e[1]

Klaus Abels and Ad Neeleman
UCL Linguistics, Chandler House, 2 Wakefield Street, London, WC1N 1PF

A long-standing question in syntax is how to deal with linear asymmetries observed in natural language. Given its inherently left-to-right nature, the parsing process is a plausible source of such asymmetries. Adopting this line of reasoning, we argue that rightward movement of obligatory material causes parsing difficulties that are absent with leftward movement and rightward movement of optional material. Evidence comes from heavy-XP shift. We discuss the results of eye-tracking experiments reported in Staub, Clifton and Frazier 2006 and replicate and extend these through a self-paced reading experiment and an experiment using the maze paradigm. We then show that the parsing constraints that explain why heavy XP-shift of obligatory material is problematic can be brought to bear on the linear asymmetry described in Greenberg's (1963) Universal 20, one of the most robust typological universals in existence. Cinque (2005) and Abels and Neeleman (2012) show that Universal 20 can be derived if movement of the noun within the extended nominal projection is uniformly leftward. This restriction to leftward movement can be modelled as resulting from difficulties in parsing that drastically suppress the typological frequency of rightward movement orders (as per Hawkins 1990, 2009 and Kirby 1999).

1. Nature abhors a vacuum

There is no doubt that there are pervasive left-right asymmetries in sentence grammar. These include asymmetries in binding (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 and Neeleman 2012, Abels 2016, Neeleman 2017, Neeleman and Payne 2020). The question of how to account for left-right asymmetries therefore remains open, despite its vital importance for syntactic theory.

In this paper, we aim to contribute to this issue by showing that certain rightward movements are problematic in parsing and hence subject to typological attrition if given enough time. This explains, we think, a class of left-right asymmetries found with movement. We will give details later, but limit ourselves here to a sketch of the proposal.

Our starting point is the assumption, widely shared in current syntactic research, that traces are unpronounced copies (or partial copies) of moved categories (Chomsky 1993, 1995).

[1] Earlier 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), and at UCL (Intermediate Generative Grammar A, 2023). We would like to thank the audiences for useful questions and comments. We benefitted in particular from comments by Adrian Staub, who prompted us to rethink the interpretation of some of our experimental findings. We would further like to thank Laura Aldridge and Samuel Duncanson-Hunter for help with the preparation of the experiments and Jamie White for discussion of statistical matters. Finally, we have received useful input from Peter Ackema, Wing-Yee Chow, Guglielmo Cinque, Liliana Nentcheva and Hans van de Koot.

This characterization of traces has important consequences for how movement depencencies are parsed. Given that traces depend for their content on the moved category, it follows 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. Thus, the copy theory of movement implies a filler-driven approach to the parsing of movement dependencies (as opposed to a gap-driven strategy). That is a good result, as the psycholinguistic literature contains a wealth of evidence for filler-driven parsing of leftward movement dependencies like *Wh*-fronting (see Stowe 1986, Frazier and Flores d'Arcais 1989, Traxler and Pickering 1996, among many others).

On the filler-driven strategy, incremental parsing of a string X-A-B-C-D-E-F, in which 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 (1a). (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 (1b) to (1c) (the symbol ***e*** represents the empty position). (iii) Subsequently, a silent copy of X is inserted and X itself is deactivated as a filler, as in (1d). (iv) Finally, the rest of the input string is parsed, as in (1e).

(1) a. $\boxed{\text{X}}$ Leftward movement
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

The thing to note about the sequence of steps in (1) is that the empty position ***e*** is filled by a copy of X immediately after it has been created (i.e. before D is parsed). This sequence of events can be contrasted with the parsing steps involved in incremental analysis of a string A-B-C-D-E-F-X, in which an element X that would normally obligatorily follow C has moved rightward. (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 (2a) to (2b). (ii) As the parser has not encountered X yet, ***e*** must remain empty while the substring D-E-F is parsed, as in (2c). (iii) It is only when X presents itself that the parser can hypothesize that it has moved, thus activating it as a filler, as in (2d). (iv) This, then, makes it possible for the parser to initiate a backward search for a position in which to insert a copy of X. (v) This search will of course identify ***e*** as a possible insertion site, after which the movement dependency can be established, as in (2e). Thus, the rightward movement of obligatory material gives rise to a temporal separation of the moment that ***e*** is created and the moment that it is filled.

(2) a. A B C Rightward movement (obligatory material)
b. A B C ***e***
c. A B C ***e*** D E F
d. A B C ***e*** D E F $\boxed{\text{X}}$
e. A B C ~~X~~ D E F X

The sequence of steps in (2) can in turn be distinguished from those necessary to parse a string A-B-C-D-E-F-X in which an element 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 (3a,b). (iii) When the parser then encounters X, it can activate this element as a filler, as in (3c), 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 (3d). (v) Subsequenly, a copy of X can be inserted, so that the movement dependency is resolved, as in (3e). Thus, this procedure resembles the leftward movement procedure rather than the procedure for rightward movement of obligatory material: like in (1) and unlike in (2), ***e*** is filled immediately after it has been created.

(3) a. A B C — Rightward movement (optional material)
 b. A B C D E F
 c. A B C D E F $\boxed{\text{X}}$
 d. A B C ***e*** D E F $\boxed{\text{X}}$
 e. A B C ~~X~~ D E F X

As this discussion demonstrates, rightward movement of obligatory material is qualitatively different from both leftward movement and rightward movement of optional material, at least on the assumptions that parsing is incremental, that movement is parsed using a filler-driven strategy, and that the parser creates empty positions only when necessary (in order to accommodate a trace or because of predictions by preceding material). We assume that this qualitative difference creates difficulties in the parsing of rightward movement of obligatory material that are absent otherwise.

It is important that we identify the exact source of the difficulty. This, we hypothesize, is not so much that in (2) ***e*** fails to be filled immediately, but rather that it fails to be filled and is removed from the leading edge of the parse tree (i.e. it is followed by the substring D-E-F before being filled). There is reason to think that the parser tolerates empty syntactic positions that remain unfilled as long as they are pushed rightward as further material is processed. For example, suppose that the substring A-B-C predicts that an item of type F will follow but not that this item will immediately follow C. Then, a empty syntactic position may be created that shifts rightward until the parser encounters F, as in (4).

(4) a. A B C ***e***
 b. A B C D ***e***
 c. A B C D E ***e***
 d. A B C D E F

Thus, the problem with rightward movement of obligatory material lies in step (2c), which violates the Leading Edge Constraint in (5). By contrast, the Leading Edge Constraint is met in leftward movement configurations or when optional material moves rightward.

(5) *Leading Edge Constraint* (LEC)
 e must be filled immediately or remain at the leading edge of the parse tree.

One reason that we are interested in parsing difficulties associated with rightward movement of obligatory material is a set of observations known as Universal 20 (Greenberg 1963; Hawkins 1983; Rijkhoff 1990, 2002; Cinque 2005, 2014; Dryer 2018). Greenberg's original formulation and data patterns are given in (6) and (7) below. In (7) we also give the number attestations (languages, genera) according to Cinque 2014 and Dryer 2018, respectively.

(6) "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."

(7) a. Dem Num A N — attested (507, 126) (113, 57)
b. N A Num Dem — attested (358, 107) (182, 85)
c. N Dem Num A — attested (69, 24) (8, 6)
d. *A Num Dem N — unattested (0,0) (0, 0)

At the heart of universal 20 is the observation that cross-linguistically the order of elements modifying the noun is constant pre-nominally but varies post-nominally. (This observation pertains to neutral word order. When marked structures are taken into account, the number of attested orders increases considerably.) It is shown in Abels and Neeleman 2012 that this pattern can be explained under the simultaneous assumption of symmetric base generation and asymmetric, leftward only, movement (see also Ackema and Neeleman 2002; for a contrasting analysis, see Cinque 1996, 2005). By symmetric base generation we mean that there is no universal ordering of a projecting category and its sister. Thus, the hierarchy in the noun phrase is presumed to be constant cross-linguistically, but demonstratives, numerals and adjectives can be linearized to the left or to the right of the noun depending on the grammar of the particular language:

(8) *Symmetric base-generation:*
a. [Dem [Num [A N]]] — attested
b. [[[N A] Num] Dem] — attested

By asymmetric movement we mean that further neutral word order can be generated through leftward but never rightward movement of the noun:

(9) *Asymmetric movement:*
a. [N [Dem [Num [A ~~N~~]]]] — attested
b. *[[[[~~N~~ A] Num] Dem] N] — unattested

We think that the LEC may explain why the movement component in the analysis of Universal 20 is asymmetric. After all, the noun is an obligatory element in the noun phrase and the movement in (9b) is rightward. Hence, by the logic outlined above the string A-Num-Dem-N will require a parsing process in which an empty position ***e*** is generated that is subsequently pushed from the leading edge of the parse tree. The hypothesis that rightward movement of the noun gives rise to parsing difficulties implies that pressure is exerted towards language change away from the offending grammar. Such pressure can explain the why the structure in (9b) is unattested, even if we assume that the language faculty can in principle generate it (see Hawkins 1990, 2009 and Kirby 1999 for discussion on how the parser can shape typology; for a related account of Universal 20, see Ackema and Neeleman 2002).

The account just sketched hinges on the assumption that the string A-Num-Dem-N is hard to parse because its generation requires rightward movement of an obligatory element. But of course that assumption cannot be tested directly; after all, the A-Num-Dem-N order is not attested and so its processing profile must remain inaccessible. However, the LEC makes predictions beyond the A-Num-Dem-N order that *are* testable. In particular, its effects should extend to a well-known rightward movement process in English, namely heavy-XP shift (see Ross 1967, Rochemont and 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 (10) and (11).

(10) a. John ate (the food that his brother prepared).
b. John ate t_1 yesterday [the food that his brother prepared]$_1$.

(11) a. John devoured *(the food that his brother prepared).
b. John devoured t_1 yesterday [the food that his brother prepared]$_1$.

While the representations of (10b) and (11b) are indistinguishable from a syntactic point of view, they are quite different as regards the parsing process they require. Heavy-XP shift of the object of an optionally transitive verb will instantiate (3) and hence it will not violate the LEC (see (12)). Heavy-XP shift of the object of an obligatorily transitive verb, however, will require early postulation of ***e*** and hence induce a violation of the LEC. The crucial parsing step is (13c).

(12) a. John ate — Heavy XP shift (optional object)
b. John ate yesterday
c. John ate yesterday $\boxed{\text{the food that his brother prepared}}$
d. John ate ***e*** yesterday $\boxed{\text{the food that his brother prepared}}$
e. John ate ~~the food that his brother prepared~~ yesterday the food that his brother prepared

(13) a. John devoured — Heavy XP shift (obligatory object)
b. John devoured ***e***
c. John devoured ***e*** yesterday
d. John devoured ***e*** yesterday $\boxed{\text{the food that his brother prepared}}$
e. John devoured ~~the food that his brother prepared~~ yesterday the food that his brother prepared

Hence, if the LEC is real, we make the following prediction:

(14) Heavy-XP shift of obligatory objects should cause parsing difficulties during the processing of material crossed by the movement; heavy-XP shift of optional objects should not.

This is not a self-evident prediction. An alternative hypothesis could be that the parser employs a gap-driven strategy for rightward movement, in which case strong evidence for a gap would help rather than hinder the parsing process.

There is a further, more subtle, prediction. In both (12) and (13) the parser identifies the shifted object as a filler and starts a backward search to find a place to insert a copy. In (12) there is one subsequent step, namely the placement of a copy in ***e***; in (13), however, there are *two* subsequent steps, as ***e*** 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 XP, there should be a reversal in parsing effort:

(15) 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.

In sum, certain assumptions about parsing can be used to explain why the movement component in the analysis of Universal 20 is asymmetric. Those assumptions can be tested by looking at the way heavy-XP shift is parsed in structures with optionally or obligatorily transitive verbs.

Given the central role of heavy-XP shift for our overall argument, we will first consider the predictions in (14) and (15). The literature already contains work that suggests these predictions 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 2, we will report on two web-based experiments intended to replicate and extend Staub et al.'s findings: a self-paced reading task (Just et al. 1982, Mitchell 1984) and a maze task (Freedman and Forster 1985; Forster et al. 2009; Boyce et al. 2020). The self-paced reading taks confirms the prediction in (14), 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 Staub et al.'s work concerns sentences in which heavy-XP shift crosses not one, but two modifiers. The LEC predicts that parsing difficulties persist over both modifiers. One can imagine a more shallow account based on surprisal when the parser does not encounter an object directly after an obligatorily transitive verb. Such surprisal should subside over time, so that any slow-down in reading is expected to decrease over the span of modifiers. Both the self-paced reading task and the maze task show, however, that the twomodifiers induce comparable parsing difficulties.

Having provided a firm empirical footing for our proposal, we develop our model of parsing in more detail in section 3 and show how it applies to heavy XP shift in section 4. In section 5, we take a closer look at Universal 20. We describe the pattern in more detail and explain how exactly our model of parsing derives the absence of rightward movement of the noun. We will also discuss why heavy-XP shift of obligatory objects survives, while rightward movement of the noun is unattested, even though the two movement operations by hypothesis give rise to the exact same complications in parsing. Section 6 concludes.

2. Testing the predictions

2.1 Literature review

The literature provides convergent evidence from two sources that confirms the predictions in (14) and (15): the corpus studies by Wasow (1997a,b) and the above-mentioned eye-tracking experiments by Staub et al. (2006).

At first sight it may seem surprising that we claim that corpus studies support claims about parsing. After all, corpora are the output of production, rather than comprehension. However, it is well known that speakers parse their own speech (Levelt 1989). For example, feedback of this type must be assumed to explain how speakers can spot errors in their production. Further evidence comes from experiments showing a breakdown in production when subjects hear their own speech slightly delayed through headphones (see Jackendoff 1987 and references given there). Consequently, parsing difficulties should express themselves not only in perception but also in production, a point made by Hawkins in a number of publications (see in particular Hawkins 1990, 2009). For written corpora copy-editing provides an additional feedback loop.

Wasow (1997a,b) carried out two corpus studies which 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 both in written and spoken corpora, as the following table shows:

Brown corpus (written) $\chi^2(1) = 6.49$, $p < .02$		Switchboard corpus (spoken) $\chi^2(1) = 10.65$, $p < .01$	
V*t*	V*p*	V*t*	V*p*
5.6%	9.3%	1.45%	3.82%

Table 1: Percentage of heavy-XP-shifted DP-objects in two corpora as determined by verb transitivity

These findings are of course consistent with the prediction that heavy-XP shift of obligatory objects causes parsing difficulties that do not arise with heavy-XP shift of optional objects.

The two eye-tracking experiments reported in by Staub et al. 2006 confirm the prediction in (14) more directly. The experiments were designed to determine how the selectional requirements of the verb influence parsing of heavy-XP shift. On the assumption that parsing difficulties give 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 slow-down 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 is 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 mere surprisal 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 the object should be there in the first place.

The findings reported in Staub et al. are summarized in Table 2.

	Verb	Adverbial	Object
V_{OblTr}	no difference	slower on all measures	faster on some measures
V_{OptTr}		faster on all measures	slower on some measures

Table 2: Comparison of reading times in three regions of sentences with a heavy-XP-shifted object

As the table shows, Staub et al. also found a difference in the processing of the shifted object. They suggest this difference is 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 slow-down on some measures compared to the processing of shifted obligatory objects. Staub et al.'s interpretation of the 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 at al. (p. 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." Thus, Staub et al.'s findings confirm the prediction in (15).

Given the importance of predictions in (14) and (15) for our overall argumentation, we ran two additional experiments. These were in the first instance intended to replicate Staub et al.'s findings using different experimental paradigms (self-paced reading and a maze task). However, we added a condition to the experimental set-up in which the heavy-XP-shifted object crosses not one but two adverbials, as shown in the following table:

	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 3: Structures tested in the self-paced reading and maze tasks

The point of this extension was to determine whether any slow-down on the material separating verb and object could be due to a shallow effect of surprisal. The LEC predicts that parsing difficulties should persist if more material is added between the verb and the object. After all, the empty position built to accommodate the object of an obligatorily transitive verb remains separated from the leading edge of the parse tree no matter how many adverbials intervene. Therefore reaction times over the adverbial of interest (boldfaced in the table) should remain constant whether a 'delay adverbial' (Adv_{delay}) is present or not. By constrast, if the slow-down 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 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:

(16) a. Reaction times over the adverbial of interest will be slower in the obligatorily transitive condition than in the optionally transitive condition, irrespective of the presence of 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.

Predictions (16a) and (16b) are operalizations of predictions (14) and (15) in terms of reaction times.

We report on the two experiments in the following subsections.

2.2 Replicating and extending Staub et al. 2006 through self-paced reading

The first experiment we ran was a web-based self-paced reading study. 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. 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 spill-over 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 makes up one experimental item and was shaped as in Table 3. Across each item, the subject, the adverbial of interests and the shifted objects 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 Staub et al.'s original items and which was held constant across items that had one. A sample item made up of four test sentences is given in (17). We have underlined relevant regions. Region 1 is the verb. Regions 2 and 4 are the adverbial of interest and the onset of the shifted object, respectively. Region 3 is the delay adverbial. In the example at hand, *attack* is optionally and *bother* obligatorily transitive.

(17) a. Sara attacked$_1$ with no mercy$_2$ the red ants$_4$ living on the windowsill.
b. Sara attacked$_1$ for ten minutes$_3$ with no mercy$_2$ the red ants$_4$ living on the windowsill.
c. Sara bothered$_1$ with no mercy$_2$ the red ants$_4$ living on the windowsill.
d. Sara bothered$_1$ for ten minutes$_3$ with no mercy$_2$ 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 2x2 design, with the nature of the verb (obligatorily transitive vs optionally transtive) 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 (16). Prediction (16a) is confirmed if (i) reading times over 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 (16b) 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 list 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 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 recuited 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 was subjected to analysis.

In data cleaning, all 307 trials with a response time below 250ms or above 2500ms 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.

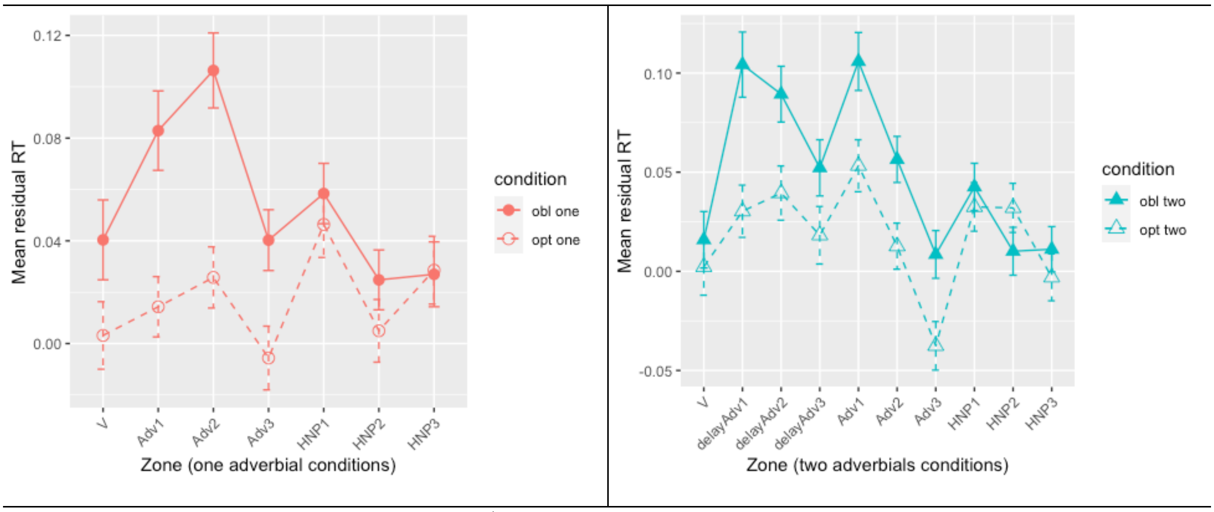


Table 4: Plots for reading times in the one/two adverbial conditions (*obl*: obligatorily transitive verb; *opt*: optionally transitive verb; *one*: one intervening adverbial; *two*: two intervening adverbials)

In Table 4, 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 Table 5.

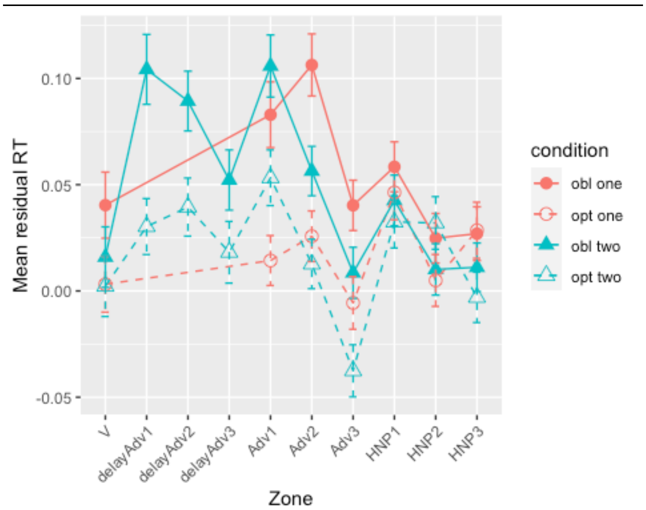


Table 5: Combined reading time plots (*obl*: obligatorily transitive verb; *opt*: optionally transitive verb; *one*: one intervening adverbial; *two*: two intervening adverbials)

We fitted seven linear mixed effects models to the resdiual reading times with contrast-coded conditions for the number of adverbials and the optional/obligatory transitivity of the verb. To do so, we used the lmer function of the lmerTest package (Kuznetsova et al. 2017) in R (version 4.3.1, 2023-06-16, R Core Team 2023): logRTresidual ~ OBL_c*PPS_c + (1 | itemnumber) + (1 | participant).

Significance testing was done using model reductions. A limited random effects structure was used to ensure that all models, including reduced ones, would converge non-defectively.

Significance levels are reported using Bonferroni correction to take into account the fact that seven models were fitted. Fixed effects for these models are summarized in table 6.

Zone	*term*	*Estimate*	*Std. Error*	*t value*	*df*	*uncorrected p*	*corrected significance*
V	(Intercept)	0.016	0.0149	1.07	50.2	0.2909	ns
V	obl_c	0.025	0.0136	1.87	1890.2	0.06218	ns
V	Adv_c	-0.013	0.0137	-0.92	1891.2	0.35516	ns
V	obl_c:Adv_c	-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_c	0.06	0.0135	4.46	1895.5	1e-5	****
Adv1	Adv_c	0.03	0.0135	2.24	1897.1	0.02518	ns
Adv1	obl_c:Adv_c	-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_c	0.062	0.0123	5.04	1895.2	0	****
Adv2	Adv_c	-0.031	0.0123	-2.52	1897	0.01168	ns
Adv2	obl_c:Adv_c	-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_c	0.046	0.0119	3.84	1896.2	1.3e-4	***
Adv3	Adv_c	-0.032	0.0119	-2.68	1898	0.00747	ns
Adv3	obl_c:Adv_c	0	0.0238	0	1895.2	0.9991	ns
HNP1	(Intercept)	0.045	0.0099	4.54	30	9e-5	***
HNP1	obl_c	0.011	0.012	0.9	1897.6	0.37019	ns
HNP1	Adv_c	-0.015	0.012	-1.23	1898.4	0.21945	ns
HNP1	obl_c:Adv_c	-0.003	0.024	-0.12	1897.1	0.90491	ns
HNP2	(Intercept)	0.018	0.0122	1.49	35.5	0.14478	ns
HNP2	obl_c	-0.001	0.0117	-0.11	1892	0.9111	ns
HNP2	Adv_c	0.006	0.0117	0.55	1892.5	0.58239	ns
HNP2	obl_c:Adv_c	-0.042	0.0234	-1.82	1891	0.06914	ns
HNP3	(Intercept)	0.017	0.0151	1.16	30.8	0.25626	ns
HNP3	obl_c	0.006	0.0118	0.5	1895.3	0.6184	ns
HNP3	Adv_c	-0.022	0.0118	-1.85	1895.9	0.06516	ns
HNP3	obl_c:Adv_c	0.014	0.0236	0.61	1894.9	0.53938	ns

Table 6: 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

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 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 (16a), but not prediction (16b). In other words, we have been 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 the verb is optionally transitive. Our results also suggests that the slowdown observed by Staub et al. cannot simply be due to surprisal when the parser encounters an adverbial where an object is expected. This is because the effect does not decrease in size when the adverbial of interest is placed after a delay adverbial. Thus, parsing difficulties persists, rather than subside, after the moment of surprisal. On our proposal, these difficulties are the result of a continuing violation of the LEC.

In the analysis reported above, the delay adverbial was excluded. However, the LEC predicts that the slowdown observed over the adverbial of interest when the verb is obligatorily transitive should extend to 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-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).

The fact that the experiment did not confirm that obligatory transitivity of the verb facilitiates 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 Staub et al.'s original study. This lack of replication is one reason why we carried out a second experiment in a different and arguably more sensitive paradigm.

2.3 Replicating and extending Staub et al. 2006 through 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). Like self-paced reading, the 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 postion. 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 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 specific instantiation of the maze task we use is known as G(rammatical)-Maze, which means that 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 which are comparable with their correct counterparts in frequence and length but which 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. validate 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 cases led to adjustments. All in all, there were twenty-four sets of four sentences shaped as in Table 3. Each set makes 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) are held constant. The verb varied and was either optionally or obligatorily transitive. The delay adverbial, which we added to Staub et al.'s original items, was the same in the items that had one. A sample quartet of test sentences is given below. The target words appear in the columns labelled a-d, with distractors to their immediate right. Words in regions relevant to the experiment appear underlined (region 1 is the verb, region 2 is the adverbial of interest, region 3 is the delay adverbial, and region 4 is the onset of the shifted object.) In this sample, *attack* is optionally and *bother* obligatorily transitive.

(18)

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
<u>attacked</u>$_1$	patience	<u>attacked</u>$_1$	patience	<u>bothered</u>$_1$	patience	<u>bothered</u>$_1$	patience
<u>with</u>$_2$	blog	<u>for</u>$_3$	tool	<u>with</u>$_2$	blog	<u>for</u>$_3$	tool
<u>no</u>$_2$	ha	<u>ten</u>$_3$	app	<u>no</u>$_2$	ha	<u>ten</u>$_3$	app
<u>mercy</u>$_2$	tends	<u>minutes</u>$_3$	improve	<u>mercy</u>$_2$	tends	<u>minutes</u>$_3$	improve
<u>the</u>$_4$	cent	<u>with</u>$_2$	blog	<u>the</u>$_4$	cent	<u>with</u>$_2$	blog
<u>red</u>$_4$	ago	<u>no</u>$_2$	ha	<u>red</u>$_4$	ago	<u>no</u>$_2$	ha
<u>ants</u>$_4$	vivo	<u>mercy</u>$_2$	tends	<u>ants</u>$_4$	vivo	<u>mercy</u>$_2$	tends
living	videos	<u>the</u>$_4$	cent	living	videos	<u>the</u>$_4$	cent
on	gain	<u>red</u>$_4$	ago	on	gain	<u>red</u>$_4$	ago
the	glad	<u>ants</u>$_4$	vivo	the	glad	<u>ants</u>$_4$	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 2x2 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 (16). Confirmation of prediction (16a) 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 (16b) requires that reading times at the onset of the heavy NP are faster in the obligatorily transitive conditions.

In line with the experiment's 2x2 design, the test items were distributed over four list 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 their participation in the experiment. Completing the task took about 40 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 participants were excluded because they indicated that their native language was not English. The data generated by the remaining 113 participants was submitted to further analysis.

The experiment generated 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 (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 below. Table 7 shows separate plots for the conditions with one and two intervening adverbials, respectively. Table 8 shows the combined plot. In all plots, the error bars indicate the standard error of the mean.

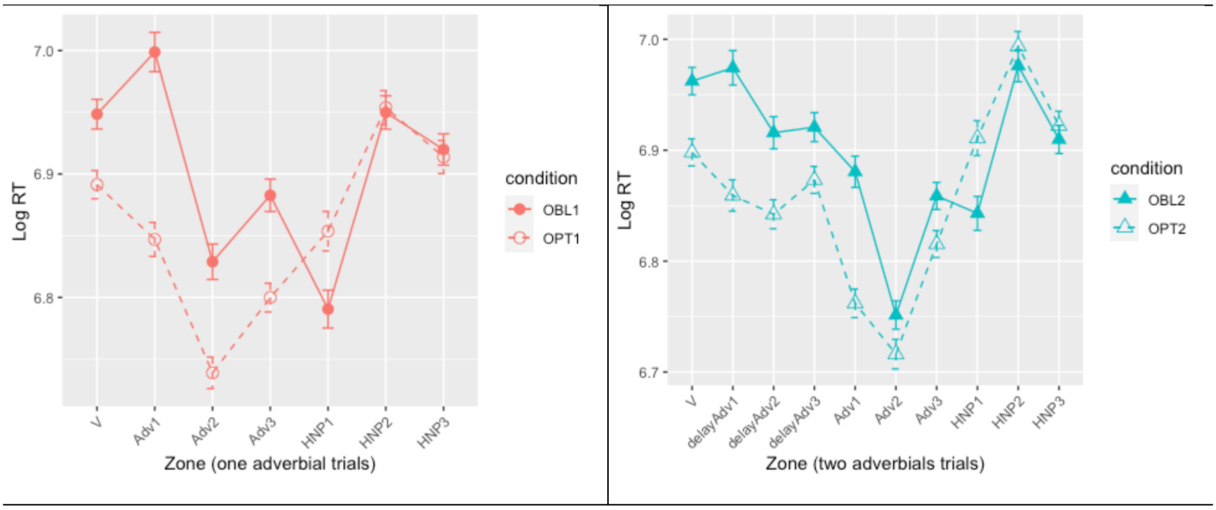


Table 7: Plots for choice times in the one/two adverbial conditions (*obl*: obligatorily transitive verb; *opt*: optionally transitive verb; *1*: one intervening adverbial; *2*: two intervening adverbials)

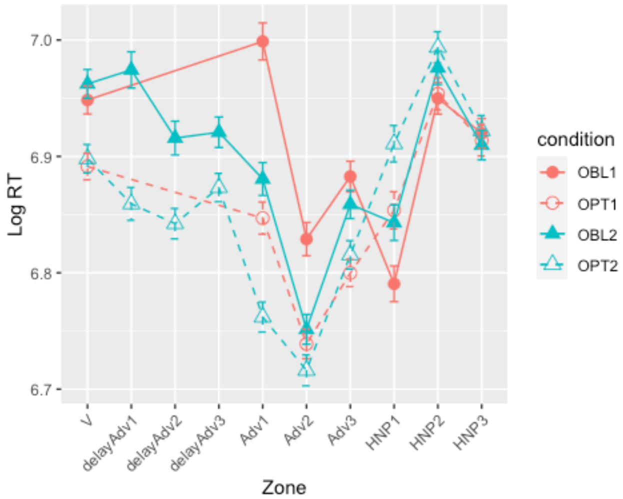


Table 8: Combined choice time plots (*obl*: obligatorily transitive verb; *opt*: optionally transitive verb; *1*: one intervening adverbial; *2*: two intervening adverbials)

Two aspects of the plot for the one-adverbial condition stand out. In the first area of interest (Adv1–Adv3), the the solid red line, which represents data for items with obligatorily transitive verbs, is systematically above the dashed red 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, choice times in the obligatorily transitive condition are slower when the intervening adverbial is parsed, faster at the onset of shifted object and similar after that. The plot for the two-adverbial condition shows a similar pattern. Over the delay adverbial and the adverbial of interest (delayAdv1–delayAdv3 and Adv1–Adv3), the solid teal line, which represents data for items with obligatorily transitive verbs, is systematically above the dashed teal line, which represents data for items with optionally transitive verbs. There is a reversal on the first word of the shifted 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 (version 4.3.1, 2023-06-16, R Core Team 2023): log(RT) ~ OBL_c*Adv_c + (1 + OBL_c | Item) + (1 | ParticipantID). Significance was tested using model reduction. The random effect structure was kept simple to ensure non-degenerate convergence of all models including reduced ones. As before, significance is reported with Boneferroni correction, given that seven separate analyses were conducted (see Table 9).

We found a significant main effect of obligatory transitivity of the verb on reaction times over the adverbial of interest (Adv1, Adv2 and Adv3). In addition, there was a significant main effect of obligatory transitivity on the onset of the shifted 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 adverbials is positive. In other words, reaction times are longer when the verb is obligatorily transitive. On the heavy NP this coefficient turns negative, however, corresponding to shorter reaction times when the verb is obligatorily transtive.

Zone	*term*	*Estimate*	*Std Error*	*t value*	*df*	*uncorrected p*	*corrected significance*
V	(Intercept)	6.929	0.023	301.27	55	0	****
V	OBL_c	-0.076	0.03	-2.55	23.2	0.01771	ns
V	Adv_c	0.016	0.012	1.34	1653.9	0.18019	ns
V	OBL_c:Adv_c	-0.021	0.023	-0.9	1651.2	0.36947	ns
Adv1	(Intercept)	6.889	0.026	261.6	55.4	0	****
Adv1	OBL_c	-0.151	0.023	-6.59	24	0	****
Adv1	Adv_c	-0.107	0.014	-7.42	1629.3	0	****
Adv1	OBL_c:Adv_c	0.055	0.029	1.93	1628.8	0.05402	ns
Adv2	(Intercept)	6.769	0.041	165.87	30.3	0	****
Adv2	OBL_c	-0.065	0.015	-4.27	23.7	2.7e-4	**
Adv2	Adv_c	-0.049	0.012	-4.15	1669.9	4e-5	***
Adv2	OBL_c:Adv_c	0.061	0.024	2.59	1672.8	0.00966	ns
Adv3	(Intercept)	6.845	0.026	264.38	55.3	0	****
Adv3	OBL_c	-0.068	0.023	-2.95	23.5	0.00704	*
Adv3	Adv_c	0	0.012	0.02	1661.2	0.98569	ns
Adv3	OBL_c:Adv_c	0.032	0.023	1.35	1660.9	0.17581	ns
HNP1	(Intercept)	6.862	0.031	219.48	46	0	****
HNP1	OBL_c	0.08	0.023	3.53	24	0.00172	*
HNP1	Adv_c	0.057	0.015	3.75	1598.8	1.8e-4	**
HNP1	OBL_c:Adv_c	-0.026	0.03	-0.85	1601.2	0.39399	ns
HNP2	(Intercept)	6.987	0.034	206.23	35.9	0	****
HNP2	OBL_c	0.017	0.018	0.97	22.2	0.34482	ns
HNP2	Adv_c	0.038	0.013	2.87	1599.9	0.00416	*
HNP2	OBL_c:Adv_c	0.032	0.026	1.2	1603.3	0.2295	ns
HNP3	(Intercept)	6.931	0.03	231.06	38	0	****
HNP3	OBL_c	0.018	0.018	1.01	22.7	0.32407	ns
HNP3	Adv_c	-0.022	0.013	-1.7	1616.6	0.08891	ns
HNP3	OBL_c:Adv_c	0.036	0.026	1.4	1618.7	0.16296	ns

Table 9: Model parameters for fixed effects of obligatoriness and number of adverbials on reactions times at the verb, the adverbial of interest and the heavy object with Bonferroni corrected significance level

Our findings fully replicate Staub et al.'s results conceptually: there is a slowing down effect on the adverbial of interest in the obligatorily transitive conditions relative to the optionally transitive ones. There is a speeding up effect at the beginning of the heavy NP in the obligatorily transitive conditions relative to the optionally transitive ones. There is also a main effect of the number of intervening adverbials but, crucially, no significant interaction, suggesting that the slowdown in the obligatorily transitive conditions should not be attributed to a superficial surprisal effect (i.e., the surprise of not finding an object where one was expected). The fact that the slowdown persists over a delay adverbial is rather more compatible with an explanation based on the LEC. In sum, the maze task confirmed both prediction (14) and prediction (15) (as operationalized in (16)).

We 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 in Staub et al.'s paper did not appear as a significant effect in our SPR experiment but does show up in the maze task.

The LEC also predicts that the verb being obligatorily transitive will affect reading times over 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 SecP1, SecP2, 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.

In this section, as well as in the previous one, we have contrasted LEC effects with the effects of surprisal (at not finding an object where one is expected). The two types of effects differed, we suggested, in that LEC effects persists until a filler is found, while one would expect surprisal effects to subside downstream from the point at which surprisal was triggered (i.e. at the onset of the first adverbial after the verb). Interestingly, our results contain a pattern suggestive of a surprisal effect. This pattern can be found not in the reaction times but in the error rates (a type of data that a maze experiment can make available, unlike a self-paced reading experiment).

Our error rate analysis included all data from the four zones of interest (verb, delay adverbial, adverbial of interest and onset of the heavy NP) from all 113 participants after an initial filter was applied: 674 datapoints were excluded as outliers because log transformed reaction times deviated more than 2.5sd from the mean. This left 21,413 datapoints for analysis. Plots showing error rates in the zones of interest are given in table 10 (for the one-adverbial and two-adverbial conditions separately) and table 11 (combined).

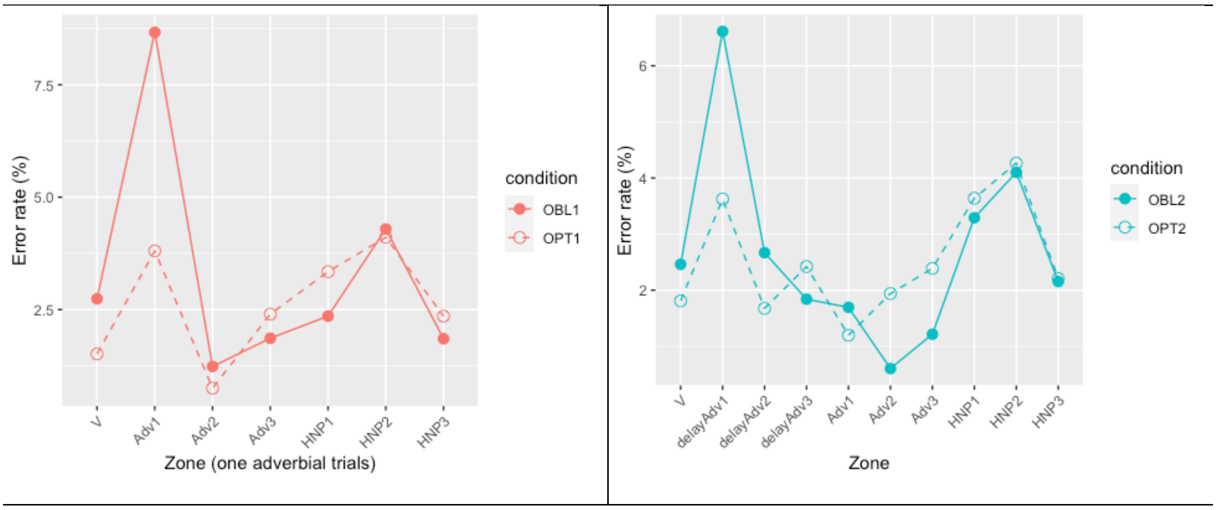


Table 10: Plots for error rates in the one/two adverbial conditions (*obl*: obligatorily transitive verb; *opt*: optionally transitive verb; *1*: one intervening adverbial; *2*: two intervening adverbials)

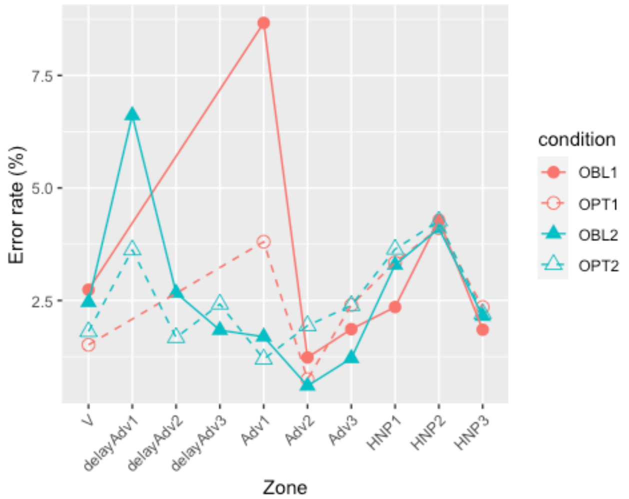


Table 11: Combined error rate plots (*obl*: obligatorily transitive verb; *opt*: optionally transitive verb; *1*: one intervening adverbial; *2*: two intervening adverbials)

In our statistical analysis we excluded data for the delay adverbial, as these were only available in the two-adverbial condition. The conditions of interest were contrast coded. Error rates were analyzed by conducting 7 separate logistic regressions using the R's lme4 package (Bates et al. 2015: Correct ~ OBL_c*Adv_c + (1 | Item) + (1 | ParticipantID). A reduced random effect structure was used to ensure convergence of the models. The results of these analyses are summarized in table 12 below. Because 7 independent analyses were conducted, significance was evaluated using Bonferroni correction.

As can be seen, the only significant effect that we found is a main effect of the number of adverbials on Adv1. This effect reflects strongly elevated error levels at Adv1 in the one-adverbial condition, independent of the obligatory or non-obligatory transitivity of the verb. It corresponds to the large red and blue spikes on Adv1 in the left graph in table 10.

The fact that no other significant effects presented themselves, and in particular the lack of any spike in error rates over the downstream adverbial in the two-adverbial condition, suggests that the peak in error rates after the verb is a consequence of surprisal. After all, one would expect surprisal to wear off as the moment of surprisal recedes.

In addition, the peak in error rates was present whether the verb was obligatorily or optionally transitive (there is no significant interaction with obligatory transitivity). This again suggests that we are dealing with a consequence of surprisal. In Staub et al's first experiment, the obligatorily transitive verbs were of course verbs used transitively one hundred percent of the time. However, the optionally transitive verbs also had a high propensity for transitive use. The mean transitivity biases of the optionally transitive verbs ranged from .66 to .93, with an overall mean of .80 (SD = .07) (Staub et al. 2006: 393). This was not a coincidence: Staub et al. wanted to show that the parser does not posit an empty object position unless it is forced to, even if statistical tendencies favor the presence of an object. As we based our experimental items on Staub et al.'s, we inherited this transitivity bias. Now, it stands to reason that surprisal has a statistical basis, as it is the defeat of expectations that are statistical in nature. Therefore, the suggestion that the post-verbal peak in error rates is a consequence of suprisal is compatible with the fact that the effect shows no sensitivity in our analysis to whether the verb is obligatorily of optionally transitive –

though the shape of the curve suggests that stronger transitivity bias produces more errors, as expected under this interpretation of the data.

Zone	*term*	*Estimate*	*Std Error*	*statistic*	*uncorrected p*	*corrected significance.*
V	(Intercept)	4.811	0.366	13.13	0	****
V	OBL_c	0.465	0.29	1.6	0.10866596	ns
V	Adv_c	-0.04	0.288	-0.14	0.88990096	ns
V	OBL_c:Adv_c	-0.393	0.59	-0.67	0.50489356	ns
Adv1	(Intercept)	4.121	0.29	14.21	0	****
Adv1	OBL_c	0.666	0.273	2.44	0.01470479	ns
Adv1	Adv_c	1.545	0.274	5.65	2e-8	****
Adv1	OBL_c:Adv_c	-0.586	0.543	-1.08	0.28077336	ns
Adv2	(Intercept)	5.566	0.502	11.09	0	****
Adv2	OBL_c	-0.352	0.411	-0.86	0.39218385	ns
Adv2	Adv_c	-0.128	0.411	-0.31	0.75571995	ns
Adv2	OBL_c:Adv_c	-1.739	0.822	-2.12	0.03442681	ns
Adv3	(Intercept)	4.303	0.264	16.29	0	****
Adv3	OBL_c	-0.473	0.294	-1.61	0.1078246	ns
Adv3	Adv_c	0.22	0.294	0.75	0.45350716	ns
Adv3	OBL_c:Adv_c	-0.429	0.587	-0.73	0.46488944	ns
HNP1	(Intercept)	4.387	0.313	14.01	0	****
HNP1	OBL_c	-0.24	0.246	-0.98	0.32918631	ns
HNP1	Adv_c	-0.259	0.248	-1.04	0.29616418	ns
HNP1	OBL_c:Adv_c	0.374	0.492	0.76	0.44739759	ns
HNP2	(Intercept)	3.691	0.241	15.33	0	****
HNP2	OBL_c	0.015	0.208	0.07	0.94148988	ns
HNP2	Adv_c	-0.013	0.207	-0.06	0.95076003	ns
HNP2	OBL_c:Adv_c	-0.091	0.415	-0.22	0.82557067	ns
HNP3	(Intercept)	4.338	0.299	14.51	0	****
HNP3	OBL_c	-0.12	0.279	-0.43	0.66561207	ns
HNP3	Adv_c	-0.041	0.278	-0.15	0.88417382	ns
HNP3	OBL_c:Adv_c	0.235	0.555	0.42	0.67250456	ns

Table 12: Model parameters for error rates with adjusted significance levels at the verb, the adverbial of interested and the heavy object

The contrasting profiles of the error rates (which decrease over time) and of the slowdown in reaction times (which persist over the intervening adverbials) suggests that the maze experiment allows us to tease apart the consequences of surprisal and the consequences of violations of the LEC. This naturally implies that the latter cannot be reduced to the former.

We close this section with a brief investigation of the parser's reaction to LEC violations. In the introduction, we claimed that the parser encounters difficulties when a parse tree violates the LEC. An empty branch represents the prediction that a category will be encountered that can replace **e**, and pushing **e** away from the leading edge of the parse tree is tantamount to a defeat of that prediction. However, as the prediction cannot be withdrawn (given that the verb is obligatorily transitive), it must remain active in further processing, giving rise to a persistent effect that we have observed. Interestingly, the slowdown caused by LEC violations is not constant across the words making up the intervening adverbials.

Table 13 gives plots of the reaction time slowdown across the areas of interest in both experiments. The absolute values are not of particular interest (and incomparable across the plots).

However, it is easy to discern that the slowdown is greatest on the first word of the adverbial of interest and then decreases substantially on the remaining two words. This is true whether the adverbial immediately follows the verb (one adverbial) or further downstream (two adverbials). The pattern repeats itself on the immediately postverbal adverbial in the two adverbial condition.

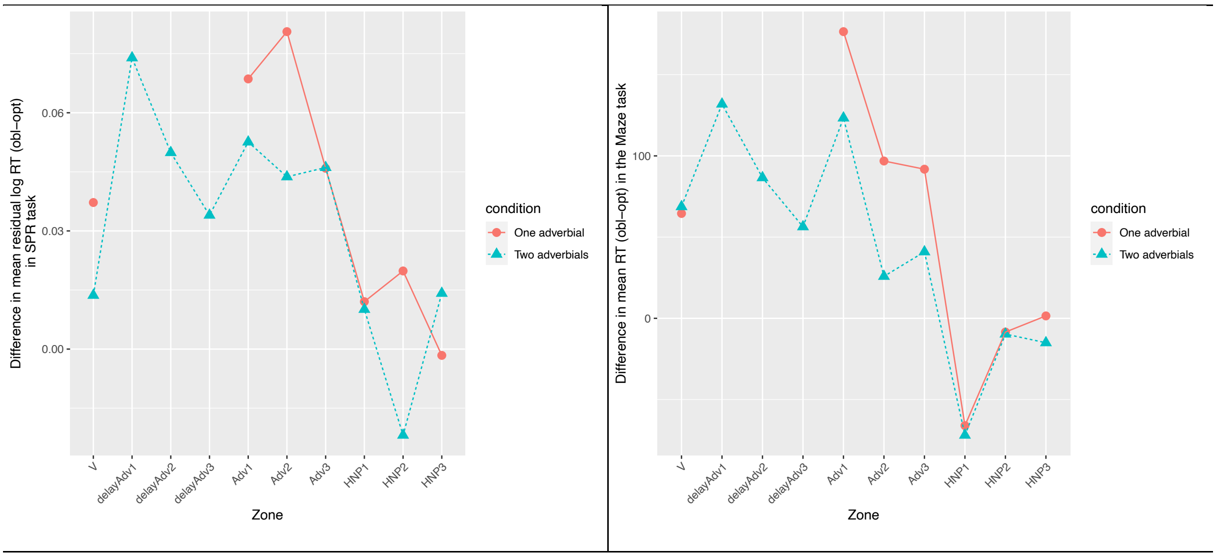


Table 13: Plots of reaction time differences in the obligatorily versus optionally transitive conditions in the self-paced reading experiment (left panel, differences of mean residual log RTs) and the maze task (right panel, difference of mean raw RTs).

In sum, there is a strong effect at the left edge of the postverbal adverbial and again at the left edge of the second adverbial. We can understand this pattern as resulting from the LEC. The parser expects to find a category to fill the postulated empty node, and such a category could in principle appear immediately after the verb or immediately after the first adverbial. There is no comparably strong effect in the middle of the adverbials. This, too, makes sense: once the parser has committed to parsing the adverbial, there is no grammatical opportunity for a filler to appear. What we see at the onset of the adverbials, then, is a 'missing filler effect': the consequence of a strongly predicted but missing filler. This missing filler effect is to rightward movement as the filled gap effect is to leftward movement.

Indeed, the pattern of slowdown in Table 13 is strongly reminiscent of the filled gap effect found with leftward movement. Forster et al. (2009), in a paper validating the Maze task, compared object and subject relatives (as in (19)). In two experiments they replicated the well-known finding that object relatives are harder to parse than subject relatives. The two experiments differed in the nature of the distractors, which were existing words in experiment 1 and nonce words in experiment 2. A detailed analysis of the time course of the difficulties connected with object relatives revealed that in both experiments this difficulty was 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 is a significant difference for the article, no significant effect for the verb, and a minor effect for the noun, which was not significant other than in the analysis by participant.

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

The missing filler effect with its notable sawtooth pattern reflects the parser's search for a filler. The sawtooth pattern is neither a short-lived reflection of surprisal at not finding an object after an obligatorily transitive verb, nor can it be explained as the cost of keeping the verb's selectional properties active in working memory. Neither hypothesis explains the slowdown at the onset of the second adverbial in the two adverbials condition.

Thus, we can distinguish three 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 the marked slowdown in the obligatorily transitive conditions at the beginning of each adverbial, reflecting violations of the LEC. (iii) There is the speed-up in the obligatorily transitive conditions at the left edge of the heavy NP, reflecting increased processing load in the optionally transitive conditions. The reduced slowdown over the second and third word of the adverbials in the obligatorily transitive conditions could be understood as resulting from the maintenance costs associated with the incomplete rightward dependency.

3. A more detailed model of the parser

In the introduction, we showed that three assumptions about parsing – Incrementality, a filler-driven approach to movement and the LEC – generate the prediction that rightward movement of obligatory material causes difficulties in processing. This prediction was tested and confirmed in section 2. With this much in place, we now develop a more detailed model of the parser and and discuss how it applies to heavy-XP.

We assume that the output of the parser is a tree and that the theory of parsing should therefore describe the process through which 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 fact that human parsing is as fast as it is cannot be understood unless the process of tree construction is incremental and monotonic, as defined below.

(20) a. *Incrementality*: The parser assigns each input symbol a structural position as soon as it comes in (Frazier and Rayner 1982; Gorrell 1995; Just and Carpenter 1992).
b. *Monotonicity*: No commitment present at stage *n* of the parsing process may be abandoned at stage *n*+1 (Marcus, Hindle and Fleck 1983).

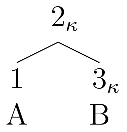
The effect of Incrementality is that when a new input symbol comes in, the parser must integrate it into the existing tree immediately, without having access to its right context. Monotonicity concerns properties of the parse tree. Once such properties are posited, any further development of the parse tree must be faithful to them.

If the output of the parser is indeed a tree, then the input string will be structured using three basic relations: domination, precedence and headedness. We opt for domination rather than immediate domination as the relation that the parser works with because it allows a certain degree of flexibility that, as we will see, facilitates incremental parsing of incoming material.[2] For concreteness' sake, we assume that the parser fixes precedence relations between any two nodes that

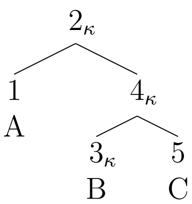
[2] If required, immediate dominance can be defined in terms of dominance.

are not in a dominance relation.[3] The notion of headedness we will use is the one employed in Grimshaw's theory of extended projection (Grimshaw 1991, 2005). This implies that for any branching non-terminal node the parser must encode whether it is part of the same extended projection as its left or its right daughter (a constraint we refer to as Endocentricity). We assume that every subtree must be headed at every stage of the parse, so that every non-terminal posited is a member of the extended projection of one of its daughters.

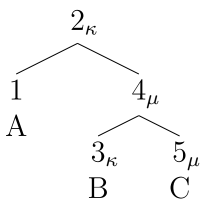
We use the following notational conventions: (i) Greek subscripts indicate nodes belonging to the same extended projection; (ii) nodes are numbered; (iii) lexical material is represented as capitals below terminal nodes. The following tree therefore encodes (i) that A precedes B; (ii) that A and B have a common ancestor; and (iii) that this ancestor belongs to the extended projection of B.

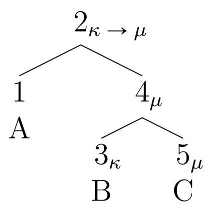
(21)


To illustrate how the system works, we consider two extensions of (21) based on an incoming additional terminal. As the parser fixes dominance but not immediate dominance relations, introduction of nodes between a mother and its daughter is information-preserving as long as headedness remains unaffected. Thus, (22) is a possible extension of (21).

(22)


The flexibility of the dominance relation does not imply that all tree extensions are allowed. In integrating C, it is impossible to insert a node between the root and B that is part of a different extended projection, as in (23a). The resulting structure violates Endocentricity, because node 2 does not have a head (it is not coindexed with either of its daughters). This cannot be repaired by changing the index on 2 to μ, as in (23b), as that violates Monotonicity (violations of parsing principles are marked by the skull-and-crossbones symbol).

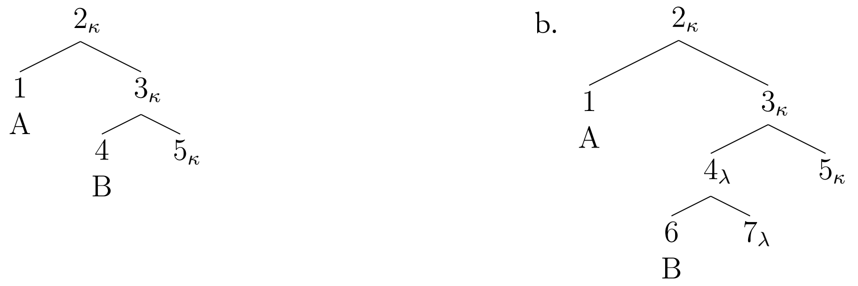
(23) a. ☠


b. ☠


[3] Given the No-Tangling Constraint, there are more compact ways of fixing precedence relations in a tree. However, there is a tradeoff between the memory requirements induced by exhaustive listing and the computational requirements of deriving precedence statements from the more compact representation when needed. The question of which representation is actually used is an empirical one that does not affect our argumentation.

Under certain circumstances, the parser may be forced to posit empty nodes. Suppose that neither A nor B can appear in the other's extended projection. In that case, an initial input A-B must be parsed as in (24a) or (24b), where 5 and 7 are empty nodes (instantiations of ***e***).

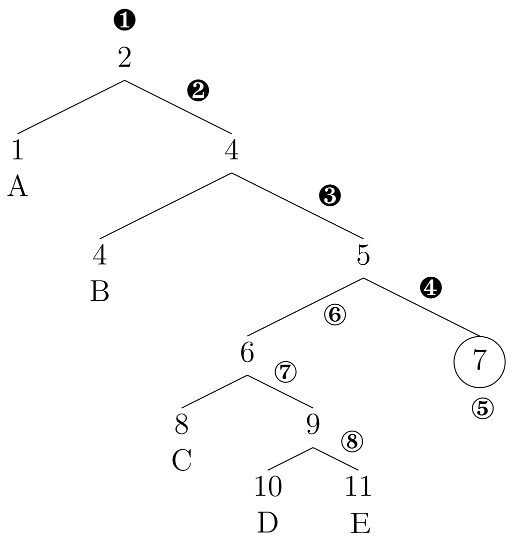
(24) a. b.



Empty nodes are predictions. They commit the parser to the existence of material that it has not encountered yet and whose content is therefore still undetermined. Hence, empty nodes should restrict the parser's future actions. This is exactly what the LEC guarantees: it requires that empty nodes be located at the right edge of the parse tree.

The benefits of the LEC can be illustrated with the tree in (25), where node 7 is empty. If we try to integrate an incoming category F into this tree, there are eight possible attachment sites: the first at the root of the tree (❶), the next three separating nodes at the leading edge (❷,❸ and ❹), the fifth under the empty node (⑤) and the last three separating nodes below the empty node (⑥, ⑦ and ⑧). If the LEC holds, only the lower four attachments sites are licit: attachment at any of the four higher sites would push the empty node away from the right edge of the tree. If the LEC does not hold, however, attachment is possible at all eight sites.

(25)



The LEC thus limits which continuations of the current parse can be considered. However, this benefit does not come for free, as restrictions imposed by empty nodes may also create difficulties for the integration of material in the parser's right context. We therefore assume that the parser adopts a conservative strategy, only generating empty nodes when forced to do so.

We should make clear that an empty node (a prediction) must be distinguished from a node that contains a category without phonological content (such as a null pronoun). Silent categories are not subject to the LEC and may therefore be found in positions other than at the right edge of the current parse. Moreover, while empty nodes may (and ultimately must) be filled by incoming material, silent categories are like other lexical material in that they cannot be overwritten without violating Monotonicity.

Finally, we discuss how movement dependencies are parsed. As mentioned in the introduction, we assume that traces are (partial) copies of the moved category. This implies that it is not possible for the parser to insert a trace before it has identified an antecedent. In line with this, we propose the following filler-driven procedure for establishing a movement dependency in parsing:

(26) a. Identify the moved category.
b. Buffer it (that is, suspend work on it and store it).
c. Insert a copy.
d. Continue work on the constituent (if need be).

Steps (26a-c) and the order in which they come are dictated by the copy theory. Step (26d) and the instruction in step (26b) that work on the constituent be suspended are necessary because moved categories may contain elements that need to enter into a syntactic dependency with material located in the parser's right context (as in structures that involve binding under reconstruction or remnant movement).

Several considerations favor a filler-driven strategy. To begin with, given that traces have no phonological content, there is no evidence for them in the input string. On the gap driven strategy, the parsing of movement dependencies must therefore rely on indirect evidence for gaps, which could come from selectional requirements of material surrounding the gap (such as a verb selecting an object). However, given that the parser must be conservative in postulating empty categories, a gap-driven strategy is not compatible with adjunct movement or movement of optional arguments.

Moreover, there are many circumstances in which the parser postulates an empty node but should not start a search for a filler. For example, a determiner in English triggers the prediction that a noun will follow (see (27)). However, it would be counterproductive if the empty branch that caches out this prediction would force the parser to start looking for a moved noun phrase. Thus, for the gap-driven strategy to work efficiently, the parser would need additional heuristics about which empty nodes trigger a search for an antecedent and which do not.

(27) 2_κ

1 the — 3_κ

Phillips and Wagers (2007) give an overview of the substantial empirical support for the filler-driven strategy. They point to the availability of supporting data from a variety of languages (including Dutch, Russian, Hungarian, Italian, German, and Japanese) and a variety of experimental paradigms (including reading time measurements, speeded grammaticality judgments, event related potentials, plausibility measures in eye tracking or self-paced reading, the 'stop making sense' task, cross-modal lexical priming, and head-mounted eye tracking). These include the filled-gap effect briefly discussed towards the end of the previous section.

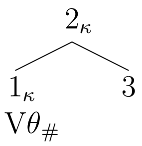
This concludes our exploration of properties of the parser as relevant for movement. Most of what we have said is uncontroversial. The main assumptions driving the analyses in the following sections are the LEC, which is novel but seems to us to follow from good parser design, and the restriction that insertion of a trace is possible only after the identification of an antecedent, which is a consequence of the copy theory of movement.

4. Parsing heavy-XP shift

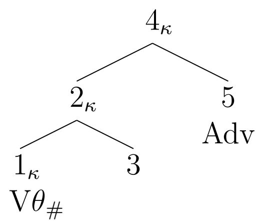
We can now be more precise about why heavy-XP shift of the object is sensitive to whether or not the verb is obligatorily transitive (as shown in section 2). As examples, we will use (28a) and (28b), starting with the former (where *devour* is obligatorily transitive and *eat* optionally transitive). The processing of these examples was already sketched in section 1.

(28) a. John devoured t_1 yesterday [the food that his brother prepared]$_1$
b. John ate t_1 yesterday [the food that his brother prepared]$_1$

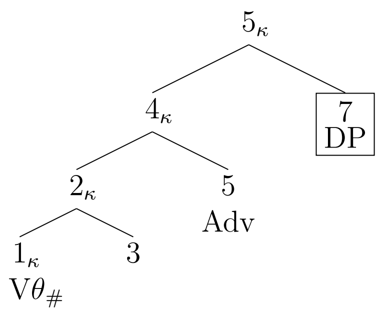
The relevant substring in (28a) is *devoured yesterday the food that his brother prepared*. Since *devour* is obligatorily transitive, the parser predicts an object once it encounters the verb. This prediction must be reified as an empty node at the leading edge of the parse tree, as in (29). (We use the #-symbol to indicate a theta role satisfied by the sister node of a given theta-role assigner.)

(29)


The next word the parser encounters is the adverbial *yesterday*. Since this adverbial cannot fill the complement position of the verb *devour*, a new position must be created for it. In (30), it is right-adjoined to the VP. Crucially, the filler for the empty node has not been encountered yet, while at the same time the empty node is pushed away from the leading edge. This means that at this stage the parser violates the LEC (which explains the parsing difficulties discussed in section 2).

(30)


The material following the adverbial forms the onset of a DP that is integrated into the VP as a second right-adjoined constituent. Once a position for this DP has been created, the parser can hypothesize that it has moved, buffer it and start the search for a position in which to insert a trace. Up to that point, the parse tree still violates the LEC:

(31)


A scan of the parser's left context will easily identify an insertion site for the trace of the shifted object, namely the empty node adjacent to the verb. The resolution of the movement dependency will be quick, given that the insertion site has been prefabricated and is likely to be highly salient in view of its violating the LEC.

(32)
$$[_{5_\kappa}\ [_{4_\kappa}\ [_{2_\kappa}\ [_{1_\kappa}\ \text{V}\theta_{\#}]\ [\cancel{3/7}\ \cancel{\text{DP}}]]\ [_{5}\ \text{Adv}]]\ [_{7}\ \text{DP}]]$$

In sum, it follows that heavy-XP shift of obligatory objects leads to difficulties in the parsing of material between the verb and the shifted DP and a speeding up once the DP has been encountered.

We now turn to the parsing of the substring *ate yesterday the food that his brother prepared* in (28b), which differs from our earlier example in the verb being optionally transitive. In order to understand the consequences of this, we need to be more explicit about the difference between obligatorily transitive, optionally transitive, and intransitive verbs. We assume that obligatorily transitive verbs must have an internal theta role, intransitive verbs cannot have an internal theta role, and optionally transitive verbs tolerate having an internal theta role. This implies that optionally transitive verbs should not be seen as lexically ambiguous between an obligatorily transitive and an intransitive variant. The way this plays out in parsing is as follows. An obligatorily transitive verbs will have a theta role at any stage of the parsing process. Optionally transitive verbs by contrast start out without an internal theta role, but may acquire one if this is necessary to accommodate incoming material. Our notation for this tolerance of an internal theta role is '(θ).' We are aware that this notation is usually used to abbreviate two separate rules or structures, but this is not how it is meant here.

As a consequence, the parser need not and will not postulate an empty node in post-verbal position when it encounters an optionally transitive verb:

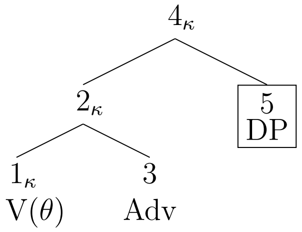
(33)
$$[_{1_\kappa}\ \text{V}(\theta)]$$

The incoming adverbial can therefore be accommodated without triggering a violation of the LEC, as in the tree below.

(34)
$$[_{2_\kappa}\ [_{1_\kappa}\ \text{V}(\theta)]\ [_{3}\ \text{Adv}]]$$

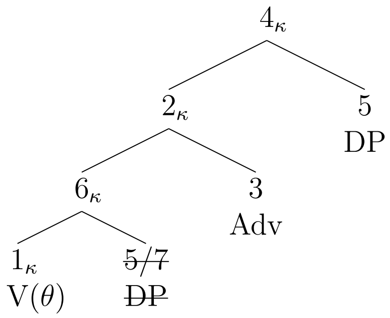
When the parser next encounters the shifted DP, it will identify it as a moved category and start the search for a position to insert a trace. Of course, the tree in (35) does not yet contain such a position, but the assumptions that we have made about tree growth allow the creation of an inserrion site that is sister to the verb.

(35)



Creation of this position involves postulation of nodes 6 and 7 in (36). This does not clash with any of the statements about dominance, precedence, or headedness that characterize (35). The order of verb, adverb, and DP remains the same. Node 2 still dominates nodes 1 and 3, and node 1 is still the head of node 2. As soon as the empty position is created, it will be filled by a copy of DP, so that at no stage of the process the LEC is violated.

(36)



We thus have an explanation for the fact that heavy-XP shift of optional arguments does not cause difficulties during the parsing of material that separates the verb and its object, in contrast to heavy-XP shift of obligatory arguments. In addition, we can understand why integration of the filler in (35) and (36) should be slower than integration of the filler in the obligatory condition in (31) and (32). The absence of a pre-fabricated gap means two things. First, the search for a gap will not be aided by the existence of an unfilled position. Second, the relevant position must be newly created.

Note that our our proposal explains why the effects uncovered by Staub et al. are categorical in nature, rather than sensitive to the propensity of the verb to take an object. This is because the conservativity required to avoid unnecessary violations of the LEC disfavors postulation of a gap in V-Adv strings where there is a choice. Therefore, as long as the verb does not obligatorily select for an object, the only available parse will not feature an empty node. If the verb is obligatorily transitive, an empty node must of course be posited (with the consequences decribed above).

5. Universal 20

We now turn to Universal 20 to consider whether the linear asymmetry it describes can be explained in terms of the LEC.

5.1 The pattern

Recall that in the introduction we considered four possible word orders in the noun phrase (see (7)). According to Greenberg's Universal 20, three of these are attested and one is not. We showed that this asymmetry can be explained using a version of phrase-structure theory that allows variation in the linearization of sister nodes and a movement rule that is uniformly leftward. Of course, the number of possible word orders of four elements (demonstrative, numeral, adjective, and noun) is not four but 24. We therefore need to show how the proposed account scales up.

As mentioned, Universal 20 is a description of attested and non-attested neutral orders in the extended nominal projection. The definition of neutral orders is a matter of debate, but the three-step decision procedure in (37) summarizes common practice in the field (see Dryer 2007 and Abels 2016 for discussion).[4]

(37) A given order as neutral if it is
  i. the only grammatical order, or
  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.

The complete paradigm of all attested and nonattested neutral orders of demonstrative, numeral, adjective and noun was described in Cinque 2005 and supported by additional typological evidence in Cinque 2014. It is characterized by the basic asymmetry originally observed by Greenberg: the order of elements before the noun is constant, but there is variation following the noun. Table 14 shows the pattern (unattested orders are shaded; all unshaded orders are attested)

	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

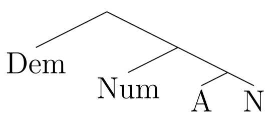
Table 14: Attested and unattested orders in the noun phrase according to Cinque 2005, 2013

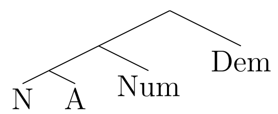
For further discussion of Universal 20 pattern, see section 5.4.

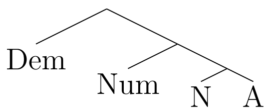
Abels and Neeleman (2012) propose a reinterpretation of Cinque's (2005) antisymmetric analysis of the data in Table 14, demonstrating that the typological pattern follows directly from a grammar that allows variation in the linearization of sister nodes but requires movement to be leftward. With regard to the structure of the noun phrase, all that needs to be assumed is that there is a universal hierarchy such that adjectives are attached lower than numerals and numerals are attached lower than demonstratives:

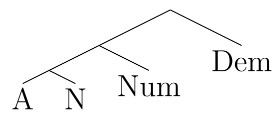
(38) Dem > Num > A > N

If languages can choose different linearizations for the resulting structure, this allows for the following representations, which all yield orders that are attested (they correspond with the orders in columns I and II in table 14):

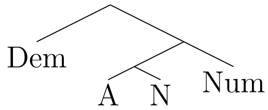
(39) Ia.


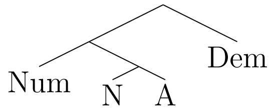
IIa.


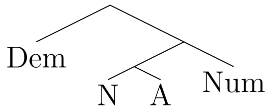
Ib.


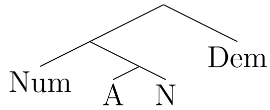
IIb.


[4] Steedman (2020) proposes an analysis that rules out only two of the twenty-four logically possible noun phrase orders. However, as his analysis is not restricted to neutral orders, a comparison with other work in this domain is far from trivial.

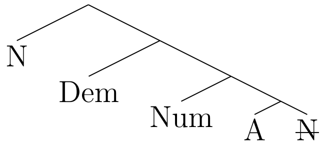
Ic.


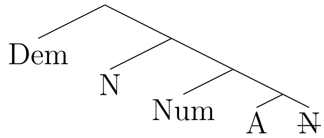
IIc.


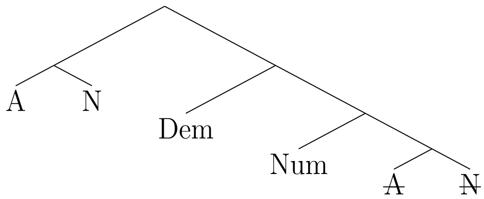
Id.


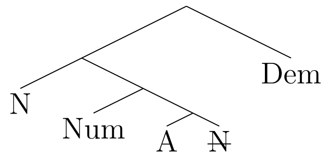
IId.


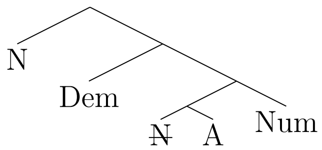
What needs to be added to capture the full paradigm is the hypothesis that neutral word orders can be derived by leftward movement of the noun or a constituent containing noun and adjective. This allows the additional representations in (40). (It is an open question why the movements that derive neutral word orders should be restricted to constituents containing the noun. For discussion, see Cinque 2005, Georgi and Müller 2010 and Steddy and Samek-Lodovici 2011.)

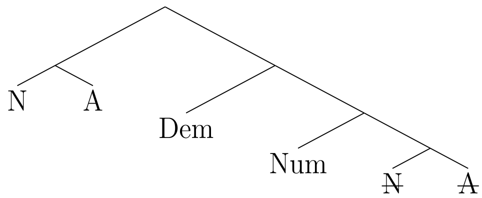
(40) IIIa.


IIIb.


IIIc.


IIId.


IIIe.


IIIf.


We next consider why the unattested orders are underivable. The constraint that movements deriving neutral word order must be leftward is crucial in answering this question. It implies that none of the trees in (40) have an attested mirror image, which is in line with observations. In table 14, the relevant mirrored orders (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 (38), 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 N-A order (see (41a)). Similarly, given that numerals must be adjacent in any base-generated structure to the substring comprising the adjective and the noun, the order in (41b) is excluded. As a consequence, all the orders in (IVa-f) are ruled out, as are the orders in (Ie,f).

(41) 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 base structures for (IIe) and (IIf) must have been either Num-Dem-A-N or Num-Dem-N-A. Neither of these base structures is grammatical, however, as they are not compatible with the hierarchy in (38) (compare (IVc,f)).

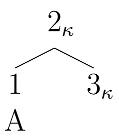
In conclusion, the constraint that movements deriving neutral order cannot be rightward is instrumental in explaining the full pattern of attestated and nonattested orders in the extended nominal projection. The crucial question that presents itself at this point is *why* movement of the noun or a constituent containing the noun must be leftward, especially in view of the fact that languages differ with respect to the linearization of sister nodes. Our first step in answering this question is to demonstrate that rightward movement of (a constituent containing) the noun necessarily leads to a violation of the LEC.

5.2 Parsing movement of (constituents containing) the noun

In order to illustrate the way our parser deals with leftward and rightward movement in the noun phrase, we consider two orders: A-N-Dem-Num for leftward movement and its unattested mirror image Num-Dem-N-A for rightward movement. The difficulties that the rightward movement derivation faces are not limited to this string, but generalize to all potential rightward movements in the extended nominal projection.

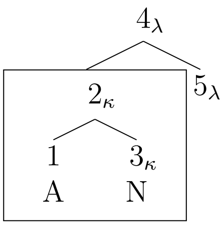
The first step in parsing the A-N-Dem-Num order involves providing a position for the adjective. Once the parser encounters the adjective, it predicts a noun and hence an empty node is generated at the leading edge of the tree, as in (42), in concurrence with the LEC.

(42)



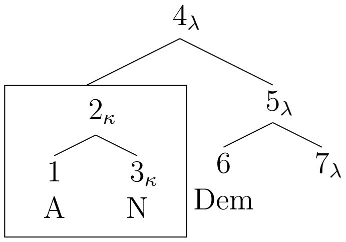
At the next step, N inserted in the empty node. On the assumption that the structure of the extended nominal projection always unfolds completely, irrespective of what modifiers are present, [A N] can be identified as a filler at this stage, which implies that it is buffered and that a trace is predicted. This prediction is instantiated by a new empty node at the leading edge of the tree in (43). (If the structure of the extended nominal projections unfolds only to accommodate modifiers actually present, the identification of [A N] as a filler would depend on the presence of the demonstrative. In this case, step (43) would be skipped.)

(43)

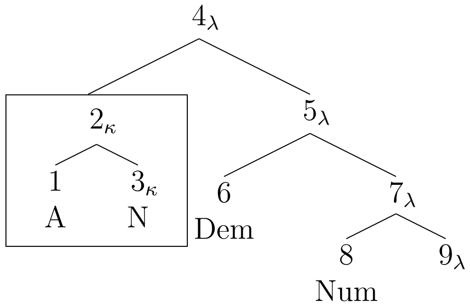


Subsequently, the demonstrative and the numeral are integrated, as in (44a,b), preserving the prediction that a trace is to follow.

(44) a.

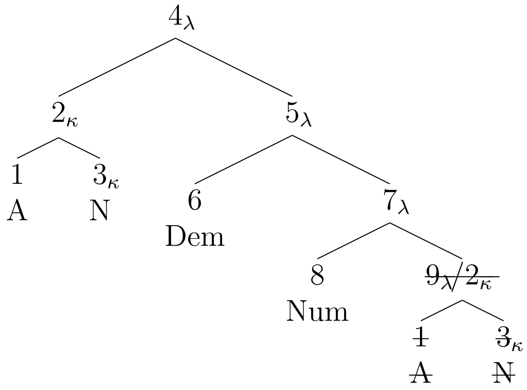


b.



Finally, a trace (a copy of [A N]) is inserted for the empty node, leading to a full and successful parse of the structure, as in (45).[5]

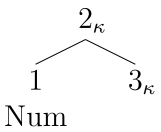
(45)



In sum, leftward movement of the [A N] constituent in the A-N-Dem-Num order is unproblematic. We next consider the mirror image of this order, Num-Dem-N-A, which is unattested and would have to be derived by rightward movement of the [N A] constituent. We will be careful to consider the best possible sequences of parsing steps for this order, but even so, a violation of the LEC (or of Monotonicity) is unavoidable.

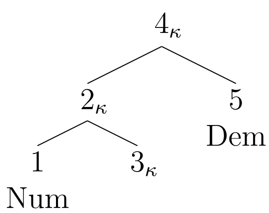
The first category that the parser encounters is the numeral. On the assumption that the numeral is licensed only in the extended projection of a noun, an empty node is predicted, which must at this stage be placed at the leading edge of the parse tree, as in (46) (see below for discussion of alternative assumptions).

(46)



The next element in the input string is the demonstrative, which, like the the numeral, must be part of the extended projection of a noun. It must be integrated in a position higher than the numeral given the universal hierarchy in (38). The only way to satisfy this requirement while respecting the order of elements in the input is to push the empty node away from the right edge of the parse tree, in contravention of the LEC:

(47) ☠



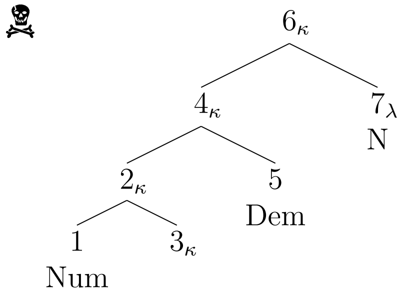
The next element the parser encounters is the noun, which can be integrated straightforwardly, as in (48a). By itself, integration of the noun does not remove the violation of the LEC. However, the position of the noun in (48a) is higher than the highest modifier in the noun's extended

---

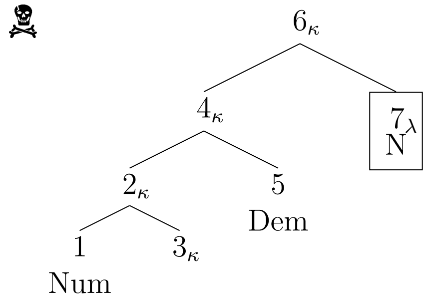
[5] In principle, debuffering of the [A N] constituent could already have happened at stage (44a) with the trace inserted in empty node 7. Subsequent integration of the numeral would then create a node between 5 and 7 (with the numeral attached as a left or right daughter of that node without an effect on word order). This possibility does not affect our argumentation below. Which sequence of parsing steps is correct is an empirical issue that could be explored by considering the point at which reactivation of the buffered constituent occurs.

projection. This implies that N must be identified as having moved from a position below this modifier, as in (48b).

(48) a.

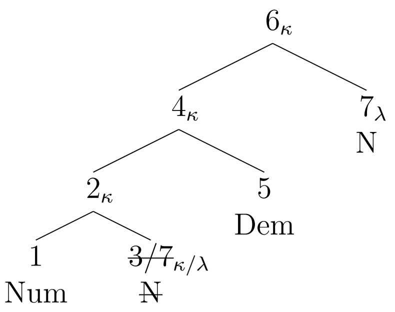


b.



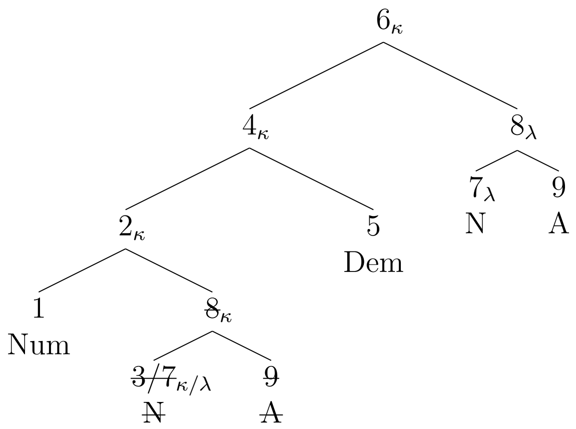
Once the noun has been identified as a filler, the parser initiates a search for a possible insertion site, which is of course present in the form of empty node 3. The movement dependency can thus be resolved, so that compliance with the LEC is re-established:

(49)



Finally, the parser encounters the adjective, which, on the hypothesized sequence of parsing steps, would have to be integrated in the filler and the trace, as in (50). This could happen simultaneously or sequentially.

(50)

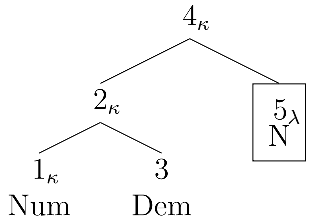


Thus, rightward movement of the noun-adjective combination in the Num-Dem-N-A order triggers a violation of the LEC. This violation has no counterpart in the sequence of parsing steps needed to analyze *leftward* movement in the A-N-Dem-Num order.

There is a crucial assumption in the derivation of this result, namely that the numeral must be licensed in the extended projection of the noun. This forces the parser to posit an empty node (node *3* in (46)) when encountering a numeral. It is important to consider what happens if we do not make that assumption, so that the parser need not (and therefore will not) postulate an empty node after the numeral. This is possible if we allow the numeral to act as the head of an extended nominal projection.[6]

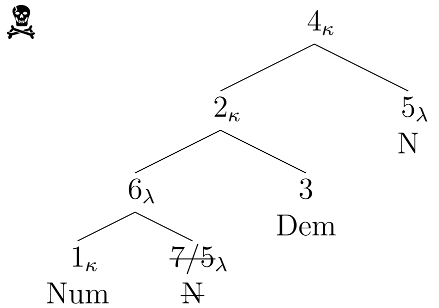
After encountering the the demonstrative and the noun, the parser can hypothesize that the noun has moved, ending up with the representation in (51).

(51)

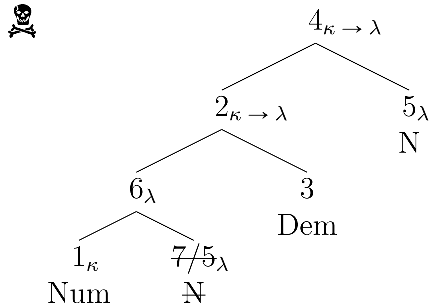


Resolution of the movement dependency leads to problems. As the noun is to be the head of the extended projection, it must percolate its index, but the nodes dominating the demonstrative and the noun are already labelled as projections of the numeral. Leaving the subscripts originally assigned to nodes 2 and 4 intact, as in (52a), leads to a violation of Endocentricity (in particular, node 2 does not have daughter from which it inherits its subscript). The Endocentricity violation can be repaired by assigning nodes 2 and 4 a new subscript that correctly identifies the noun as the lexical head of the structure, as in (52b). However, this repair requires retraction of information and therefore violates Monotonicity:

(52) a.



b.



In conclusion, if we design a parsing sequence for the Num-Dem-N-A order that does not violate the LEC, we end up violating either Endocentricity or Monotonicity or both (for related discussion, see Ackema and Neeleman 2002). We demonstrated this on the assumption that the numeral can project, but it is easy to see that exactly the same problems arise if we allow the demonstrative to project. Our overall conclusion is therefore that while leftward movement of (a constituent containing) the noun within the extended nominal projection is unproblematic, rightward movement necessarily leads to parsing difficulties.

Although we have illustrated the reasoning behind this conclusion using only two word orders, it is clear that the result is general and robust. In the case of rightward movement, difficulties will arise no matter what modifier the parser encounters first, no matter whether it is the noun or a constituent containing the noun that moves, and no matter what material is crossed.

There is one type of rightward movement that does not run into parsing difficulties. Such difficulties arise where the moving constituent crosses other material. As long as no material is crossed, that is, as long as the movement is string-vacuous, any postulated empty node will

[6] We do not think that this is actually possible. The contexts in which a demonstrative and a numeral form a constituent in the absence of a noun are typically ones that would license an elided noun. The sentence *these three arrived late*, for instance, requires a context that provides an antecedent for the missing noun.

remain at the leading edge of the parse tree. Whether string-vacuous rightward head movement exists is an empirical issue that we will not pursue here.

5.3 Synchronic and diachronic consequences of the LEC

We have argued above that rightward movement of obligatory categories creates parsing difficulties, because such movement necessarily leads to a violation of the LEC. We have seen two instances of this. First, experimental findings reported in Staub et al. 2006 and replicated and extended in section 2 show that heavy-XP shift of an obligatory object is harder to parse than heavy-XP shift of an optional object, a fact that falls into place exactly because obligatory selection of the object forces the postulation of an empty node that cannot remain at the leading edge of the parse tree. Second, 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 a violation of the LEC.

This leads to an urgent question. Why should the very same parsing difficulties give rise apparent ungrammaticality 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 parsing difficulties, namely when the verb only optionally selects an object. On the other hand, rightward movement of a constituent containing the noun in the extended nominal projection will always cause parsing difficulties (as long as it is not string-vacuous). 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.

A substantial body of work suggests that difficulty in parsing corresponds to typological frequency. As Hawkins 2004 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 2004, 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. Such a mechanism has been proposed by Kirby 1999.

The foundation of Kirby’s analysis is the distinction between input and intake. The language-learning child is confronted with utterances in its environment (the input). However, in order to acquire a grammar, the child must parse those utterances to come to representations that can be used for the acquisition of grammar (see Fodor 1998, Gagliardi 2012, Omaki 2010 and Valian 1990). Structures that are hard to parse will have a reduced frequency in the intake compared to their frequency in the input. Kirby shows that over time this reduction will eliminate structures that are hard to parse, as long as the system is seeded with some initial variation.

This mechanism straightforwardly militates against the relevant rightward movements in the extended nominal projection; they systematically cause parsing difficulties and will therefore be suppressed in the child’s intake. Heavy-XP shift of optional objects, by contrast, does not cause parsing difficulties and therefore contributes to a presence of relevant inputs in the intake that is – we propose – sufficiently frequent to acquire a general process of rightward shift. 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 and 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 also apply to obligatory objects. A distinction could only be made by referring to the thematic make-up of the verb, which is exactly the kind of conditioning that the move-α program claims does not exist.

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.

5.4 The frequency of Universal 20 orders

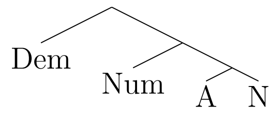
The account of Universal 20 developed in section 5.1-5.3 is based on the notion that the LEC radically reduces the typological frequency of structures derived by rightward movement of the noun (or a constituent containing the noun). This is a departure from the accounts in Cinque 2005 and Abels and Neeleman 2012, where the ban on rightward noun movement was taken to be grammatical and hence unviolable. In this section, we will explore how the LEC fits in with other factors that determine typological frequency.

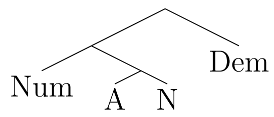
The typological distribution of Universal 20 orders follows a power law, with some orders much more frequent than others. It is possible to give a coherent account of the data based purely on formal factors. A good fit with the known distribution can be achieved if there is (i) a preference for harmonic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

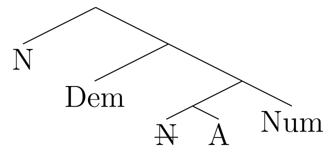
The effects of harmony are well known (see Cinque 2005, Culbertson et al. 2012, Culbertson and Enochson Newport 2015 and Dryer 2018). Our definition of harmony as it applies to extended projections is given in (53).

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

The definition in (53) characterizes the structure in (53a) as harmonic and the structure in (53b) as disharmonic. Any structure in which the noun moves is also disharmonic (unless the movement is string-vacuous). For example, if the noun sits in a head position in (53c), its sister is a nonprojecting category on right branch. However, the demonstrative, across which the noun has moved, is a nonprojecting category on a left branch. (This logic extends to structures in which the noun-adjective unit moves if, as Georgi and Müller's (2010) argue, it reprojects in its landing site. Reprojection will require slight adjustments of the sequence of parsing steps in section 5.3, but does not affect the conclusions drawn there.)

(54) a.


b.


c.


The preference for N-A order holds, we assume, of the base positions of these elements. Evidence for it 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 in N-Adj or Adj-N phrases the noun is interpreted first, followed by the adjective, regardless of syntactic word order. This sequence of interpretive steps is forced by the fact that the interpretation of gradable adjectives depends on the noun: a big butterfly, for instance, does not have the same size as a big elephant, but is rather a butterfly that is big for a butterfly. Thus, if the noun comes first, the adjective can be interpreted immediately, but if the adjective comes first, it has to be buffered for interpretation until after the noun is interpreted, and so the latter order is dispreferred as less efficient in parsing. (A preference for N-A order has been invoked before in Dryer's (2018) account of the frequency of Universal 20 orders, but Dryer treats it as a surface effect.)

The preference for nonmovement structures is tied to the fact that movement is a costly operation in parsing, requiring buffering and gap finding (as described in section 5.2).

We can understand the ranking of the fourteen Universal 20 orders if we assume that following hierarchy of preferences holds:

(55) Harmony > N-before-A > No-Movement

The predictions 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 is not (or not necessarily) merged to the left of the noun are more frequent than those in which the adjective *is* unambiguously merged to the left of the noun. (iii) Within each of the two disharmonic classes, nonmovement orders outrank movement orders. Thus, the ranking predicted by (55) is as given in table 14.

Notice that, like other accounts of typological frequency based on formal factors, the above presupposes a certain degree of 'linguistic entropy', possibly as a consequence of language contact. Without such a counteracting fluctuation in word order, all languages would 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.

	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 15: Ranking of Universal 20 orders as predicted by (55)

The hierarchy of preferences in (55) is crucial in generating the ranking in table 15. 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-course 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 (55) 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 order 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 frequency of genera containing languages of the relevant type as a measure. Rankings based on these measures are given in table 16 below (there are two counts of languages and two counts of genera, taken from Dryer 2018 and Cinque 2023, respectively; Cinque's sample is accessible at https://iris.unive.it/handle/10278/3750226). As the table shows, the five rankings do not always agree. In fact, the spread of rankings across the five measures goes up to six ranks for some orders. Nonetheless, the predicted ranking is a good match with the rankings based on each of the individual measures (mismatches appear in bold), and a perfect match with the average rankings.

	AF(D)	*LF(D)*	*GF(D)*	*LF(C)*	*GF(C)*	*Spr*	*A.R.*	*P.R.*
N-A-Num-Dem	1 (44.17)	1 (182)	1 (85)	1 (630)	1 (136)	1	1 (1)	1
Dem-Num-A-N	2 (35.56)	2 (113)	2 (57)	2 (442)	2 (115)	1	2 (2)	2
Dem-N-A-Num	3 (29.95)	4 (53)	3 (40)	4 (204)	3 (89)	2	3 (3.4)	3-5
Dem-Num-N-A	4 (22.12)	5 (40)	4 (32)	5 (184)	4 (76)	2	4-5 (4.4)	3-5
Num-N-A-Dem	**6 (14.54)**	3 (67)	5 (27)	3 (239)	5 (49)	4	4-5 (4.4)	3-5
N-A-Dem-Num	**5 (14.80)**	6 (36)	6 (19)	6 (103)	6 (35)	2	6 (5.8)	6-10
Dem-N-Num-A	7 (9.75)	8 (12)	8 (10)	10 (50)	9 (29)	4	7 (8.4)	6-10
N-Num-A-Dem	8-9 (9.00)	10 (11)	9 (9)	8 (70)	8 (34)	3	8 (8.7)	6-10
N-Dem-A-Num	8-9 (9.00)	7 (13)	7 (11)	**12 (32)**	**12 (17)**	6	9-10 (9.3)	6-10
N-Dem-Num-A	10 (5.67)	**11-12 (8)**	**11 (6)**	7 (83)	7 (25)	6	9-10 (9.3)	6-10
Dem-A-N-Num	11 (5.34)	**9 (12)**	**10 (7)**	11 (48)	**10 (27)**	3	11 (10.2)	11-13
Num-A-N-Dem	12 (4.00)	11-12 (8)	12 (5)	**9 (55)**	11 (21)	4	12 (11.1)	11-13
A-N-Num-Dem	13 (3.00)	13-14 (5)	13-14 (3)	13 (33)	13 (13)	2	13 (13.2)	11-13
A-N-Dem-Num	14 (2.50)	13-14 (5)	13-14 (3)	14 (20)	14 (8)	2	14 (13.8)	14

AF(D): adjusted frequency (Dryer 2018); LF(D): language frequency (Dryer 2018); GF(D): genus frequency (Dryer 2018); language frequency (Cinque 2023); GF(C): genus frequency (Cinque 2023); Spr: spread of rankings; AR: average ranking over the five measures; PR: predicted ranking according to table 14.

Table 16: A comparison between the predicted ranking and the actual ranking of the fourteen Universal 20 orders compatible with the LEC

We conclude that typological frequency in the Universal 20 domain is determined by a hierarchy of formal preferences. Among these, we find preferences that are associated with parsing costs (in particular the N-A preference and the preference for non-movement orders), alongside Harmony, which presumably reflects a bias in language acquisition.

Given that rightward movement orders, if they exist at all, are scarcer than the orders in table 15, it must be that case that the LEC affects typological frequency more strongly than the other factors mentioned:

(56) LEC > Harmony > N-before-A > No-Movement

We can partly understand why this should be so. The N-A preference and the preference for non-movement orders emerge because an increase in parsing effort will lead to a reduced presence of relevant orders in the intake (along the lines of the previous section). Clearly, the violation of a constraint regulating structure building should have a bigger effect on presence in the intake than a mere increase in effort, and hence it should influence typological frequency more profoundly. It is less clear why the bias for harmonic orders should be weaker than the push away from orders that violate the LEC, as there is no a priori expectation regarding the effects of learing biases vis-à-vis parsing factors.

Whether the LEC suppresses orders derived by rightward movement altogether 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 (2023) systematically goes through these languages and argues that none of them provide 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, it is interpreted as a partitive construction (Wendy Liz Arbey López Márquez, p.c.) and is 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 LEC are strong enough to reduce the typological frequency of rightward movement orders to zero or very close to it.

6. Conclusion

In closing, we highlight the consequences of our analysis of Universal 20 for the wider theory of syntax. As mentioned in the introduction, Universal 20 is one of the most intensely researched generalizations in linguistic typology and has turned out to be quite robust. In addition, the signature asymmetry of the pattern – fixed order to the left of the lexical head of a linguistic domain and variable order to the right – has been found well beyond the extended nominal projection. Cinque (2009, 2014) discusses instantiations of the Universal 20 pattern with adjectives and nouns (based on the adjectival hierarchy in (57a)), with PPs and verbs (based on the hierarchy in (57b)), with adverbials and verbs, based on the hierarchy in (57c)) and with verbs and mood, tense and aspect markers (mood comprises declarative markers, interrogative markers, imperative markers, etc.).

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

In addition, Abels (2017) demonstrates that word order in Germanic verb clusters closely matches the Universal 20, while Neeleman (2016) shows that the same is true of the distribution of prepositional phrases with regard to the verb in Dutch. It seems, then, that the pattern is one that is 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 right assessment, we can remove from universal grammar one of the stipulations in Abels and Neeleman 2012, namely that movements that derive neutral word orders are uniformly leftward. This is obviously an explanatory gain.

References

Abels, Klaus. 2016. The fundamental left-right asymmetry in the Germanic verb cluster. *Journal of Comparative Germanic Linguistics* 19: 179-220.

Abels, Klaus, and Ad Neeleman. 2012. Linear asymmetries and the LCA. *Syntax* 15: 25–74.

Ackema, Peter, and Ad Neeleman. 2002. Effects of short-term storage in processing rightward movement. In *Storage and computation in the language faculty*, ed. by Sieb Nooteboom, Fred Weerman and Frank Wijnen (219–256). Dordrecht: Kluwer.

Bach, Emmon. 1971. Questions. *Linguistic Inquiry* 2: 153–166.

Barker, Chris. 2012. Quantificational binding does not require c-command. *Linguistic Inquiry* 43: 614–33.

Bates, Douglas, Martin Mächler, Ben Bolker and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48.

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.
Bruening, Benjamin. 2014. Precedence-and-command revisited. *Language* 90: 342–388.
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. In *The view from Building 20*, ed. by Kenneth Hale and Samuel Jay Keyser (1–52). Cambridge, MA: MIT Press..
Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
Cinque, Guglielmo. 1996. The antisymmetric programme: theoretical and typological implications. *Journal of Linguistics* 32: 447–464.
Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36: 315–332.
Cinque, Guglielmo. 2009. The fundamental left–right asymmetry of natural languages. In *Universals of language today*, ed. by Sergio Scalise, Elisabetta Magni and Antonietta Bisetto (165–184). Dordrecht: Springer.
Cinque, Guglielmo. 2014. On the movement account of Greenberg's Universal 20: Refinements and replies. Ms. University of Venice.
Cinque, Guglielmo. 2023. Again on the order of demonstrative, numeral, adjective, and noun. Ms. University of Venice.
Crain, Stephen, and Janet Dean Fodor. 1985. How can grammars help parsers? In *Natural language parsing: Psychological, computational, and theoretical perspectives*, ed. by David R. Dowty, Lauri Karttunen and Arnold M. Zwicky (94–128). Cambridge: Cambridge University Press.
Culbertson, Jennifer, and Kelly Enochson. 2015. Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PLoS ONE* 10: 0116946.
Culbertson, Jennifer, and Elissa Newport. 2015. Harmonic biases in child learners: In support of language universals. *Cognition* 139: 71–82.
Culbertson, Jennifer, Paul Smolensky and Geraldine Legendre. 2012. Learning Biases Predict a Word Order Universal. *Cognition* 122: 306–329.
Den Dikken, Marcel. 1995. Extraposition as Intraposition, and the Syntax of English Tag Questions. Ms. Vrije Universiteit, Amsterdam.
Dryer, Matthew. 2007. Word order. In *Language typology and syntactic description. Vol. 1, Clause structure* (2nd edition), ed. by Timothy Shopen (61–131). Cambridge: Cambridge University Press.
Dryer, Matthew. 2018. On the order of demonstrative, numeral, adjective and noun. *Language* 94: 798–833.
Fodor, Janet Dean. 1998. Parsing to learn. *Journal of Psycholinguistic Research* 27: 339–374.
Forster, Kenneth, Christine Guerrera and Lisa Elliot. 2009. The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods* 41: 163–171.
Frazier, Lyn, and Giovanni Flores D'Arcais. 1989. Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language* 28: 331–344.
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.
Freedman, Sandra, and Kenneth Forster. 1985. The psychological status of overgenerated sentences. *Cognition* 19: 101–131
Gagliardi, Ann. 2012. *Input and intake in language acquisition*. Doctoral dissertation, University of Maryland.
Georgi, Doreen, and Gereon Müller. 2010. Noun-phrase structure by reprojection. *Syntax* 13: 1–36.
Gorrell, Paul. 1995. *Syntax and parsing*. Cambridge, MA: Cambridge University Press.

Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of language*, ed. by Joseph Greenberg (73–113). Cambridge, MA: MIT Press.
Grimshaw, Jane. 1991. Extended projection. Ms. Rutgers University.
Grimshaw, Jane. 2005. *Words and structure*. Stanford, CA: CSLI.
Haider, Hubert. 2005. How to turn German into Icelandic – and derive the OV–VO contrasts. *Journal of Comparative Germanic Linguistics* 8: 1–56.
Haider, Hubert. 2014. *Symmetry breaking in syntax*. Cambridge: CUP.
Hawkins, John. 1983. *Word order universals; Quantitative analyses of linguistic structure*. New York: Academic Press.
Hawkins, John. 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21: 223–261.
Hawkins, John 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
Hawkins, John 2009. Language universals and the performance-grammar correspondence hypothesis. In *Language Universals*, ed. by Morten Christiansen, Chris Collins and Shimon Edelmann (54–78). Oxford: Oxford University Press.
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. Blogpost available at https://hlplab.wordpress.com/2008/01/23/modeling-self-paced-reading-data-effects-of-word-length-word-position-spill-over-etc/.
Just, Marcel, and Patricia Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory". *Psychological Review* 99: 122–149.
Just, Marcel, Patricia Carpenter and Jacqueline Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111: 228–238.
Kamp, Hans, and Barbara Partee. 1995. Prototype theory and compositionality. *Cognition* 57: 129–191.
Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
Kirby, Simon. 1999. *Function, selection, and innateness: The emergence of language universals*. Oxford: Oxford University Press.
Kuznetsova, Alexandra, Per Brockhoff and Rune Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82: 1–26.
Lasnik, Howard, and Mamoru Saito. 1992. *Move a: Conditions on its application and output*. Cambridge, MA: MIT Press.
Lee, Ming-Wei. 2004. Another Look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research* 33: 51–73.
Levelt, Willem. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
MacDonald, Maryellen. 1994. Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes* 9: 157–201.
MacDonald, Maryellen, N. J. Pearlmutter, and Mark Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101: 676–703.
Marcus, Mitchell, Donald Hindle and Margaret Fleck. 1983. D-Theory: Talking about talking about trees. In *21st Annual Meeting of the Association for Computational Linguistics* (129–136). Cambridge, MA: Association for Computational Linguistics.
Michel, Jean-Baptiste, et al. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331: 176–182.
Mitchell, Don. 1984. An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In *New methods in reading comprehension research*, ed. by David Kieras and Marcel Just (69–89). Hillsdale, NI: Erlbaum.
Neeleman, Ad. 2017. PP-over-V meets Universal 20. *Journal of Comparative Germanic Linguistics* 20: 3–47.
Neeleman, Ad, and Amanda Payne 2020. PP Extraposition and the Order of Adverbials in English. *Linguistic Inquiry* 51: 471–520.

Omaki, Akira. 2010. *Commitment and flexibility in the developing parser*. Doctoral dissertation, University of Maryland.
Overfelt, Jason. (2015) *Rightward movement; A study in locality*. Doctoral dissertation, University of Massachusetts Amherst.
Pearson, Matthew. 2000. Two types of VO languages. In *The derivation of VO and OV*, ed. by Peter Svenonius (327–364). Amsterdam: John Benjamins.
Perlmutter, David. 1983. *Studies in Relational Grammar 1*. Chicago, IL: The University of Chicago Press.
Phillips, Colin, and Matthew Wagers. 2007. Relating structure and time in linguistics and psycholinguistics. In *The Oxford handbook of psycholinguistics*, ed. by Gareth Gaskell (739–756). Oxford: Oxford University Press.
R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Rijkhoff, Jan. 1990. Explaining word order in the noun phrase. *Linguistics* 28: 5–42.
Rijkhoff, Jan. 2002. *The noun phrase*. Oxford: Oxford University Press.
Rochemont, Michael, and Peter Culicover. 1990. *English focus constructions and the theory of grammar*. Cambridge: Cambridge University Press.
Ross, John. 1967. *Constraints on variables in syntax*. Doctoral dissertation, MIT.
Staub, Adrian, Charles Clifton 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.
Steddy, Sam, and Vieri Samek-Lodovici. 2011. On the ungrammaticality of remnant movement in the derivation of Greenberg's Universal 20. *Linguistic Inquiry* 42: 445–469.
Stowe, Laurie. 1986. Parsing WH-constructions: Evidence for online gap location. *Language and Cognitive Processes* 1: 227–245.
Traxler, Matthew, and Martin Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35: 454–475.
Trueswell, John, and Albert Kim. 1998. How to prune a garden-path by nipping it in the bud: Fast-priming of verb argument structures. *Journal of Memory and Language* 39: 102–123.
Trueswell, John, Michael Tanenhaus and Christopher Kello. 1993. Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of experimental psychology: Learning, memory, and cognition* 19: 528–553.
Valian, Virginia. 1990. Logical and psychological constraints on the acquisition of syntax. In *Language processing and language acquisition*, ed. by Lyn Frazier and Jill de Villiers (119–145). Dordrecht: Kluwer.
Wasow, Thomas. 1997a. End-weight from the speaker's perspective. *Journal of Psycholinguistic Research* 26: 347–361.
Wasow, Thomas. 1997b. Remarks on grammatical weight. *Language Variation and Change* 9: 81–105.

1 December 2023