, leading edge corollary effects are predicted at the onset of every category separating the gap from the filler in the obligatorily transitive condition. On a uniform analysis, the only predicted effect is category-based surprisal. The data clearly support our proposal over the uniform account.

But this does not mean that category-based surprisal has no empirical footprint. In fact, we believe that there is a clear signal in our data of the effect of finding an adverbial where an object was strongly expected. However, this signal comes from error rates, rather than reaction times (a type of data obtained in maze tasks but not in self-paced reading tasks).

Figure 10 plots the error rates for each word position in the four conditions of the maze task. Figure 11 plots the differences in error rates between the obligatorily and optionally transitive conditions.

Figures 10 and 11 show a sharp spike in error rates and error rate differences on the first word after an obligatorily transitive verb. These spikes are clear outliers. The mean error rate for the entire experiment is roughly 0.02, with a standard deviation of about 0.013. If we calculate outliers as those error rates that are three or more standard deviations from the mean, then the only two outliers are the error rates for the first

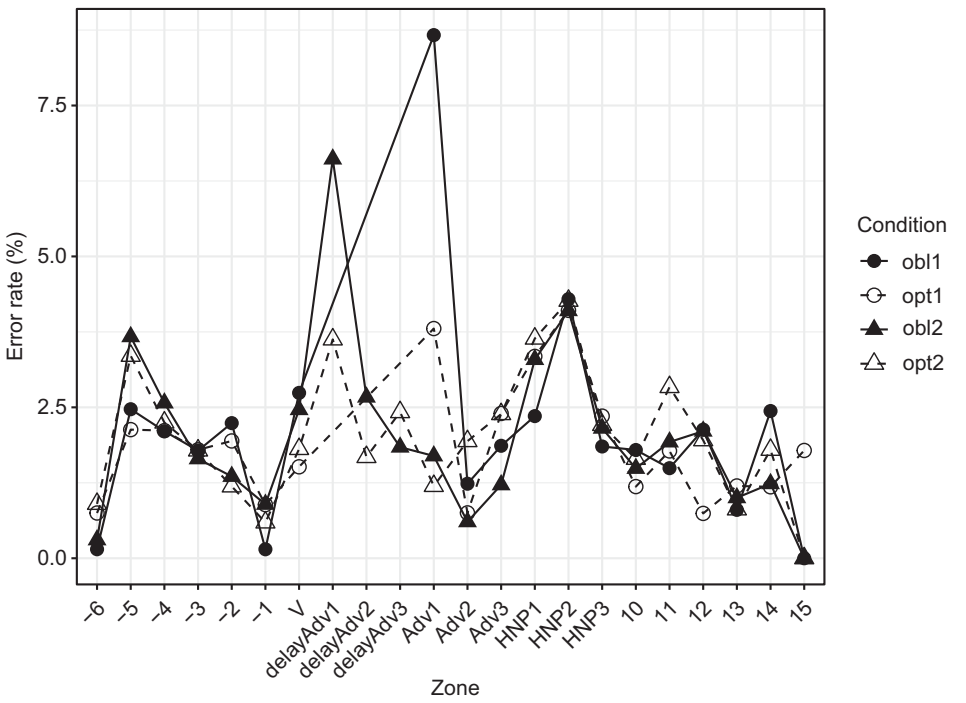


Figure 10. *Plot of error rates in the maze task. Note: obl: obligatorily transitive verb, opt: optionally transitive verb; 1: one intervening adverbial, 2: two intervening adverbials.*

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

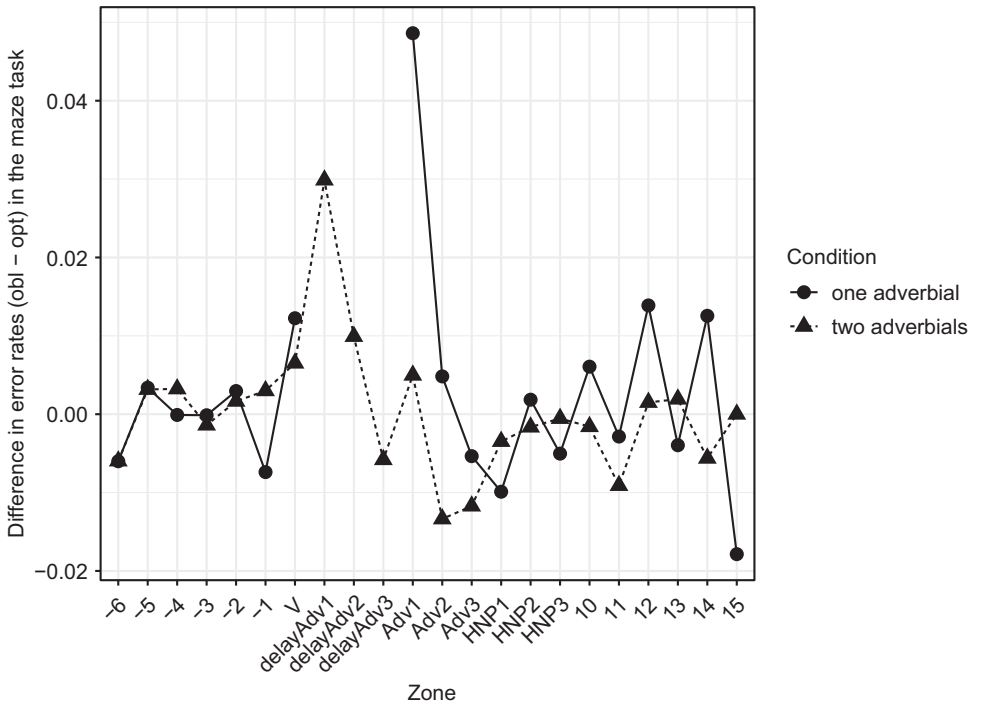


Figure 11. *Plot of differences in error rates in the maze task in the one-adverbial and two-adverbials conditions.*

word that follows the obligatorily transitive verb in the one-adverbial and two-adverbials conditions. Other ways of identifying outliers (such as absolute deviation around the median) converge on the same two data points. An obvious interpretation of the error rate spikes immediately after the verb in the obligatorily transitive conditions is that they result from the thwarted strong expectation to find a DP after the obligatorily transitive verbs, that is, from category-based surprisal. Notice that the error rates also spike after optionally transitive verbs but to a lesser degree. This, again, is as expected, given that many of the verbs in the optionally transitive conditions have a strong transitivity bias.

Category-based surprisal needs to be distinguished from what one might call **word-based surprisal**, the surprisal at finding a specific word W_n given a prior context consisting of the specific words $W_1 \ldots W_{n-1}$. Category-based surprisal abstracts away from concrete lexical items: it expresses the defeat of analytical expectations regarding the category to present itself next. Word-based surprisal is a more diffuse notion: it amalgamates various types of expectation, ranging from semantic and pragmatic expectations to expectations associated with style and register. Of course, category-based surprisal and word-based surprisal are not independent. Each word belongs to one or more categories, and hence any calculation of word-based surprisal will inextricably involve category-based surprisal in one way or another.

The weight attached to word-based surprisal in psycholinguistics has grown in recent years as a consequence of two developments. On the one hand, it has been argued to correlate well with various behavioral measures, including reaction times and reading times (see Levy 2008 and subsequent work). On the other hand, the rise of large language models has made available a reliable and practical operationalization of word-based surprisal as the negative log likelihood assigned to a word by the model given a preceding string.

The finding that there is a good fit overall between behavioral measures and word-based surprisal calculated in this way should be seen, we believe, as a consequence of the fact that parsing is an important force shaping the corpus from which large language models learn (although it is of course not the only such force). As already argued in our discussion of Wasow's (1997a,b) investigation of heavy-XP shift at

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

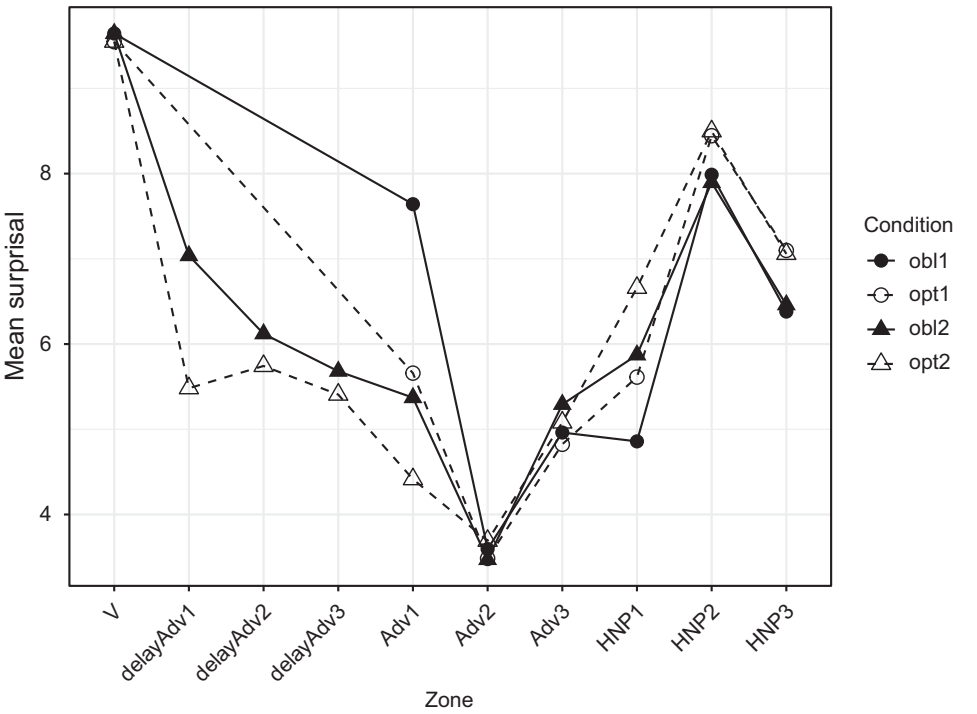


Figure 12. *Combined word-based surprisal plots.* Note*: obl: obligatorily transitive verb, opt: optionally transitive verb; 1: one intervening adverbial, 2: two intervening adverbials.*

the beginning of this section, parsing affects production and hence the profile of natural-language corpora. The corpora on which a large language model is trained in turn dictate its output. The relevance of this for (psycho)linguists is that we can use the output of large language models as a sophisticated, though indirect, test of hypotheses about the human language faculty. We therefore explored whether word-based surprisal, as calculated by a large language model, reflected our predictions about the pinch points in the parsing of heavy-XP shift. One can see this exercise as an attempt to validate Wasow's corpus findings.

We carried out this additional test by extracting the probability of each word in the stimulus sentences for our maze experiment as determined by GPT-2 (small) given the preceding word sequence.[10] We then calculated the average surprisal for each position in each condition. The results are plotted in Figure 12.

A number of features of this plot are immediately obvious. While in the error rates we observed a sharp spike at the word that immediately follows an obligatorily transitive verb (attributed to category-based surprisal), no such spike is present here. However, there is a marked difference in surprisal between the first word after an obligatorily transitive verb (solid lines) and the first word after an optionally transitive verb (corresponding dashed lines). Just like reaction times, surprisal is higher after obligatorily transitive verbs. Furthermore, the reversal of the lines that we observed in the reaction time data reappears in Figure 12. Over the first adverbial and the onset of the second adverbial, the dashed lines (representing the optionally transitive conditions) are below their counterpart solid lines (representing the obligatorily transitive conditions). However, over the onset of the heavy NP, the dashed lines are above the solid lines.

At this aggregate level, then, the output of the large language model confirms the predictions of the leading edge corollary (in conjunction with minimal attachment): the obligatorily transitive conditions come with a disadvantage at Adv1 (expressed in Figure 12 as a higher average surprisal value) and an advantage at HNP1 (expressed as a lower average surprisal value). Note that this effect is there whether a

[10] A Python script allowing us to extract these values was provided to us by Sebastian Schuster. We use GPT-2 (small) for the reasons discussed in Oh & Schuler 2023.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

delay adverb is present or not and hence cannot be attributed to the influence of category-based surprisal (which, as noted, is a necessary component of word-based surprisal). We should point out that the aggregate level is arguably the correct measure to test analytical predictions. This is because taking the average over surprisal values allows us to abstract away from aspects of word-based surprisal that are irrelevant (as they have to do with individual words in individual test sentences), thereby elevating components of surprisal values associated with the parsing process.

It is striking that the effect of obligatory transitivity in the output of the large language model—unlike in our human participants—does not persist across the entire adverbial. We hypothesized above that the continuing slowdown in humans (as depicted in Figure 9) has to do with maintenance costs. These are unlikely to affect the profile of the corpus, however, and are therefore not found in the word-based surprisal values extracted from the large language model.

To assess in more detail to what extent word-based surprisal predicts our results, we added surprisal at each word to our original reaction time data and created models of reaction time as before but with surprisal as a co-variate: log(RT) ~ OBL*Adv + surprisal + (1 | Item) + (1 | ParticipantID). As before, *p*-values were computed using the Satterthwaite approximation and significance corrected with Bonferroni correction.

The results can be summarized as follows: word-based surprisal is a highly significant predictor of reaction time at every word analyzed. The effect of obligatory transitivity remains significant in every position for which it was significant in the original analysis (i.e. Adv1, Adv2, Adv3, HNP1). There is an additional significant effect of obligatory transitivity on the verb. The effect of the number of adverbs remains significant on the second word of the heavy NP but ceases to be significant on either the first word of the adverbial or the first word of the heavy NP. The direction of these effects remains unchanged. Otherwise, nothing changes. No new interactions emerge.

The fact that the effect of obligatory transitivity can still be observed in this analysis shows that word-based surprisal does not fully account for the parsing effects we are trying to study. This need hardly surprise us. Reaction times are a direct measure of parsing effort, but the surprisal values as generated by a large language model are ultimately derived from corpora shaped by many pressures in addition to parsing effort (as explained above). Furthermore, some parsing effects (such as maintenance cost) have no plausible influence on the profile of the corpus.

The argumentation just laid out is based on the premise that parsing pressures partly shape the corpus. One might be tempted to construe the relationship between parsing and the corpus in the opposite way; that is, one could suppose that parsing difficulties arise because particular structures are rare in language use and hence rare in the corpus. We believe that this would be a serious mistake. For a start, this mode of reasoning would allow little space for the type of discrepancy between reaction times and word-based surprisal that is evident in our results. More importantly, however, it leaves theoretically meaningful aspects of the corpus as arbitrary, unexplained, and ultimately unexplainable facts (historical accounts of course lead to regress).

This point carries over to the question at the heart of this article: why are certain orders in the noun phrase not found in any language? This is a question about the profile of the typological record (which is a corpus of sorts). Its essence lies in the fact that the unattested data do not form some arbitrary subset of the logically possible orders but yield to a precise theoretical description (as shown in Section 1). Thus, the typological record presents us with a problem, and the solution to that problem cannot be found in the typological record itself.

In conclusion, we can distinguish four effects in our data. (i) There is surprisal at the lack of an object right after the verb, as seen in the spike in error rates. (ii) There is a marked slowdown in the obligatorily transitive conditions at the beginning of each adverbial, as predicted by the leading edge corollary. (iii) There is relative speeding up in the obligatorily transitive conditions at the left edge of the heavy NP, reflecting the relatively higher processing load in the optionally transitive conditions. (iv) There is a reduced but continuing slowdown over the second and third word of the adverbials in the obligatorily transitive conditions, which plausibly reflects maintenance costs associated with the incomplete rightward dependency.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

4. Universal 20

We now return to the derivation of universal 20. In Section 1, we showed that neutral orders of demonstratives, numerals, adjectives, and nouns can be captured by a combination of two hypotheses: (i) symmetric base generation under a universal hierarchy Dem > Num > A > N, and (ii) asymmetric, leftward-only, movement of the noun (or constituents containing the noun). As we pointed out, this account raises an important conceptual issue: if phrase structure is symmetric, why should movement of the noun be asymmetric? We argue that the answer to that question should start with the assumption that movement of the noun can in principle be rightward as well as leftward. However, grammars that employ rightward noun movement systematically cause parsing difficulties that are absent in grammars in which the noun remains in situ or moves leftward. These parsing difficulties make grammars with rightward noun movement harder to acquire, and this, given enough time, leads to their disappearance from the typological record.

In what follows, we first identify the parsing difficulties inherent in rightward movement of a head (Section 4.1). This discussion is informed by the heavy-XP shift experiments reported in Section 3. Then we consider the relationship between parsing difficulties and typological frequency, basing ourselves on earlier work by Kirby (1999) and Hawkins (2004, 2009) (Section 4.2). Finally, we explore how the leading edge corollary fits in with other factors known to affect the typological frequencies of noun phrase orders, showing that four ranked constraints (including the leading edge corollary) are sufficient to provide an adequate model of the known data (Section 4.3).

4.1. Parsing head movement

It is obvious that structures that can be built without movement will require less processing effort than otherwise comparable structures that do involve movement, simply because of the costs of the movement routine. If parsing effort is reflected in typological frequency, this may explain why the structures in 29a,b are more frequent than those in 29c,d (parentheses are used for nodes whose internal order cannot be determined; numbers in parentheses represent languages and genera according to Dryer 2018 and Cinque 2025, respectively).

(29) a. [Dem [Num [A N]]] (attested (113, 57) (457, 116))
b. [[[N A] Num] Dem] (attested (182, 85) (657, 141))
c. N [Dem [Num (A ~~N~~)]] (attested (8, 6) (93, 27))
d. [[(~~N~~ A) Num] Dem] N (unattested (0,0) (0, 0))

Why are leftward and rightward movement of the noun not on a par? We have argued in Section 2 that, beyond the movement routine, additional processing costs may arise because of the filled-gap effect (in the case of leftward movement) and the missing-filler effect (in the case of rightward movement). We now argue that while rightward movement of the noun systematically gives rise to missing-filler effects, leftward noun movement never gives rise to filled-gap effects.

As stated in the leading edge corollary (which we derived from the combination of minimal attachment and the minimal chain principle), rightward movement of obligatory material will trigger a costly search for a filler (costly because a gap cannot be stored in the way that the content of a filler can be).

As an example, consider how the structure in 29d would have to be parsed. When the adjective presents itself, the parser must treat it either as a predicate or as an attributive modifier.[11] Since we are dealing with a noun phrase, the second option is the correct one (the first would give rise to a garden-path effect). But if the AP is an attributive modifier, it must be accompanied by a nominal category. Hence, the parser is forced to posit an empty position, as in 30a. The parser will then attempt to integrate the next element it

[11] The distinction is possibly encoded grammatically, with adjectives that function as attributive modifiers derived from predicative adjectives by an operator dubbed JOIN (Chierchia & Turner 1988, Baker 2003, and Truswell 2004). Both Baker and Truswell argue that JOIN derives a modifier that selects a nominal category to account for the fact that APs can act as modifiers of nouns, but not as modifiers of other adjectives (cf. **The bus is big red*).

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

encounters (the numeral) into, or to the left of, this empty position. This attempt must fail (resulting in a first missing-filler effect), because the numeral is not a noun and the structure *[A [Num N]] violates the hierarchy in 4. The only remaining option is to remove the empty position from the leading edge of the representation, as in 30b. This triggers a search for the missing filler, implemented by subjecting each subsequent constituent to the movement regime. Hence, an unsuccessful attempt is made to insert a copy of the demonstrative in the empty position, as in 30c (resulting in a second missing-filler effect). Only when the noun is attached and hypothesized to have moved will a trace be inserted in the empty position and the movement dependency resolved, as in 30d,e.

(30) a. (A *e*)
b. [(A *e*) Num]
c. [[(A *e*) Num] Dem]
d. [[[(A *e*) Num] Dem] N]
e. [[[(A ~~N~~) Num] Dem] N]

The logic of this example carries over to all structures in which the noun moves rightward across other material. This is because all orders that require rightward movement include one of three substrings: A-Num-N, A-Dem-N, or Num-Dem-N. We have just shown that the orders A-Num-N and A-Dem-N trigger a missing-filler effect (because Num and Dem cannot fill the empty nominal position triggered by the adjective's selectional properties). The problem with Num-Dem-N is much the same. Like attributive adjectives, numerals must attach to a nominal category. Hence, when the parser encounters the numeral, it will build a representation (Num *e*). However, the demonstrative cannot be integrated into, or to the left of, the empty position, as it is not a noun and the structure [Num [Dem N]] violates the hierarchy in 4. A missing-filler effect is hence inescapable.

How does this state of affairs compare with the parsing of leftward movement of the noun? At first blush, one might think that leftward noun movement should give rise to filled-gap effects. Consider the representation in 29c and suppose that the parser is in state 31a. By the minimal chain principle, an attempt should now be made to insert the trace of the noun into the empty position, as in 31b. However, the next element in the input string is the numeral, which (one may think) will require that the trace is withdrawn and the noun reactivated as a filler, as in 31c.

(31) a. N [Dem *e*]
b. N [Dem ~~N~~]
c. N [Dem [Num *e*]]

The sequence of steps in 31, however, is based on the unnecessary assumption that the trace of a moved head can only be inserted once the parser has reached the linear position corresponding to the launching site of movement. In fact, standard conditions on the movement of heads—such as the head movement constraint (Travis 1984) and chain uniformity (Chomsky 1995)—entail that the parser can uniquely identify the structural position of the trace as soon as it has identified a head as a filler. In the case at hand, once it has buffered a copy of the noun, it can immediately offload the trace, as it knows that the trace must be inserted as the unique lexical head of the extended nominal projection. Thus, step 32a is immediately followed by step 32b. When the parser subsequently encounters the demonstrative, the numeral, and the adjective, these elements can be attached in positions preceding the lower copy of the noun, as in steps 32c–e. At no point is it necessary to withdraw the trace.[12]

[12] Note that in standard examples of the filled-gap effect (such as Stowe's (1986) *My brother wanted to know who Ruth will bring **us** home to at Christmas*, where the effect is found at the boldfaced pronoun), the trace erroneously inserted cannot remain in place, but must be withdrawn and inserted elsewhere.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

(32) a. N

b. N [~~N~~]

c. N [Dem ~~N~~]

d. N [Dem [Num ~~N~~]]

e. N [Dem [Num (A ~~N~~)]]

The rightward shift of the trace in 32 is possible because the trace has no phonological content. A similar rightward shift of an overt head would not be permitted, as the parser must respect the order of elements in the input string.

There is experimental evidence that supports this view of the processing of displaced heads. This evidence does not come from the noun phrase, but from **verb second** in German. In German, the verb is base-generated in clause-final position, yielding SOV as the neutral order in finite embedded clauses. Verb second is the result of two movement operations that take place in declarative main clauses (Williams 1974, Koster 1975, Emonds 1976, den Besten 1983, and much subsequent work). The operation of interest to us here shifts the finite verb to the left periphery of the clause. An additional phrasal movement places a constituent (frequently the subject) in a position preceding the verb's landing site, with the consequence that the verb ends up in second position. Thus, main-clause word order is SVO if the subject is fronted.

On our assumptions, the SOV order of embedded clauses can be parsed straightforwardly, as in 33 (note that C must be followed by a verbal projection, which triggers the postulation of an empty position). Importantly, the sequence of steps in 33 implies that thematic integration of the arguments must wait until the verb is inserted in *e* (steps 33d,e; we indicate theta-role satisfaction by adding a hash mark to the theta-role in question).

(33) a. C [*e*]

b. C [DP_S *e*]

c. C [DP_S [DP_O *e*]]

d. C [DP_S [DP_O V⟨θ θ⟩]]

e. C [DP_S [DP_O V⟨$\underline{\theta}_{\#}$ $\theta_{\#}$⟩]]

Suppose—contrary to our assumptions—that in main clauses the trace of the verb was inserted only when the parser reaches the point in the string where the verb would be in embedded clauses (that is, following the object). Then, the sequence of parse steps would presumably be as in 34 (on the assumption that a clausal structure is predicted early on, with an empty position to host the verb's trace). Given that there is overwhelming syntactic evidence that the verb performs its thematic functions in its base position (or at any rate, in a position much lower in the clausal architecture than the landing site of verb second), the most straightforward prediction of a proposal along these lines would be that even though the identity of the verb is known early on, this information cannot be used by the parser while analyzing the material that separates the landing and launching sites of verb movement.

(34) a. DP_S [*e*]

b. DP_S V [*e*]

c. DP_S V [~~DP_S~~ [*e*]]

d. DP_S V [~~DP_S~~ [DP_O *e*]]

e. DP_S V [~~DP_S~~ [DP_O ~~V~~⟨θ θ⟩]]

f. DP_S V [~~DP_S~~ [DP_O ~~V~~⟨$\underline{\theta}_{\#}$ $\theta_{\#}$⟩]]

It is, of course, awkward that the parser has information available but cannot use it. Indeed, Freitag (2021) argues that verb movement may yield early benefits (despite the fact that the verb performs its functions in

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

its base position) if the verb's trace is inserted immediately after the verb has been identified as a filler. As Freitag puts it (2021:273): 'the parser places the finite verb in clause-final position very early and inserts incoming phrases before the verb'. On this view, an SVO main clause will be parsed as in 35. This parse permits thematic integration of the subject as soon as the verb presents itself (steps 35b–e) and thematic integration of the object as soon as it is identified (steps 35f,g).

(35)
a. $\boxed{\text{DP}_S}$ [*e*]
b. $\boxed{\text{DP}_S}$ $\boxed{\text{V}}$ [*e*]
c. $\boxed{\text{DP}_S}$ V [~~V~~⟨θ̲ θ⟩]
d. DP_S V [~~DP~~$_S$ ~~V~~⟨θ̲ θ⟩]
e. DP_S V [~~DP~~$_S$ ~~V~~⟨θ̲$_{\#}$ θ⟩]
f. DP_S V [~~DP~~$_S$ [DP_O ~~V~~⟨θ̲$_{\#}$ θ⟩]]
g. DP_S V [~~DP~~$_S$ [DP_O ~~V~~⟨θ̲$_{\#}$ θ$_{\#}$⟩]]

As Freitag remarks, several studies support early insertion of the verb's trace; we briefly review two here. Friederici and Frisch (2000) report on two event-related potential (ERP) experiments that explore the processing of German main and embedded clauses, respectively. Both experiments compare well-formed sentences with sentences that feature a mismatch between object and verb (in semantics, in case, or in the number of arguments versus the arity of the verb). In embedded clauses, mismatches triggered neural responses at the clause-final verb, as expected. These varied per condition and consisted of early negativity (N400 or left-lateralized negativity), followed by positivity (P600), a sequence interpreted as indicative of detection of the mismatch and subsequent repair. In verb-second clauses, neural responses came earlier, at the object. They again varied according to the type of violation and differed from those in the earlier experiment. Semantic mismatches and number-of-argument mismatches triggered an N400 response, taken to be indicative of difficulties with integration of the argument given the prior verb. Number-of-argument mismatches additionally triggered later positivity (presumably signifying an attempt at repair). Finally, case mismatches did not trigger early negativity, but resulted in clear positivity in the 600–800 ms range (Freitag (2021:243) suggests that the lack of early negativity may be because of grammatical continuations that exist in the case-mismatch condition).

Friederici and Frisch (2000) conclude that verb second affects the way the object is parsed. In main clauses (where the verb precedes the object) thematic integration takes place as soon as the object is identified, while in embedded clauses (where the object precedes the verb) thematic integration must wait till the verb presents itself. If the verb's trace is inserted early, as in 35, this pattern can be reconciled straightforwardly with the uncontested syntactic fact that the verb selects the object from its base position.

Konieczny et al. (1997) ran three eye-tracking experiments that explore attachment ambiguities in main and embedded clauses. The items all had the structure in 36. The boldfaced XP was a PP or a DP with ambiguous case marking that in principle could be attached either in the preceding object DP or to the verbal projection. The XP and verb were chosen to favor one or the other attachment. The experiments warranted three conclusions. (i) In embedded clauses, XPs that favor attachment in the DP are easier to parse than XPs that favor attachment to the verbal projection. (ii) In verb-second clauses, the opposite pattern was found: XPs that favor attachment in the DP are **harder** to parse than XPs that favor attachment in the verbal projection. Thus, when the verb is in its base position, there is a preference for attachment in the object DP, but when the verb undergoes Verb second, there is a preference for attachment to the verbal projection. (iii) Parsing of the XP took longer in verb-second clauses.

(36) a. C [DP_S [DP_O **XP** V]]
b. DP_S V [DP_O **XP** ~~V~~]]

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

The attachment preferences can be understood if in general attachment to the last active category is the default (this preference subsumes various late-closure effects). In embedded clauses, the last active category before the XP is parsed is the DP object (step 37b), so attachment in the DP is preferred, as indicated by the arrow in 37c.

(37) a. C [DP_S *e*]
b. C [DP_S [DP_O *e*]]
c. C [DP_S [DP_O $\underset{\leftarrow}{\mathbf{XP}}$ *e*]]

In verb-second clauses, however, the DP is immediately thematically integrated as the object, which means that the last active category when the XP presents itself is the verb's trace, where the internal theta role has just been marked as satisfied (step 38c). This favors attachment to the verbal projection.

(38) a. DP_S V [~~DP~~$_S$ ~~V~~⟨$\underline{\theta}_{\#}$ θ⟩]
b. DP_S V [~~DP~~$_S$ [DP_O ~~V~~⟨$\underline{\theta}_{\#}$ θ⟩]]
c. DP_S V [$\underline{\text{DP}}_S$ [DP_O $\underline{\text{V}}$⟨$\underline{\theta}_{\#}$ $\theta_{\#}$⟩]]
d. DP_S V [~~DP~~$_S$ $\underset{\rightarrow}{\mathbf{XP}}$ ~~V~~⟨$\underline{\theta}_{\#}$ $\theta_{\#}$⟩]]

As Freitag (2021:164) suggests, the increased reading times of XP in verb-second clauses may be attributed to the fact that such clauses are in principle complete after the DP has been integrated as object, while embedded clauses are not (the verb is still missing). Hence, wrap-up processes may have started that must be interrupted when the PP is encountered.

In conclusion, the experiments on the processing of heavy-XP shift discussed in Section 3 suggest that rightward head movement will systematically give rise to missing-filler effects. By contrast, the assumption that traces of leftward head movement are inserted early (for which there is independent motivation in existing experimental work) implies that such movement will not give rise to filled-gap effects. Thus, when we look back at the typology of unmarked orders in the noun phrase, parsing effort allows us to make two distinctions: a first distinction between structures with and without movement, and a second distinction between structures with leftward and structures with rightward movement. We explore the typological consequences of this in the next two subsections.

4.2. Synchronic and diachronic consequences of the leading edge corollary

We have argued that rightward movement of obligatory categories systematically gives rise to missing-filler effects (as expressed by the leading edge corollary). Confirmation of this hypothesis comes from the findings reported by Staub et al. (2006) regarding heavy-XP shift and from the replication and extension of those findings in Section 3. We have suggested that, as a consequence, heavy-XP shift of an obligatory object is harder to parse than heavy-XP shift of an optional object (even though the latter shows a slowdown over the onset of the heavy NP). The results reported in Table 6 directly support this suggestion.[13] The model coefficients in a linear mixed-effects model correspond to relative (rather than standardized) effect sizes. The model coefficient for the relative slowdown at the first word of the adverbial in the obligatory conditions (0.151) is roughly twice as large as the relative slowdown on the first word of the heavy NP in the optional conditions (0.08). Moreover, the large effect persists on the following words, while the small effect vanishes immediately. This justifies the notion that the parsing

[13] Indirect evidence comes from the fact that in the eye-tracking experiments by Staub et al. (2006), the difficulties in the obligatorily transitive condition affected all measures, while the slowdown in the optionally transitive condition affected only some measures. In addition, our self-paced reading experiment replicated the effects of obligatory transitivity, but not those of optional transitivity (presumably because self-paced reading is not as sensitive a measure as a maze task or eye tracking).

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

difficulties associated with heavy-XP shift of obligatory material are more severe than those associated with heavy-XP shift of optional material.

In the extended nominal projection, rightward movement of the noun or a constituent containing it creates parsing difficulties of much the same nature. As the noun is obligatory in the extended nominal projection, gap-filler order unavoidably leads to missing-filler effects.

At this point, we face an urgent question. Why should the very same parsing difficulties give rise to unattested orders in the noun phrase and to no more than reduced frequency in the case of heavy-XP shift of obligatory objects? The answer to this question cannot be found in the mechanisms of parsing. As far as the parser is concerned, both movements are disfavored, though grammatical in principle. We need to look for the distinguishing factor elsewhere.

There is, in fact, a very obvious distinguishing factor. Heavy-XP shift has instances that do not cause missing-filler effects, namely, when the verb only optionally selects an object. By contrast, missing-filler effects will systematically be found with rightward movement of a constituent containing the noun in the extended nominal projection (as long as it is not string-vacuous).[14] Our suggestion is that this difference is crucial for the acquisition of heavy-XP shift versus the acquisition of rightward movement in the noun phrase.[15]

A substantial body of work suggests that difficulty in parsing corresponds to typological frequency. As Hawkins (2004:3) puts it: 'Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments' (see also Hawkins 2009). Of course, correlation is not causation, and in order to understand the link between parsing and grammar explored by Hawkins, we need a mechanism that links parsing difficulty in individuals to typological frequency. The basis for such a mechanism must lie in the fact that typological frequency is tied to language change and that language change in turn is tied to language acquisition. Therefore, properties of structures that hinder acquisition will suppress their typological frequency. We have suggested above that speakers learn to avoid (to some degree) production of difficult-to-parse structures. This affects the input to the language acquisition process.

Furthermore, in order to acquire a grammar, the child must parse the utterances in the input to come to representations that can be used for the acquisition of grammar (Valian 1990, Fodor 1998, Omaki 2010, and Gagliardi 2012): the intake. This is schematically represented in 39, where LAD stands for **language acquisition device**, the innate mental system that permits humans to acquire a grammar.

[14] String-vacuous rightward movement of objects (or its equivalent) is frequently invoked in the analysis of wide object scope in English (see Huang 1982, Mayr & Spector 2011, and Steedman 2011). Such movement still creates a chain where none is strictly necessary, which might help explain why surface scope is preferable to inverse scope.

[15] Martina Wiltschko (p.c.) raises the question of whether the differences we observed in the obligatorily and optionally transitive conditions of our experiments might arise only in the written modality (which is how our stimuli, as well as those used by Staub et al. (2006), were presented). The question comes in two parts: (i) is there a clear intonational cue for a gap right after the verb in spoken examples of heavy-XP shift?, and (ii) is there a clear cue for a shifted object at the end of the material shifted across? The latter would facilitate integration of the heavy NP with both obligatorily and optionally transitive verbs, something that would be neutral as regards our proposal. The former, however, would provide evidence for a gap even with optionally transitive verbs and would hence undermine the suggestion in the main text that heavy-XP shift is relatively easy to learn because it frequently does not trigger missing-filler effects.

Göbbel (2024) reports on a prosodic study of heavy-XP shift. He includes pitch tracks of minimally paired examples with and without heavy-XP shift which suggest that there is no prosodic signal of the operation on the verb (see Göbbel 2024:65, figure 5). Indeed, he confirms (in personal communication) that '[t]he gap is not signalled at all'. Therefore, in the spoken modality, too, only obligatorily transitive verbs should trigger missing-filler effects.

Göbbel (p.c.) does note that there are differences in prosody at the end of the material crossed (that is, right before the heavy NP). Instead of a sentence-final L-L% tone, other intonational patterns can be observed, with an L-H% continuation rise a prominent option when the heavy NP forms a separate intonation phrase. These contrasting intonational patterns may be useful in processing, as they signal that there is still material to come (although they probably do not signal that this material will form a heavy NP). However this may be, prosodic marking right before the heavy NP comes too late to trigger missing-filler effects with optionally transitive verbs.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

(39) input → PARSER → intake → LAD → grammar

Kirby (1999) points out that, given that the parser mediates between the input and the language acquisition device, structures that are hard to parse will have a reduced frequency in the intake compared to their frequency in the input. He shows that this has consequences for typology: over many generations, a reduction of evidence in the intake will eliminate structures that are hard to parse, as long as the system is seeded with some initial variation and learners replicate the frequencies from the intake. (Given that structures may be introduced as well, such as under language contact or by random innovation, the filtering effect of the parser must presumably reach a certain threshold in order to eliminate a structure from the typological record; see Section 4.3 for discussion.)

Kirby's account of typological distribution as the result of language change mediated by language learning straightforwardly militates against the relevant rightward movements in the extended nominal projection; they systematically cause missing-filler effects and will therefore be suppressed in the child's intake. Heavy-XP shift of optional objects, by contrast, does not cause the same kind of parsing difficulties and therefore contributes to a presence of relevant inputs in the intake that is—we propose—sufficiently frequent for a general process of rightward shift to be acquired. This general process can subsequently apply to optional and obligatory objects alike.

The applicability of the process beyond the triggering data is a direct consequence of core assumptions about the grammar of movement. From the inception of transformational grammar, researchers have attempted to eliminate contextual constraints that restrict the application of movement transformations, culminating in the Move-α framework (see Chomsky 1981, 1986 and Lasnik & Saito 1992). In this framework, there is only one general transformation that can apply when permitted by the principles of grammar and language-specific constraints; that is, there can be no construction-specific conditions on movement. In our assessment, this program of research has been very successful. Few, if any, of the displacements in language require a description that mentions a specific context. (This is remarkable in view of the fact that many other kinds of rules required for adequate grammatical description do need to have such contextual constraints; examples are rules of impoverishment and vocabulary insertion used for morphological description in frameworks based on late lexical insertion.)

If the Move-α program is on the right track, the acquisition of heavy-XP shift cannot be the acquisition of a rule, but must be the acquisition of the knowledge that a certain landing site is available and that traces are allowed in object position. It is therefore not possible to define a process of rightward shift that can be applied to optional objects only. Any movement process that can apply to optional objects can automatically apply to obligatory objects as well. A distinction could only be made by referring to the thematic make-up of (a variant of) the verb, which is exactly the kind of conditioning that the Move-α program claims does not exist.[16]

A second factor that is likely to be relevant has to do with the fact that certain movements come with a payoff. This is certainly true of heavy-XP shift, which introduces a division in the clause that can be

[16] Jan-Wouter Zwart (p.c.) reminds us that heavy-XP shift in English, unlike leftward movement, does not strand prepositions. Given that prepositions typically take obligatory complements, heavy-XP shift from PPs looks like a candidate for a rule that systematically causes parsing difficulties and that will, therefore, be lost over time. The process that shifts PPs as a whole is not subject to attrition in the same way, since PPs may or may not be obligatory. Drummond et al. (2010) note that an account explicitly stipulating conditions on heavy shift from PPs is not compatible with the Move-α program. However, the prohibition on preposition stranding under heavy-XP shift can be implemented, following Bresnan 1976, by extending the A-over-A condition (or its successor in minimalism) to the operation. This would require that the grammar identifies a triggering feature shared by heavy PPs and heavy DPs. The nature of this feature requires further research, but there can be little doubt that PPs can indeed undergo heavy-XP shift (see Neeleman & Payne 2020:480–82 for evidence). We find the criticism by Drummond et al. (2010:689–90) of Bresnan's idea beside the point: in all contexts where heavy shift of the complement of P would be licensed, heavy shift of the PP is also licensed. The contexts brought up by Drummond et al. are ones where heavy-XP shift would not be licensed but PP shift is; these are irrelevant to the type of competition resolved by the A-over-A condition. (For further discussion and a very different grammatical approach, see Drummond et al.'s paper.)

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

exploited for information-structural and/or prosodic purposes (see Göbbel 2024 for recent discussion and references). By contrast, movement of lexical heads, at least when used to derive neutral word orders, does not have any interpretive effects and has no obvious prosodic advantages. Therefore, it is not protected by functional considerations in the way that heavy-XP shift is. (See fn. 4 on the option that some movements of heads do have information-structural effects and could be rightward.)

Consequently, there is pressure to abandon grammars in which neutral word order in the noun phrase is derived by rightward movement, but no similar pressure to abandon grammars that allow heavy-XP shift of obligatory objects. This makes the first class of grammars natural victims of language change, implying that given enough time they will become extinct.

4.3. The frequency of universal 20 orders

The account of universal 20 suggested above is based on the notion that the leading edge corollary radically reduces the typological frequency of structures derived by rightward movement of the noun (or a constituent containing the noun). This is a departure from the accounts in Cinque 2005 and Abels & Neeleman 2012, where the ban on rightward noun movement was taken to be grammatical and hence inviolable.

Of course, the leading edge corollary is not the only factor that increases parsing effort and hence reduces typological frequency. We suggested at the onset of Section 4.1 that because the movement routine is costly, orders derived by leftward movement are at a disadvantage compared to orders that can be generated without movement. This may explain the contrast in typological frequency between N-A-Num-Dem and Dem-Num-A-N, on the one hand, and N-Dem-Num-A, on the other (see 29). Thus, while the leading edge corollary reduces the typological frequency of rightward movement orders to zero (or close to it), the cost of the movement regime reduces the typological frequency of leftward movement orders, but to a lesser extent, given that it is a inherently weaker factor.

We may then account for typological frequency through a system of ranked constraints, as in 40 (where LEC stands for leading edge corollary and **no-movement** militates against movement). This system must be combined with what one might call 'linguistic entropy': the introduction of new word orders as a consequence of language contact or random innovation. Without such a counteracting fluctuation in word order, all languages would eventually converge on the least marked order. The relative strength of the factors suppressing frequency then determines whether entropy will give rise to a given order often, rarely, or never.

(40) LEC > no-movement
↑
cut-off point of attestation

The model in 40 is clearly too simplistic. The remainder of this section aims to improve on it by exploring which other factors determine typological frequency in the domain of universal 20.

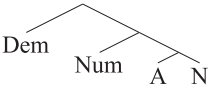
The typological distribution of universal 20 orders follows a power law, with some orders much more frequent than others. It is possible to give a coherent account of the data based on purely formal factors. A good fit with the known distribution can be achieved if in addition to the factors mentioned in 40, there is (i) a preference for harmonic orders and (ii) a preference for N-A order.

Harmony is more likely associated with biases in language acquisition than with parsing preferences (harmonic linearization allows simpler grammars than disharmonic linearization). However that may be, its empirical effects are well motivated (see Cinque 2005, Culbertson et al. 2012, Culbertson & Newport 2015, Enochson & Culbertson 2015, and Dryer 2018). In order to incorporate it into our model, an explicit definition of harmony as it applies to extended projections is necessary. We give this definition in 41.

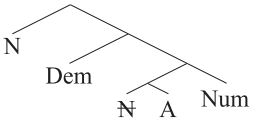
(41) Harmony: An extended projection is harmonic if nonprojecting categories are either uniformly located on right branches or uniformly located on left branches.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

Assuming that Dem, Num, and A are nonprojecting and (following Georgi & Müller 2010) that moved nouns and moved noun-adjective units reproject in their landing site, the definition in 41 characterizes the structure in 42a as harmonic and the structure in 42b as disharmonic. Any structure in which the noun moves is also disharmonic (unless the movement is string-vacuous). For example, in 42c, the noun's sister is a nonprojecting category on a right branch. However, the demonstrative, across which the noun has moved, is a nonprojecting category on a left branch. The same applies to structures in which the noun-adjective unit moves, because by assumption these too reproject.

(42) a. 

b. 

c. 

Evidence for the preference for N-A orders comes from an artificial language-learning experiment reported in Culbertson et al. 2012. Culbertson et al. discuss two explanations, one of which can be discarded, as it assumes (incorrectly in our view) that the adjective is a head in the extended nominal projection. The other explanation is based on Kamp and Partee's (1995) (Semantic-)Head Primacy Principle, which states that nouns are interpreted before attributive adjectives, regardless of word order. This sequence of interpretive steps is forced by the fact that the interpretation of many adjectives depends on the noun: a big butterfly, for instance, does not have the same size as a big elephant, but rather describes a butterfly that is big for a butterfly. Similarly, the red in *a red brick* does not describe the same color as the red in *a red traffic light.* Thus, if the noun comes first, the adjective can be interpreted immediately (even if the noun has moved; see Section 4.1). If the adjective comes first, however, it has to be buffered for interpretation until after the noun is interpreted, so the latter order is dispreferred as less efficient in parsing. This parsing advantage does not hold for all adjectives. However, grammars with a single order for noun-adjective combinations are simpler and therefore preferred in acquisition. Consequently, the preference for N-A order will generalize to all adjectives. (A preference for N-A order, but based on a generalization from relative clauses to adjectives, has also been invoked in Dryer's (2018) account of the frequency of universal 20 orders.)

We can capture the ranking of universal 20 orders if the model in 40 is extended as in 43.

(43) LEC > harmony > N-before-A > no-movement
↑
cut-off point of attestation

The predictions for the fourteen attested orders are as follows. (i) Harmonic orders outrank disharmonic orders. (ii) Within the class of harmonic orders, N-A-Num-Dem is more frequent than Dem-Num-A-N, and within the class of disharmonic orders, those in which the adjective follows the noun are more frequent than those in which the adjective precedes the noun. (iii) Within each of the two disharmonic classes, nonmovement orders outrank movement orders. Thus, the ranking predicted by 43 is as given in Table 7.

Note that the hierarchy of preferences in 43 is crucial in generating the ranking in Table 7. If harmony, N-before-A, and no-movement were all equal in status, the ranking would presumably be a function of the number of violations. This would lead to a more coarse-grained picture in which not six, but four groups of orders are distinguished: those with zero violations (N-A-Num-Dem), those with one violation (Dem-Num-A-N, Dem-N-A-Num, Dem-Num-N-A, and Num-N-A-Dem), those with two violations (N-A-Dem-Num, Dem-N-Num-A, N-Num-A-Dem, N-Dem-A-Num, N-Dem-Num-A, Dem-A-N-Num, Num-A-N-Dem, and A-N-Num-Dem), and those with three violations (A-N-Dem-Num). If the hierarchy in 43 was altered, the predicted ranking would shift accordingly. For example, if the hierarchy were N-before-A > harmony > no-movement, Dem-Num-A-N would not be ranked second but tenth.

How well does the predicted ranking fit the data? There are several measures of typological frequency in the universal 20 domain. Dryer (2018) favors a notion of adjusted frequency, which takes account of both genealogical and geographical distance. Alternatively, one could take frequency of language type or

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

	Harmony	N-before-A	No-movement	Predicted ranking
N-A-Num-Dem				1
Dem-Num-A-N		*		2
Dem-N-A-Num	*			3–5
Dem-Num-N-A	*			3–5
Num-N-A-Dem	*			3–5
N-A-Dem-Num	*		*	6–10
Dem-N-Num-A	*		*	6–10
N-Num-A-Dem	*		*	6–10
N-Dem-A-Num	*		*	6–10
N-Dem-Num-A	*		*	6–10
Dem-A-N-Num	*	*		11–13
Num-A-N-Dem	*	*		11–13
A-N-Num-Dem	*	*		11–13
A-N-Dem-Num	*	*	*	14

Table 7. *Ranking of universal 20 orders as predicted by 43.*

	AF(D)	LF(D)	GF(D)	LF(C)	GF(C)	Spr	AR	PR
N-A-Num-Dem	1 (44.17)	1 (182)	1 (85)	1 (652)	1 (141)	0	1 (1)	1
Dem-Num-A-N	2 (35.56)	2 (113)	2 (57)	2 (457)	2 (116)	0	2 (2)	2
Dem-N-A-Num	3 (29.95)	4 (53)	3 (40)	4 (214)	3 (90)	1	3 (3.4)	3–5
Dem-Num-N-A	4 (22.12)	5 (40)	4 (32)	5 (190)	4 (77)	1	4–5 (4.4)	3–5
Num-N-A-Dem	**6 (14.54)**	3 (67)	5 (27)	3 (253)	5 (50)	3	4–5 (4.4)	3–5
N-A-Dem-Num	**5 (14.80)**	6 (36)	6 (19)	6 (105)	6–7 (35)	2	6 (5.9)	6–10
Dem-N-Num-A	7 (9.75)	8 (12)	8 (10)	11 (49)	8 (29)	4	7–8 (8.4)	6–10
N-Num-A-Dem	8–9 (9.00)	10 (11)	9 (9)	8 (73)	6–7 (35)	4	7–8 (8.4)	6–10
N-Dem-A-Num	8–9 (9.00)	7 (13)	7 (11)	**13 (33)**	**12 (17)**	6	9 (9.5)	6–10
N-Dem-Num-A	10 (5.67)	**11–12 (8)**	**11 (6)**	7 (93)	10 (27)	5	**11 (9.9)**	6–10
Dem-A-N-Num	11 (5.34)	**9 (12)**	**10 (7)**	10 (52)	**9 (28)**	2	**10 (9.8)**	11–13
Num-A-N-Dem	12 (4.00)	11–12 (8)	12 (5)	**9 (55)**	11 (21)	3	12 (11.1)	11–13
A-N-Num-Dem	13 (3.00)	13–14 (5)	13–14 (3)	12 (34)	13 (13)	2	13 (13)	11–13
A-N-Dem-Num	14 (2.50)	13–14 (5)	13–14 (3)	14 (21)	14 (9)	1	14 (13.8)	14

Table 8. *A comparison between the predicted ranking and the actual ranking of the fourteen universal 20 orders compatible with the leading edge corollary.* Note*: AF(D): adjusted frequency (Dryer 2018), LF(D): language frequency (Dryer 2018), GF(D): genus frequency (Dryer 2018), language frequency (Cinque 2025), GF(C): genus frequency (Cinque 2025), Spr: spread of rankings (highest minus lowest ranking), AR: ranking based on average ranking over the five measures, PR: predicted ranking according to Table 7.*

frequency of genera containing languages of the relevant type as a measure. Rankings based on these measures are given in Table 8 (there are two counts of languages and two counts of genera, taken from Dryer 2018 and Cinque 2025, respectively). As the table shows, the five measures do not always agree. In fact, the spread of rankings across the five measures goes up to five or six ranks for some orders. This variation suggests to us that there is still a great deal of uncertainty about the actual typological frequency of the orders in this domain.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

Nonetheless, we can observe that the predicted ranking is a good match with each of the five rankings in Table 8. Each contains only one pair of orders whose ranking is not as predicted (these mismatches appear in bold). The same is true of the average ranking over the five measures. Importantly, the mismatches between the predicted rankings and the rankings produced by the five measures do not converge on the same pair of orders (which would be indicative of a robust disagreement between predictions and data).

We conclude that typological frequency in the universal 20 domain can be captured through a hierarchy of formal preferences. Among these, we find preferences that are associated with parsing costs (in particular, the leading edge corollary, the N-A preference, and the preference for nonmovement orders), alongside harmony, which may reflect a bias in language acquisition.

We can only partly understand why the relations between the constraints in 43 should be as given. The N-A preference and the preference for nonmovement orders emerge because an increase in parsing effort will lead to a reduced presence of relevant orders in the intake. However, this increase in effort does not match the difficulties associated with missing-filler effects. Hence, it is no surprise that the leading edge corollary outweighs N-before-A and no-movement. We have no insight into the relative positions of the latter two factors, however, nor do we understand why harmony should be placed where it is, given that there is no a priori expectation regarding the effects of learning biases vis-à-vis parsing factors. These stipulated subrankings require further thought.

Whether the leading edge corollary suppresses orders derived by rightward movement altogether (as suggested above) or permits such orders to emerge very infrequently is an empirical question that is hard to answer on the basis of the available evidence. Dryer (2018) identifies eight languages that potentially violate the analysis of universal 20 in Cinque 2005. Cinque (2025) systematically goes through these languages and argues that none of them provides a convincing counterexample. For example, the Num-Dem-A-N order found in Sierra Popoluca would require rightward movement of the A-N unit if the order corresponds to a single extended nominal projection. However, noun phrases with this order are interpreted as partitive (Wendy Liz Arbey López Márquez, p.c.) and are therefore likely to correspond to two extended nominal projections, with the numeral in the higher one and the rest of the material in the lower one. While the issue of very infrequent neutral orders is likely to require considerable additional research, Cinque's findings suggest that the effects of the leading edge corollary are strong enough to reduce the typological frequency of rightward movement orders to zero or very close to it.

5. Conclusion

In closing, we highlight the consequences of our analysis of universal 20 for the wider theory of syntax. As mentioned in the introduction (Section 1), universal 20 is one of the most intensely researched generalizations in linguistic typology and has turned out to be quite robust. The word-order restrictions it imposes have been replicated in artificial language-learning experiments (see Martin et al. 2019 and Martin et al. 2024), and the signature asymmetry of the pattern—fixed order to the left of a lexical head and variable order to the right—has been found well beyond the extended nominal projection. Cinque (2009) discusses instantiations of the universal 20 pattern with adjectives and nouns (based on the adjectival hierarchy in 44a), with PPs and verbs (based on the hierarchy in 44b), with adverbials and verbs (based on the hierarchy in 44c), and with verbs and mood, tense, and aspect markers (based on the hierarchy in 44d; mood comprises declarative markers, interrogative markers, imperative markers, etc.).

(44) a. $A_{SIZE} > A_{COLOR} > A_{NATIONALITY} > N$
b. $PP_{TIME} > PP_{PLACE}\ PP_{MANNER} > V$
c. *no longer* > *always* > *completely* > V
d. Mood > Tense > Aspect > V

In addition, Abels (2016) demonstrates that word order in Germanic verb clusters closely matches universal 20, while Neeleman (2017) shows that the same is true of the distribution of prepositional

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

phrases with regard to the verb in Dutch. It seems, then, that the pattern is ubiquitous and fundamental to syntax.

The analysis developed in the previous sections leads to the conclusion that the universal 20 pattern emerges from the interaction of two factors. First, the structure of the noun phrase is characterized by a universal order of attachment of modifiers combined with variability in the linearization of sister nodes. This combination of universal hierarchy and variable linearization is of course typical of phrase structure theory in generative grammar. The second factor, we have argued, consists of the parsing difficulties associated with rightward movement of obligatory material. If this is the correct assessment, we can remove from universal grammar one of the stipulations in Abels & Neeleman 2012, namely, that movements that derive neutral word orders are uniformly leftward. This is obviously an explanatory gain.

Data availability statement. Stimuli and tasks for the self-paced reading and maze experiments are available at Gorilla (https://app.gorilla.sc/openmaterials/1263243 and https://app.gorilla.sc/openmaterials/1288197, respectively). Raw data and analysis scripts are available at Trolling (https://doi.org/10.18710/HWWGL5).

Acknowledgments. Versions of this paper were presented at GLOW 35 (Potsdam, 2012), at UC Santa Cruz (2013), at the University of Vienna (2023), at UMass (2023), at the 34th Colloquium on Generative Grammar (Madrid, 2025), at the 16th International Conference on Linguistic Research and Applications (INALCO, Paris, 2025), and at the University of Geneva (2026). We would like to thank the audiences for useful questions and comments. Constructive criticism, initially from Adrian Staub and subsequently from two referees, editor John Beavers, and associate editor Michael Putnam, prompted us to rethink the interpretation of our experimental findings. We would also like to thank Laura Aldridge and Aira Duncanson-Hunter for help with the preparation of the experiments, Jamie White for discussion of statistical matters, Sebastian Schuster for help with the calculation of AI-based measures of surprisal, and Andrew Lamont and Edward Göbbel for help with determining the prosody of heavy-XP shift. Finally, we received useful feedback on earlier drafts from Peter Ackema, Wing-Yee Chow, Guglielmo Cinque, Liliana Nentcheva, and Hans van de Koot. [Full editorial history: Received 02 December 2023; revision invited 15 December 2024; revision received 18 July 2025; revision invited 10 December 2025; revision received 26 February 2026; accepted 28 March 2026.]

Competing interests. The authors have no competing interests to declare.

Ethics statement. The experiments reported in this article were carried out in line with UCL's ethics guidelines and were approved by the UCL Research Ethics Committee (Project ID number: LING-2015-03-23).

References

Abels, Klaus. 2016. The fundamental left-right asymmetry in the Germanic verb cluster. *Journal of Comparative Germanic Linguistics* 19.179–220. https://doi.org/10.1007/s10828-016-9082-9.

Abels, Klaus, and Ad Neeleman. 2012. Linear asymmetries and the LCA. *Syntax* 15.25–74. https://doi.org/10.1111/j.1467-9612.2011.00163.x.

Ackema, Peter, and Ad Neeleman. 2002. Effects of short-term storage in processing rightward movement. *Storage and computation in the language faculty*, ed. by Sieb Nooteboom, Fred Weerman, and Frank Wijnen, 219–56. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-010-0355-1_8.

Bach, Emmon. 1971. Questions. *Linguistic Inquiry* 2.153–66. https://www.jstor.org/stable/4177622.

Baker, Mark. 2003. *Lexical categories: Verbs, nouns and adjectives*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511615047.

Barker, Chris. 2012. Quantificational binding does not require c-command. *Linguistic Inquiry* 43.614–33. https://doi.org/10.1162/ling_a_00108.

Boyce, Veronica; Richard Futrell; and Roger Levy. 2020. Maze made easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language* 111:104082. https://doi.org/10.1016/j.jml.2019.104082.

Bresnan, Joan. 1976. On the form and functioning of transformations. *Linguistic Inquiry* 7.3–40. https://www.jstor.org/stable/4177911.

Bruening, Benjamin. 2014. Precedence-and-command revisited. *Language* 90.342–88. https://doi.org/10.1353/lan.2014.0037.

Chierchia, Gennaro, and Raymond Turner. 1988. Semantics and property theory. *Linguistics and Philosophy* 11.261–302. https://doi.org/10.1007/BF00632905.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Knowledge of language*. New York: Praeger.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. by Kenneth Hale and Samuel Jay Keyser, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

Christianson, Kiel; Andrew Hollingworth; John F. Halliwell; and Fernanda Ferreira. 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42.368–407. https://doi.org/10.1006/cogp.2001.0752.
Cinque, Guglielmo. 1996. The antisymmetric programme: Theoretical and typological implications. *Journal of Linguistics* 32.447–64. https://doi.org/10.1017/S0022226700015966.
Cinque, Guglielmo. 2005. Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry* 36.315–32. https://doi.org/10.1162/0024389054396917.
Cinque, Guglielmo. 2009. The fundamental left–right asymmetry of natural languages. *Universals of language today*, ed. by Sergio Scalise, Elisabetta Magni, and Antonietta Bisetto, 165–84. Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-8825-4_9.
Cinque, Guglielmo. 2025. The superiority of a movement approach to Greenberg's universal 20 (a reply to Dryer 2018). *Syntactic Theory and Research* 1(1):7. https://doi.org/10.16995/star.17346.
Culbertson, Jennifer, and Elissa L. Newport. 2015. Harmonic biases in child learners: In support of language universals. *Cognition* 139.71–82. https://doi.org/10.1016/j.cognition.2015.02.007.
Culbertson, Jennifer; Paul Smolensky; and Géraldine Legendre. 2012. Learning biases predict a word order universal. *Cognition* 122.306–29. https://doi.org/10.1016/j.cognition.2011.10.017.
de Vicenzi, Marica. 1991. *Syntactic parsing strategies in Italian: The minimal chain principle*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-3184-1.
den Besten, Hans. 1983. On the interaction of root transformation and lexical deletive rules. *On the formal syntax of the Westgermania*, ed. by Werner Abraham, 47–131. Amsterdam: John Benjamins.
Drummond, Alex; Norbert Hornstein; and Howard Lasnik. 2010. A puzzle about P-stranding and a possible solution. *Linguistic Inquiry* 41.689–92. https://doi.org/10.1162/LING_a_00018.
Dryer, Matthew S. 2007. Word order. *Language typology and syntactic description, vol. 1: Clause structure*, 2nd edn., ed. by Timothy Shopen, 61–131. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511619427.002.
Dryer, Matthew S. 2018. On the order of demonstrative, numeral, adjective, and noun. *Language* 94.798–833. https://doi.org/10.1353/lan.2018.0054.
Emonds, Joseph. 1976. *A transformational approach to English syntax: Root, structure-preserving, and local transformations*. New York: Academic Press.
Enochson, Kelly, and Jennifer Culbertson. 2015. Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PLoS ONE* 10:0116946. https://doi.org/10.1371/journal.pone.0116946.
Ferreira, Fernanda. 2003. The misinterpretation of noncanonical sentences. *Cognitive Psychology* 47.164–203. https://doi.org/10.1016/S0010-0285(03)00005-7.
Fodor, Janet Dean. 1998. Parsing to learn. *Journal of Psycholinguistic Research* 27.339–74. https://doi.org/10.1023/A:1023255705029.
Forster, Kenneth I.; Christine Guerrera; and Lisa Elliot. 2009. The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods* 41.163–71. https://doi.org/10.3758/BRM.41.1.163.
Frazier, Lyn. 1978. *On comprehending sentences: Syntactic parsing strategies*. Storrs: University of Connecticut dissertation. https://digitalcommons.lib.uconn.edu/dissertations/AAI7914150.
Frazier, Lyn. 1987. Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory* 5.519–59. https://doi.org/10.1007/BF00138988.
Frazier, Lyn, and Giovanni B. Flores d'Arcais. 1989. Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language* 28.331–44. https://doi.org/10.1016/0749-596X(89)90037-5.
Frazier, Lyn, and Janet Dean Fodor. 1978. The sausage machine: A new two-stage parsing model. *Cognition* 6.291–326. https://doi.org/10.1016/0010-0277(78)90002-1.
Frazier, Lyn, and Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements and the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14.178–210. https://doi.org/10.1016/0010-0285(82)90008-1.
Freedman, Sandra E., and Kenneth I. Forster. 1985. The psychological status of overgenerated sentences. *Cognition* 19.101–31. https://doi.org/10.1016/0010-0277(85)90015-0.
Freitag, Constantin. 2021. *Verb-second as a reconstruction phenomenon: Evidence from grammar and processing*. Berlin: De Gruyter. https://doi.org/10.1515/9783110725018.
Friederici, Angela D., and Stefan Frisch. 2000. Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language* 43.476–507. https://doi.org/10.1006/jmla.2000.2709.
Gagliardi, Ann. 2012. *Input and intake in language acquisition*. College Park: University of Maryland dissertation.
Georgi, Doreen, and Gereon Müller. 2010. Noun-phrase structure by reprojection. *Syntax* 13.1–36. https://doi.org/10.1111/j.1467-9612.2009.00132.x.
Göbbel, Edward. 2024. Heavy NP shift in English: Phonological solutions. *Phonological Externalization* 9.39–74. https://sapporo-u.repo.nii.ac.jp/record/2000117/files/SULB00001437.pdf.
Gorrell, Paul. 1995. *Syntax and parsing*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627682.
Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*, ed. by Joseph Greenberg, 73–113. Cambridge, MA: MIT Press.
Grillo, Nino, and João Costa. 2014. A novel argument for the universality of parsing principles. *Cognition* 133.156–87. https://doi.org/10.1016/j.cognition.2014.05.019.
Grimshaw, Jane. 1991. *Extended projection*. New Brunswick, NJ: Rutgers University, ms.
Grimshaw, Jane. 2005. *Words and structure*. Stanford, CA: CSLI Publications.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

Haider, Hubert. 2005. How to turn German into Icelandic—and derive the OV–VO contrasts. *Journal of Comparative Germanic Linguistics* 8.1–53. https://doi.org/10.1007/s10828-004-0293-0.

Haider, Hubert. 2014. *Symmetry breaking in syntax*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139084635.

Hawkins, John A. 1983. *Word order universals: Quantitative analyses of linguistic structure*. New York: Academic Press.

Hawkins, John A. 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21.223–61. https://www.jstor.org/stable/4178670.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199252695.001.0001.

Hawkins, John A. 2009. Language universals and the performance-grammar correspondence hypothesis. *Language universals*, ed. by Morten H. Christiansen, Christopher Collins, and Shimon Edelmann, 54–78. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195305432.003.0004.

Huang, C.-T. James. 1982. *Logical relations in Chinese and the theory of grammar*. Cambridge, MA: MIT dissertation. http://hdl.handle.net/1721.1/15215.

Jackendoff, Ray. 1987. *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Jaeger, Florian. 2008. Modeling self-paced reading data: Effects of word length, word position, spill-over, etc. https://hlplab.wordpress.com/2008/01/23/modeling-self-paced-reading-data-effects-of-word-length-word-position-spill-over-etc/.

Just, Marcel A., and Patricia A. Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99.122–49. https://doi.org/10.1037/0033-295X.99.1.122.

Just, Marcel A.; Patricia A. Carpenter; and Jacqueline D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111.228–38. https://doi.org/10.1037/0096-3445.111.2.228.

Kamp, Hans, and Barbara Partee. 1995. Prototype theory and compositionality. *Cognition* 57.129–91. https://doi.org/10.1016/0010-0277(94)00659-9.

Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Kimball, John. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2.15–47. https://doi.org/10.1016/0010-0277(72)90028-5.

Kirby, Simon. 1999. *Function, selection, and innateness: The emergence of language universals*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198238119.001.0001.

Konieczny, Lars; Barbara Hemforth; Christoph Scheepers; and Gerhard Strube. 1997. The role of lexical heads in parsing: Evidence from German. *Language and Cognitive Processes* 12.307–48. https://doi.org/10.1080/016909697386871.

Koster, Jan. 1975. Dutch as an SOV language. *Linguistic Analysis* 1.111–36.

Kuznetsova, Alexandra; Per B. Brockhoff; and Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82.1–26. https://doi.org/10.18637/jss.v082.i13.

Lasnik, Howard, and Mamoru Saito. 1992. *Move α: Conditions on its application and output*. Cambridge, MA: MIT Press.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106.1126–77. https://doi.org/10.1016/j.cognition.2007.05.006.

Marcus, Mitchell; Donald Hindle; and Margaret Fleck. 1983. D-theory: Talking about talking about trees. *Proceedings of the 21st annual meeting of the Association for Computational Linguistics*, 129–36. https://doi.org/10.3115/981311.981337.

Martin, Alexander; David Adger; Klaus Abels; Patrick Kanampiu; and Jennifer Culbertson. 2024. A universal cognitive bias in word order: Evidence from speakers whose language goes against it. *Psychological Science* 35.304–11. https://doi.org/10.1177/09567976231222836.

Martin, Alexander; Theeraporn Ratitamkul; Klaus Abels; David Adger; and Jennifer Culbertson. 2019. Cross-linguistic evidence for cognitive universals in the noun phrase. *Linguistics Vanguard* 5:20180072. https://doi.org/10.1515/lingvan-2018-0072.

Mayr, Clemens, and Benjamin Spector. 2011. Not too strong! Generalizing the scope economy condition. *Proceedings of Sinn und Bedeutung* 14.305–21. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/474.

Michel, Jean-Baptiste; Yuan Kui Shen; Aviva Presser Aiden; Adrian Veres; Matthew K. Gray; The Google Books Team; Joseph P. Pickett; Dale Hoiberg; Dan Clancy; Peter Norvig; et al. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331.176–82. https://doi.org/10.1126/science.1199644.

Mitchell, Don. 1984. An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. *New methods in reading comprehension research*, ed. by David Kieras and Marcel Just, 69–89. Hillsdale, NJ: Lawrence Erlbaum.

Nchare, Abdoulaye Laziz. 2012. *The grammar of Shupamem*. New York: New York University dissertation.

Neeleman, Ad. 2017. PP-over-V meets universal 20. *Journal of Comparative Germanic Linguistics* 20.3–47. https://doi.org/10.1007/s10828-017-9085-1.

Neeleman, Ad, and Amanda Payne. 2020. PP-extraposition and the order of adverbials in English. *Linguistic Inquiry* 51.471–520. https://doi.org/10.1162/ling_a_00352.

Oh, Byung-Doh, and William Schuler. 2023. Why does surprisal from larger transformer-based language models provide a poorer fit to human reading times? *Transactions of the Association for Computational Linguistics* 11.336–50. https://doi.org/10.1162/tacl_a_00548.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press

Omaki, Akira. 2010. *Commitment and flexibility in the developing parser*. College Park: University of Maryland dissertation. http://hdl.handle.net/1903/10911.
Overfelt, Jason. 2015. *Rightward movement: A study in locality*. Amherst: University of Massachusetts Amherst dissertation. https://doi.org/10.7275/7536511.0.
Partee, Barbara H.; Alice ter Meulen; and Robert E. Wall. 1993. *Mathematical methods in linguistics*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-009-2213-6.
Perlmutter, David. 1983. *Studies in relational grammar 1*. Chicago: University of Chicago Press.
Phillips, Colin, and Matthew Wagers. 2007. Relating structure and time in linguistics and psycholinguistics. *The Oxford handbook of psycholinguistics*, ed. by Gareth Gaskell, 739–56. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198568971.013.0045.
R Core Team. 2023. R: A language and environment for statistical computing. Version 4.3.1, 2023-06-16. Vienna: R Foundation for Statistical Computing.
Rijkhoff, Jan. 1990. Explaining word order in the noun phrase. *Linguistics* 28.5–42. https://doi.org/10.1515/ling.1990.28.1.5.
Rijkhoff, Jan. 2002. *The noun phrase*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237822.001.0001.
Rochemont, Michael, and Peter Culicover. 1990. *English focus constructions and the theory of grammar*. Cambridge: Cambridge University Press.
Ross, John Robert. 1967. *Constraints on variables in syntax*. Cambridge, MA: MIT dissertation. http://hdl.handle.net/1721.1/15166.
Santi, Andrea; Nino Grillo; Emilia Molimpakis; and Michael Wagner. 2019. Processing relative clauses across comprehension and production: Similarities and differences. *Language, Cognition and Neuroscience* 34.170–89. https://doi.org/10.1080/23273798.2018.1513539.
Staub, Adrian; Charles Clifton Jr.; and Lyn Frazier. 2006. Heavy NP shift is the parser's last resort: Evidence from eye movements. *Journal of Memory and Language* 54.389–406. https://doi.org/10.1016/j.jml.2005.12.002.
Steddy, Sam, and Vieri Samek-Lodovici. 2011. On the ungrammaticality of remnant movement in the derivation of Greenberg's universal 20. *Linguistic Inquiry* 42.445–69. https://doi.org/10.1162/LING_a_00053.
Steedman, Mark. 2011. *Taking scope*. Cambridge, MA: MIT Press.
Steedman, Mark. 2020. A formal universal of natural language grammar. *Language* 96.618–60. https://doi.org/10.1353/lan.2020.0043.
Stowe, Laurie A. 1986. Parsing WH-constructions: Evidence for online gap location. *Language and Cognitive Processes* 1.227–45. https://doi.org/10.1080/01690968608407062.
Townsend, David, and Thomas G. Bever. 2001. *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/6184.001.0001.
Travis, Lisa deMena. 1984. *Parameters and effects of word order variation*. Cambridge, MA: MIT dissertation. http://hdl.handle.net/1721.1/15211.
Traxler, Matthew J., and Martin J. Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35.454–75. https://doi.org/10.1006/jmla.1996.0025.
Truswell, Robert. 2004. Attributive adjectives and the nominals they modify. Oxford: University of Oxford M.Phil. thesis.
Valian, Virginia. 1990. Logical and psychological constraints on the acquisition of syntax. *Language processing and language acquisition*, ed. by Lyn Frazier and Jill de Villiers, 119–45. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-3808-6_5.
Wasow, Thomas. 1997a. End-weight from the speaker's perspective. *Journal of Psycholinguistic Research* 26.347–61. https://doi.org/10.1023/A:1025080709112.
Wasow, Thomas. 1997b. Remarks on grammatical weight. *Language Variation and Change* 9.81–105. https://doi.org/10.1017/S0954394500001800.
Williams, Edwin. 1974. *Rule ordering in syntax*. Cambridge, MA: MIT dissertation. http://hdl.handle.net/1721.1/14770.

Klaus Abels is Professor of Linguistics at University College London. Email: k.abels@ucl.ac.uk.

Ad Neeleman is Professor of Linguistics at University College London. Email: a.neeleman@ucl.ac.uk.

Cite this article: Abels, Klaus, and Ad Neeleman. 2026. Linear asymmetries and incremental parsing. *Language* 102, 329–371. https://doi.org/10.1017/S009785072600010X.

https://doi.org/10.1017/S009785072600010X Published online by Cambridge University Press