Manuscript version March 2023
Accepted for publication in Konietzko, Andreas & Susanne Winkler (eds.). *Information structure and Discourse in Generative Grammar. Mechanisms and Processes* (Studies in Generative Grammar 146). Berlin & Boston: de Gruyter Mouton.

Álvaro Cortés Rodríguez

Which syntactician which kind of ellipsis: An experimental investigation of multiple sluicing

Abstract: This paper investigates the syntactic structure of multiple sluicing (MS) in English and German. The acceptability status of this construction in English remains controversial, and the judgments reported in the literature vary significantly (cf., Takahashi 1994; Nishigauchi 1998; Merchant 2001; Lasnik 2014; Kotek & Barros 2018). Therefore, obtaining experimental results on the acceptability status of MS is crucial for successful future investigations of this configuration. Here, I focus on two factors that have been reported to influence the acceptability of English multiple sluicing: *prepositionhood*, which is the presence or absence of a prepositional wh-phrase as the non-initial remnant, and *weight,* which refers to the number of heads following the wh-word in a non-initial remnant. The results of my experiments show that, despite multiple sluicing generally being a marked construction in English, speakers exhibit clear preferences for some configurations over others. First, the presence of a prepositional wh-remnant improves the overall acceptability of multiple sluicing sentences, which confirms claims from the literature (e.g., Richards 2010; Lasnik 2014). I argue that this preference is observed because prepositions provide additional cues which facilitate the retrieval of the information presented in the antecedent (cf. Martin & Mcelree 2008, 2011). Second, the results presented here reveal a significant negative effect on the overall acceptability of English multiple sluicing in the presence of a *heavy* wh-phrase, *contra* Lasnik (2014). I will interpret these results as motivated by information-structural factors, namely the reiteration of given material causing a penalty in the perceived acceptability of multiple sluicing sentences. Moreover, my experiment on German confirms that the multiple sluicing in this language is not modulated by the factors of prepositionhood and weight. I claim that the differences observed for prepositionhood and weight in the two examined languages can be captured by an implementation of a cue-based retrieval approach to ellipsis (Martin & Mcelree 2008, 2011; Nykiel, Kim & Sim 2022), once coupled with stipulations about the differing grammaticality status of MS in English and German.

Keywords: ellipsis, multiple sluicing, multiple wh-interrogatives, experimental syntax

1 Introduction

Sluicing refers to an utterance that contains a seemingly nonsentential wh-phrase that conveys a propositional meaning and which displays the syntactic distribution of a wh-clause (Ross 1969; Lasnik 1999), see (1) and (2B). The propositional meaning of a sluice is retrieved by reference to an *antecedent* in the proximate discourse context. Analyses of sluicing fall into two camps: those claiming that the wh-phrase is the only pronounced phrase in an otherwise unpronounced yet syntactically-present wh-clause (Ross 1969; Levin 1982; Chung, Ladusaw & McCloskey 1995; Merchant 2001; Stjepanović 2003; Aelbrecht 2010; Chung 2013; Lasnik 2014; Abels 2018a, 2018b; among many others), and those claiming that no unpronounced linguistic material associates with the wh-phrase whatsoever (Ginzburg & Sag 2000; Culicover & Jackendoff 2005; Sag & Nykiel 2011). The prevailing variant of the former *sententialist* view is that the wh-phrase in sluicing configurations is the only surviving *remnant* of a phonological suppression operation (called *ellipsis*) that, when extraneous licensing conditions are satisfied, applies unselectively to the TP node of a wh-clause (Ross 1969, Merchant 2001), see (3).

(1) Nora painted something, but I just don't know **what**. (where the sluiced wh-phrase is boldfaced)

(2) A: Someone has stolen Oier's book.
B: **Who**?

(3) … but I just don't know **what**$_{1}$ ~~[$_{TP}$ Nora painted t_{1}]~~. (sententialist analysis of (1))

The focus of this paper is *multiple sluicing* (MS), i.e., configurations displaying two (or more) sluiced wh-phrases (see (4)). Reading from right to left, I henceforth refer to sluiced wh-phrases[1] in an MS configuration as *swh1* and *swh2*. I refer to the phrases in the antecedent with which swh1 and swh2 respectively co-refer (which are italicized in (4)) as *correlate1* and *correlate2*.

(4) *Everyone* ate *something*, but I just don't know **who what**.

The extant literature on English MS is characterized by an absence of agreement on the basic data. Researchers either report MS in English as impossible (Takahashi 1994; see also Takahashi & Lin 2012), marginal (Nishigauchi 1998; Merchant 2001; Hoyt & Teodorescu 2012; Lasnik 2014), or subject to a significant inter-speaker variation (Kotek & Barros 2018; Barros & Frank 2016, to appear). Without agreement on the basic data points, it is difficult to adjudicate between competing analyses. Putting aside some exceptions, all extant analyses of MS are based on acceptability judgments that were collected informally.[2] This collection method involves eliciting acceptability judgments in uncontrolled settings (e.g., via email or during casual conversation) by asking a few native speakers (perhaps only one native speaker, the linguist herself) for a non-numerical reaction – usually 'acceptable' or 'unacceptable' – to a target utterance. Since researchers on English MS concede that MS yields subtle judgments (e.g., Lasnik 2014), which informal collection methods struggle to reliably reflect, research on MS would greatly benefit from results obtained using formal experimental methods. In this paper, I initiate the process of placing research on MS on firmer empirical ground by reporting the results of acceptability judgment experiments conducted in English and German using formal experimental methods.

Various independent causal factors have been claimed to influence the perceived acceptability of MS in English. In the present study, I will focus on two factors: *prepositionhood* and *weight*. Here, *prepositionhood* refers to whether swh2 is accompanied by a preposition, and weight refers to how many syntactic heads constitute swh2. I concentrate on prepositionhood because there is substantial disagreement regarding whether or not the presence of a preposition with swh2 yields higher acceptability ratings for English MS (cf., e.g., Bolinger 1978; Merchant 2001; Richards 2010; Lasnik 2014; Kotek & Barros 2018). I test weight because influential sententialist analyses of English MS that claim that swh2 undergoes rightward extraposition to escape the ellipsis (Lasnik 2014; Chang & Kim 2013; Park 2015) predict that increased weight will yield higher ratings for English MS.

Because it is important to determine whether an analysis proposed for English can extend to other languages or is inherently language-specific, I introduce a cross-linguistic comparative element to the current study by also experimentally testing the two above-mentioned factors in German. It has been suggested that German accepts MS more readily (Merchant 2006; Richards 2010; Winkler 2013). For instance, Richards (2010) argues that, unlike in English, languages with a rich case morphology like German allow for multiple sluices involving two DPs. Richards' claim has the direct implication that swh2 being a PP or a DP should not affect the acceptability of MS in German. With regards to weight, no predictions have been offered for German. However, it is worth noting that, in their discussion of German MS, Abels & Dayal (to appear) employ both prepositional and nominal swh2s, as well as what I call 'bare' and 'explicit' swh2s but without making concrete mention to their acceptability. Yet those authors provided no further information about those factors. In addition, German offers a good ground for determining the possible ameliorating effect of having morphological case-marking on nominal wh-phrase remnants beyond the presence of prepositions as case-assigning heads.

The organization of this paper is as follows. In Section 2, I will give an overview of the research on MS, focusing on the two main properties under investigation in the upcoming experiments, namely *prepositionhood* and *weight* of swh2, and two additional properties related to weight, namely *harmony* and *congruence*. The latter are included as factors of insurance against the possibility that either property contributes to influencing acceptability judgments. In Section 2, I also provide more information on the cross-linguistic perspective on MS. Section 3 presents three acceptability judgment studies in English and one acceptability judgment study in German that test the interplay of *prepositionhood* and

---

[1] The terms (wh)-remnant and sluiced wh-phrase (swh) will be used interchangeable without a making a standpoint to a particular framework or underlying analysis assumption.

[2] The most notable report on informally collected MS data can be found in Lasnik (2014), who provides an appendix with pooled data of 29 subjects. The data was not collected using formal experimental methods but rather by informal judgments collected per email for a small set of items. Informants could view all items together while evaluating them. Lasnik's study included a total of 12 sentences, out of which five were instances of MS (four <DP, PP> and one <DP, DP>, none of which included a "heavy" non-initial sluiced wh-phrase).

To my knowledge, only Chung & Park (2017) provide a formal investigation on MS. Their study, however, focuses on MS configurations with two prepositional complements. They compare the interaction of preposition pied-piping and *swiping* (see Merchant 2002) between native and non-native speakers of English.

weight of swh2. Additionally, this section explores the possible interaction between the form of the initial and the non-initial wh-remnants (i.e., the factor *congruence*). The results obtained are that (i) for English, prepositionhood improves the acceptability of MS, whereas weight has the opposite effect, but (ii) for German, neither prepositionhood nor weight has a significant effect on the overall acceptability of MS. Here, I also discuss the importance of experimental results in the current theoretical landscape. I argue that the results severely undermine Lasnik's (2014) rightward exposition analysis of English MS.[3] I also show that, although the results are compatible with Richard's (2010) *Distinctness* theory (trivially so, in the case of the factor *weight*), there are independent reasons to be skeptical of this theory as applied in the domain of ellipsis. I suggest that *prepositionhood* and *weight* are both non-grammatical influencers of MS acceptability: they are reflections of the (un)availability of cues used by the hearer during processing to link a sluiced wh-phrase to its correlate.

2 Background

2.1 Aspects (possibly) affecting the acceptability of multiple sluicing

2.1.1 Prepositionhood

Fox & Pesetsky (2003) and Richards (2010) claim that English MS sentences combining a nominal sluiced wh-phrase and a PP sluiced wh-phrase are acceptable. In contrast, those MS sentences involving two nominal sluiced wh-phrases are unacceptable. This contrast in acceptability is suggested to hold regardless of whether the wh-phrase is a bare wh-pronoun or a wh-phrase composed of *which* and a restrictor NP (Richards 2010); see the contrasts in (5) and (6).

(5) a. I know everyone danced with someone, but I don't know who with whom.
b. * I know everyone insulted someone, but I don't know who whom. (Richards 2010:3)

(6) a. I know every man danced with a woman, but I don't know which man with which woman.
b. * I know every man insulted a woman, but I don't know which man which woman. (ibid.)

Richards (2010) explains this disparity in terms of *Distinctness* (7), which stipulates that two nodes of the same syntactic category cannot be linearized in the same Spell-Out domain.[4] Thus, Distinctness is not sensitive to linear adjacency but to Spell-Out domains. Additionally, Richards argues that Distinctness only affects functional projections but not lexical ones.[5] Although Richards (2010) does not provide an example of English MS with two PPs (cf. Chung & Park 2017, see also Richards 2001), he discusses extraneous structures in which adjacent PPs are possible. He preliminarily argues that the well-formedness of (8) can be explained if Ps are considered to introduce a phase boundary.

(7) If a linearization statement $< \alpha, \alpha >$ is generated, the derivation crashes. (ibid.:5)

(8) the gift [of a book] [to John] [by Mary] (ibid.:9)

Although two linearly adjacent DPs always trigger a Distinctness violation in English, this configuration triggers no such violation in languages with richer morphological cases, such as German or Japanese (Richards 2010). Thus, Richard's account predicts that having two linearly adjacent sluiced DP *wh*-phrases in German does not cause a degradation in

---

[3] Lasnik's explanation has a very English-specific character since extraposition of wh-phrases is not possible, for instance, in German or Hindi (Abels & Dayal 2017, 2021). I would like to thank an anonymous reviewer for pointing this out to me.

[4] Richards (2010) assumes that Spell-Out is a cyclic operation that is triggered whenever a strong phase, i.e., CP, *v*P, PP, or K(ase)P, is completed.

[5] On the one hand, the following have been considered lexical categories: N, V, A (Halle & Marantz 1993; Embick & Noyer 2007; Zwarts 2011), and on the other hand, abstract morphemes (Embick & Noyer 2007) and D, C, I/T (Zwarts 2011) have been classified as functional categories. However, the characterization of prepositions has been under much debate in the generative grammar, where Ps have been considered to have properties for being both a lexical or a functional category (Corver & Van Riemsdijk 2001; Zwarts 2011). Richards (2010) seems to consider Ps as functional categories, thus they should be affected by the Distinctness rule.

acceptability, and therefore his account predicts that whether swh2 is a PP or a DP has no effect on the acceptability of German MS.

Lasnik (2014) also reports that, all other things being equal, English MS is more acceptable when swh2 is a PP than a DP (9). He interprets this disparity as evidence that swh2 undergoes rightward extraposition in English MS, as this displacement operation is independently known to be more readily available with PPs than DPs (compare the examples in (10), in which a PP or DP undergoes extraposition to a rightward position structurally higher than the temporal adverb *yesterday*).

(9) a. ? Someone talked about something, but I can't remember who about what.
b. ?* Someone saw something, but I can't remember who what. (Lasnik 2014:8)

(10) a. Some students met yesterday with some professors.
b. * Some students met yesterday some professors. (ibid.)

The opinion of Fox & Pesetsky (2003), Richards (2010), and Lasnik (2014) that English MS with two DP sluiced *wh*-phrases is degraded is not unanimously shared. For example, Merchant (2001, 2006), Kotek & Barros (2018), and Barros & Frank (2016, to appear) report that such configurations are either mildly deviant (Merchant) or fully acceptable (Kotek & Barros, and Barros & Frank):

(11) a. ? Everyone bought something, but I couldn't tell you who what. (Merchant 2006:284)
b. Every boy likes some girl, but I don't know which boy which girl. (Kotek & Barros 2018:799)

Kotek & Barros (2018) report that their informally collected acceptability judgment data on English MS shows substantial interspeaker variation, with some native speakers consistently judging MS sentences with two DP sluiced wh-phrases as fully acceptable.

2.1.2 Weight

As mentioned in the previous section, Lasnik (2014) claims that English MS sentences in which both of the sluiced wh-phrases are DPs are highly degraded. Lasnik also claims that, if swh2 is 'heavy' in such a configuration, this heaviness has an ameliorating effect on acceptability judgments, thus raising the MS configuration from "?*" to "?".

(12) a. ?* Someone saw something, but I can't remember who what. (Lasnik 2014:8)
b. ? Some linguist criticized yesterday some paper about sluicing, but I don't know which linguist which paper about sluicing. (ibid:9)

Because it is independently known that DPs can only undergo rightward extraposition in English if they are 'heavy' (Drummond, Hornstein & Lasnik 2010) (see (13)), Lasnik views the apparent presence of a similar condition on swh2 in English MS as further proof that swh2 undergoes rightward extraposition.

(13) a. ?* I saw yesterday Jim.
b. ?? I saw yesterday Harry.
c. I saw yesterday Alexander. (Drummond, Hornstein & Lasnik 2010:690)

To my knowledge, Lasnik (2014) is the only researcher who claims that the weight of a nominal swh2 influences the acceptability of English MS. Lasnik presents only one example featuring a heavy swh2, which has already been repeated here in (12b) above.[6] Moreover, the extant literature on MS contains no cited examples in which swh1 and swh2 are radically different in terms of their weight. For instance, I have found no cited examples of the form "*…who which student in linguistics*". Whether or not this absence of variation in the examples from the literature hints toward a

[6] Provided the mixed judgments found in the literature for different MS structures, and prompted by one of the reviewers, Lasnik (2014) runs a small acceptability judgment experiment with 29 subjects. This study included 12 sentences, five out of which included an instance of MS. None of his experimental sentences included a heavy non-initial wh-remnant.

previously unnoticed restriction/preference on English MS (namely, that swh1 and swh2 should be of roughly equal weight) is entirely unknown. To determine if this factor, which I henceforth call *congruence*, impacts acceptability judgments, I include it as an 'across-sub-experiment' factor in my experimental design (see Section 3.1.3).

2.2 Multiple sluicing cross-linguistically

So far, we have inspected multiple sluicing in English and discussed that this construction seems to be marked and subject to high interspeaker variation. Nevertheless, it is worth mentioning that multiple sluicing is found in languages with different mechanisms for wh-question formation. For instance, multiple sluicing, perhaps unsurprisingly, is more productive in multiple wh-fronting languages; for example, in Slavic languages such as Serbo-Croatian (Stjepanović 1999), Slovenian (Marušič & Žaucer 2013), Russian (Grebenyova 2009), but also in Romanian (Hoyt & Teodorescu 2012). Additionally, multiple sluicing also seems to be productive in wh-in-situ languages like Japanese (Takahashi 1994; Nishigauchi 1998; Hiraiwa & Ishihara 2002) or Mandarin Chinese (Bai & Takahashi under review; Takahashi & Lin 2012), although the acceptability status of MS in this latter language has been debated (cf. Chiu 2007). In other single-fronting languages like Spanish, MS has been reported to be allowed (Rodrigues, Nevins & Vicente 2009; Ortega-Santos 2016; Gallego 2017). However, my experimental investigation also showed that native speakers of Peninsular Spanish consider it marginal at best and subject to prepositionhood (Cortés Rodríguez 2021). Interestingly, German seems to allow MS more readily thanks to a richer case morphology (Richards 2010; see also Merchant 2006; Winkler 2013). For this reason, I will include an acceptability judgment study in German to compare its results to the results obtained for English. Additionally, testing multiple sluicing in German will allow us to investigate the effects of case marking and prepositionhood independently. A direct contrast in the acceptability of MS in English and German can be seen in (14) and (15).

(14) ?* Someone saw something, but I can't remember who what. (Lasnik 2014:8)

(15) Jemand hat [et]was gesehen, aber ich weiß nicht, wer was. (Merchant 2006:285)
someone has something seen but I know not who what
'Someone has seen something, but I don't know who what.'

Dutch also makes the crucial differentiation. Richards (2010) (see also Merchant 2001) argues that Dutch is like German in that it allows MS with two DPs. This is surprising given the arguments about case morphology, and we would expect Dutch to be more similar to Spanish and English than to German in this regard. In fact, in an informal survey with 20 participants conducted by De Vries (2020), he finds that Dutch native speakers show a preference for MS involving a non-initial preposition wh-phrase (see (16)). Those results go against Richards' (2010) predictions for Dutch and are in-line with the arguments presented here for case-marking and retrieval cues.

(16) a. Jan heeft een boek gegeven aan een of ander meisje, mar ik weet niet welk boek aan
Jan has a book given to one or other girl, but I know not which book to
welk meisje.
which girl.

b. %? Een inbreker heeft een buurman bestolen, mar ik weet niet welke inbreker welke buurman.
a burglar has a neighbor roobed but I know not which burglar which burglar
(De Vrie 2020:4)

This informal sketch suggests that properties specific to a certain language may affect the acceptability of MS in that language. By conducting the upcoming experiment in English and German, we may obtain formal confirmation of this suggestion. The precise motivation for a formal investigation in German will become clearer after the discussion of the English results in Section 3.1.3.

3 Experimental evidence

3.1 Experiment series: Multiple sluicing in English

3.1.1 Methods

Design and Materials

I conducted three sub-experiments to test the acceptability of MS in English when modulated by the factors of PREPOSITIONHOOD and WEIGHT. Each sub-experiment investigated the acceptability of MS under the presence or absence of a preposition in swh2. In addition, each of these experiments compared the influence of swh2's weight – in terms of the number of syntactic heads – on acceptability. This design allows for determining how strongly prepositionhood and weight independently affect judgments. Furthermore, I investigated whether the interaction of both factors increases acceptability beyond what is expected for each independent factor. Moreover, swh1 was modified for weight across each sub-experiment. I implemented this manipulation in order to determine whether the effect caused by the factor modifying swh2 is independent and thus unaffected by the weight of swh1. I constructed 30 test items per experiment using a 2×3 design with PREPOSITIONHOOD and WEIGHT as within-item and within-subject factors. Both factors were manipulated on swh2 only. PREPOSITIONHOOD has two levels, *+P* and *–P*, indicating the presence or absence of a preposition. WEIGHT refers to the number of syntactic heads constituting the sluiced wh-phrase. It has three levels, *bare* (a single wh-pronoun, *who*), *explicit* (*which* plus a simplex NP restrictor, *which student*), and *heavy* (*which* plus a complex NP restrictor, *which student of Chemistry*).[7] The experiment was divided into three sub-experiments to test the possible effect of varying swh1's weight on acceptability judgments. Swh1 is *bare*, *explicit*, and *heavy* in sub-experiments 1, 2, and 3, respectively.

Other potential influences on acceptability were controlled for. For instance, although there is disagreement in the literature about what scopal relations can obtain between correlate1 and correlate2 in the antecedent of an MS sentence,[8] scholars agree that, when all extraneous conditions are satisfied, an MS utterance will be judged as acceptable if correlate1 is (or contains) a universal quantifier and correlate2 is (or contains) an existential quantifier. Thus, each test sentence employed in the experiment displays correlates of this type.

Each test sentence also displays *harmony* between the two correlates and their corresponding sluiced wh-phrases. Here, *harmony* refers to equal weighting in each correlate/wh-phrase pair. Concretely, bare sluiced wh-phrases are always paired with bare correlates (e.g., *someone* and *who*), explicit wh-phrases are always paired with explicit correlates (e.g., *some student*, *which student*), and heavy wh-phrases are always paired with heavy correlates (e.g., *some student of Chemistry*, *which student of Chemistry*). The test sentences were constructed in this manner to avoid the possible detrimental effect of disharmony on acceptability judgments (see Nykiel 2013 for experimental evidence that "maximal overlap" between the correlate and the remnant yields higher acceptability ratings).

Also, the sluiced wh-phrases will be immediately preceded by *know* rather than any other embedded question selecting verb. This restriction is emplaced to align with Nishigauchi's (1998) claim that English MS is most acceptable in a '*know*' environment. Additionally, all test sentences obey the clausemate condition,[9] which states that correlates in an MS configuration must originate in the same clause (e.g., Takahashi 1994; Nishigauchi 1998; Merchant 2001; Lasnik 2014; Abels & Dayal 2017, to appear; see also Cortés Rodríguez in press for an overview and an experimental investigation). It is widely believed that several other syntactic phenomena are constrained to occur within local clausal

[7] Several definitions of grammatical weight have been offered in prior literature, ranging from string length (Behaghel 1909) to grammatical complexity based on the number of nodes in a tree structure (Chomsky 1975). From the results of a corpus study, Wasow (1997) concluded that all definitions of heaviness yield similar results. With this in mind, I use Lasnik's example from (12) above as an exemplar of heaviness. For the purposes of the upcoming experiment, a 'heavy' wh-phrase is one that contains the wh-word *which* plus a nominal restrictor that itself contains a prepositional expression (e.g., *which boy with long trousers*).

[8] On the one hand, some scholars defend that it is a requirement for the first correlate in an MS construction to have a universal force (Nishigauchi 1998; Merchant 2001; see also Abels & Dayal 2017, to appear). On the other hand, some other scholars admit that MS configurations can be composed of two existential quantifiers (Lasnik 2014; Kotek & Barros 2018; Barros & Frank 2016, to appear). For the sake of simplicity, all experimental material presented here follows the former pattern.

[9] It is important to note that Lasnik's rightward extraposition analysis derives the clausemate condition. See however Cortés Rodríguez (in press) for an alternative analysis and empirical investigation.

domains as well, such as multiple questions, gapping, and quantifier raising (see, e.g., (Farkas & Giannakidou 1996; Grano & Lasnik 2018).

Each experimental sentence consists of three clauses: the antecedent, *intro,* and sluice, as diagrammatically represented in Table 1. Tables 2-4 provide an example item for each of the sub-experiments.

Table 1: Structure of the critical items

Antecedent			**Intro**	**Sluice**		
Every-	*verb (preposition)*	some-,	but I just don't know	wh-	*(preposition)*	wh-
Correlate 1 = universal quantifier	*(prepositional) Verb*	*Correlate 2 = existential quantifier*	*governing expression*	*Remnant 1*	*absence or presence of preposition*	*Remnant 2*

Table 2: One example item from Sub-experiment 1

Sentences	**Condition**
Everyone completed something, but I just don't know who what.	[-P/bare]
Everyone completed some essay, but I just don't know who which essay.	[-P/explicit]
Everyone completed some essay about colonialism, but I just don't know who which essay about colonialism.	[-P/heavy]
Everyone commented on something, but I just don't know who on what.	[+P/bare]
Everyone commented on some essay, but I just don't know who on which essay.	[+P/explicit]
Everyone commented on some essay about colonialism, but I just don't know who on which essay about colonialism.	[+P/heavy]

Table 3: One example item from Sub-experiment 2

Sentences	**Condition**
Every undergrad completed something, but I just don't know which undergrad what.	[-P/bare]
Every undergrad completed some essay, but I just don't know which undergrad which essay.	[-P/explicit]
Every undergrad completed some essay about colonialism, but I just don't know which undergrad which essay about colonialism.	[-P/heavy]
Every undergrad commented on something, but I just don't know which undergrad on what.	[+P/bare]
Every undergrad commented on some essay, but I just don't know which undergrad on which essay.	[+P/explicit]
Every undergrad commented on some essay about colonialism, but I just don't know which undergrad on which essay about colonialism.	[+P/heavy]

Table. 4: One example item from Sub-experiment 3

Sentences	**Condition**
Every activist of Greenpeace praised something, but I just don't know which activist of Greenpeace what.	[-P/bare]
Every activist of Greenpeace praised some documentary, but I just don't know which activist of Greenpeace which documentary.	[-P/explicit]
Every activist of Greenpeace praised some documentary on penguins, but I just don't know which activist of Greenpeace which documentary on penguins.	[-P/heavy]
Every activist of Greenpeace talked about something, but I just don't know which activist of Greenpeace about what.	[+P/bare]

Sentences	Condition
Every activist of Greenpeace talked about some documentary, but I just don't know which activist of Greenpeace about which documentary.	[+P/explicit]
Every activist of Greenpeace talked about some documentary on penguins, but I just don't know which activist of Greenpeace about which documentary on penguins	[+P/heavy]

Items in each experiment were distributed across six lists according to a Latin square design and randomized within each list. Participants in all three sub-experiments saw in total five items in each condition. In addition to the 30 critical items, 90 fillers were included. Fifteen out of those filler sentences corresponded to the 5-degree standardized items from Gerbrich, Schreier & Featherston (2019), which range from A-type fillers (most natural) to E-type fillers (least natural). Furthermore, among the filler sentences, 16 fillers displaying a *gapping* configuration were included (8 <DP, DP> and 8 <DP, PP>); examples are presented in (17) and (18). I will discuss these fillers and their relation to MS in more detail in the discussion section. The experimental design was identical for all three sub-experiments.

(17) In July, David will vote for the greens and Lisa for the liberals. [prepositional]

(18) On Halloween, Steve will wear a hat and William a mask. [gapping nominal]

Participants and Procedure

A web-based acceptability judgment task was designed using *PsychoPy 3* experiment creation application (Peirce et al. 2019) and *Pavlovia* as the experiment hosting platform.[10] 37_{Exp1} | 36_{Exp2} | 41_{Exp3} participants[11] were recruited via *Prolific*[12]. Participants were adult native speakers of English residing in the UK or USA (self-reported) and were naïve to the experiment's purpose. Furthermore, participants were prohibited from participating in more than one sub-experiment. The task had to be performed on a computer screen: smartphones were ruled out. Participants read and rated 120 sentences using a 7-point Likert scale ranging from 1 'very unnatural' to 7 'very natural'. They were also instructed that there was no "right" or "wrong" answer and were asked to follow their intuitions. Participants received monetary compensation of £2.50 for their participation in this study, which lasted approximately 15 minutes. In Sub-experiment 1, two participants had to be removed for disclosing a native language other than English.Additionally, based on the judgments given by the participants to the standardized items (Gerbrich, Schreier & Featherston 2019), 7_{Exp1} | 3_{Exp2} | 9_{Exp3} participants were excluded from the analysis for misusing the scale. Therefore, data from 28_{Exp1} | 33_{Exp2} | 32_{Exp3} participants were included in the analysis and will be reported in the results. Finally, every participant completed a practice round comprising five sentences before judging the experimental items.

Predictions

(19) **Predictions regarding PREPOSITIONHOOD**

[Pi] English MS configurations in which swh2 is a PP are rated significantly higher than those in which swh2 is a DP.

[Pii] For English, there is substantial interspeaker variation, with some speakers rating DP swh2 configurations as fully acceptable.

(20) **Predictions regarding WEIGHT**

[Wi] In English MS configurations, *heavy* DP swh2s are rated significantly higher than their *bare* and *explicit* counterparts.

[Wii] English MS configurations in which the swh1 and swh2 are *congruent* are rated significantly higher than those in which no such congruence obtains.

---

[10] https://pavlovia.org/

[11] Throughout the experimental section, I will be using the sub-indices "Expl", "Exp2", and "Exp3" to refer to the three sub-experiments I have conducted. I will include them next to the numbers I am reporting to facilitate the readability.

[12] https://prolific.co/

3.1.2 Data analysis and results

All data presented in this paper were analyzed using the R statistics software, Version 3.6.1 (R Core Team 2020). To test for significant effects, I analyzed the judgment data employing logistic regression, using the *ordinal* package and the *clmm* function (which stands for Cumulative Link Mixed Modeling) (Christensen 2019). As a first step, experimental factors and interactions were entered as fixed effects, as well as random effects for items and subjects with maximal random slopes. For each iment, I report the model with the best maximal fixed and random effect structure supported by the data, which is obtained using backward model selection.[13] The corresponding formulas are included in the tables showing the statistical analyses.

Figure 1 shows the mean acceptability ratings obtained in Sub-experiment 1. The corresponding statistical analysis is given in Table 5. There are two main effects. Prepositional swh2s received significantly higher ratings than nominal swh2s. The main effect WEIGHT shows that heavier expressions receive significantly lower ratings. All WEIGHT levels significantly differ from their preceding level ($p < .001$).

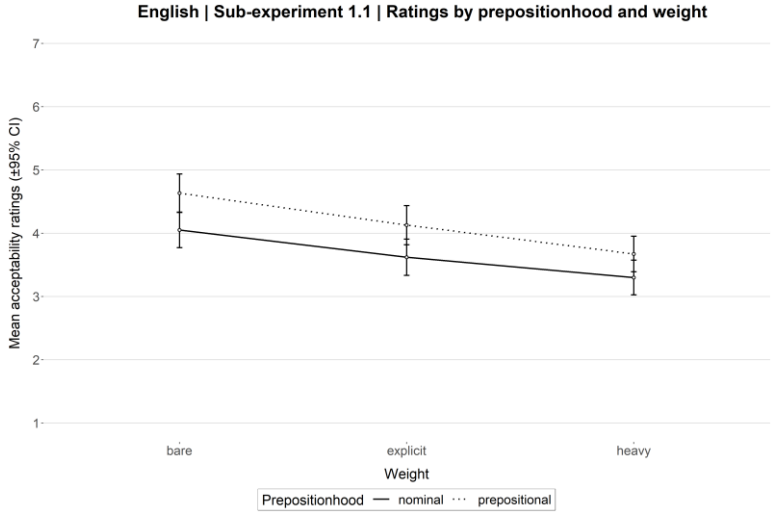

Figure 1: Mean acceptability rating for Sub-experiment 1 (n = 28). Error bars show 95% confidence interval.

Table 5: Cumulative Link Mixed Model fitted with the Laplace approximation for Sub-experiment 1

	Estimate	Std. error	*z* value	Pr(>|*z*|)
PREPOSITIONHOOD*(prepositional)*	0.8821	0.2600	3.316	0.000912 ***
WEIGHT*(explicit-bare)*	-0.8591	0.1894	-4.537	5.71 e-06 ***
WEIGHT*(heavy-explicit)*	-0.7182	0.2173	-3.305	0.000951 ***

Formula: *rating ~ prepositionhood + weight + (prepositionhood ∗ weight | subject), threshold="flexible"*

Figure 2 shows the mean acceptability ratings obtained in Sub-experiment 2, and the results of its statistical analysis are presented in Table 6. The results show a main effect for the factor PREPOSITIONHOOD ($p < .001$), whereby conditions containing a PP swh2 were judged significantly more acceptable than conditions with a DP swh2. Concerning WEIGHT, no significant difference ($p = .49$) is observed between *bare* and *explicit* swh2s. Conversely, there is a significant effect ($p < .001$) between the levels *explicit* and *heavy*, showing a detriment in acceptability for the *heavy* conditions. Additionally, since the model reported in Table 6 was coded for s*liding differences* contrast, the difference between the levels *bare* and *heavy* is not present. In a separate analysis using *effect coding*, the difference between *bare* and *heavy* was also found to be significant.

[13] This backward model selection was performed by starting with a full model including all predictors as fixed effects and the interaction between them, as well as all random effects, which comprised random intercepts and random slopes for subjects and items. The model was subsequently simplified one factor at a time and compared to its minimally diverging more complex model using the *anova* function.

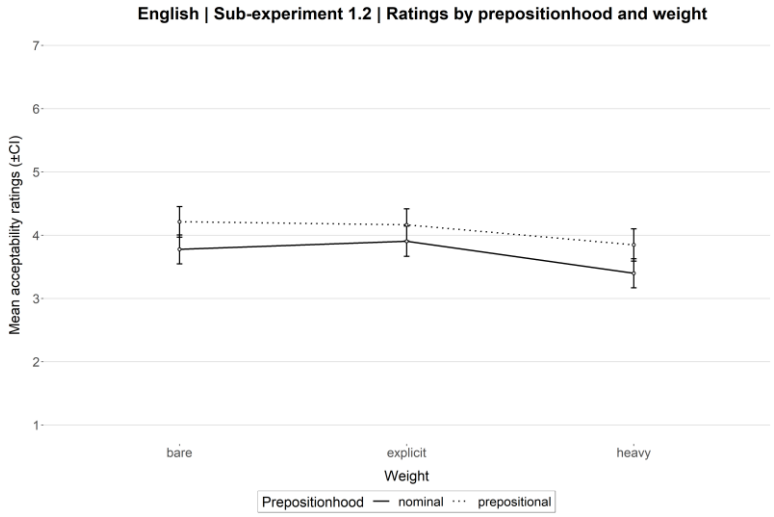

Figure 2: Mean acceptability rating for Sub-experiment 2 (n = 33). Error bars show 95% confidence interval.

Table 6: Cumulative Link Mixed Model fitted with the Laplace approximation for Sub-experiment 2

	Estimate	**Std. error**	***z* value**	**Pr(>|*z*|)**
PREPOSITIONHOOD*(prepositional)*	0.7083	0.2096	3.380	0.000726 ***
WEIGHT*(explicit-bare)*	0.1077	0.1566	0.688	0.491633
WEIGHT*(heavy-explicit)*	-0.8214	0.2275	-3.611	0.000305 ***

Formula: *rating ~ prepositionhood + weight + (prepositionhood + weight | subject), threshold="symmetric2"*

The mean acceptability ratings obtained in Sub-experiment 3 are presented in Figure 3, and Table 7 displays the corresponding statistical analysis. Similar to the results obtained for Sub-experiment 2, those results show a main effect for PREPOSITIONHOOD ($p < .005$). Concerning WEIGHT, the difference between the *bare* and *explicit* levels is not significant ($p = 0.26$). The difference between the factor levels for the levels *explicit* and *heavy*, however, is significant ($p < .001$). For completeness, it is worth noting that in a separate analysis using *effect coding*, the difference between *bare* and *heavy* did not show a significant effect.

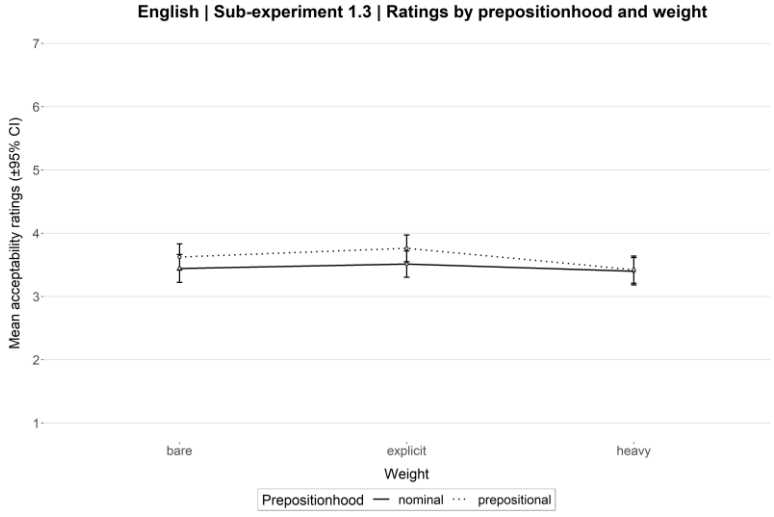

Figure 3: Mean acceptability rating for Sub-experiment 3 (n = 32). Error bars show 95% confidence interval.

Table 7: Cumulative Link Mixed Model fitted with the Laplace approximation for Sub-experiment 3

	Estimate	Std. error	*z* value	Pr(>|*z*|)
PREPOSITIONHOOD*(prepositional)*	0.3370	0.1176	2.865	0.00417 **
WEIGHT*(explicit-bare)*	0.1824	0.1611	1.133	0.25742
WEIGHT*(heavy-explicit)*	-0.3735	0.1703	-2.193	0.02834 *

Formula: *rating ~ prepositionhood + weight + (prepositionhood + weight | subject), threshold="symmetric"*

Lastly, Figure 4 shows the combined mean acceptability ratings obtained across the three sub-experiments. I follow Gerbrich, Schreier & Featherston's (2019) approach for comparison across experiments. Using the 15 standard items I included across all three sub-experiments, the ratings were *z*-score transformed using the methodology described by these authors. The *z*-score ratings were obtained by subtracting from each score the mean of the standard fillers and dividing the result by the standard deviations of the standard fillers.

I analyzed the normalized responses using a linear mixed-effect model (LMEM) with the *lme4* package and the *lmer* function (Bates et al. 2015). First, all experimental factors and their corresponding interactions were entered into the model, as well as including random intercepts for subjects and items and by-subject and by-item random slopes slopes for both subjects and items. Here I report the model with maximal random effect structure supported by the data (Barr et al. 2013), which I obtained by performing a backward model selection with *anova*. *P*-values were obtained using the *lmerTest* package (Kuznetsova, Brockhoff & Christensen 2017), which uses Satterthwaite's method. The results of the statistical analysis are presented in Table 8. Like in each individual sub-experiment, the effect for PREPOSITIONHOOD is highly significant ($p < .001$). As for the WEIGHT factor, there is no significant effect between the levels *bare* and *explicit*, and there is only a significant detrimental effect ($p < .001$) in acceptability between the *explicit* and *heavy* conditions (which entails a significant effect between *bare* and *heavy*).

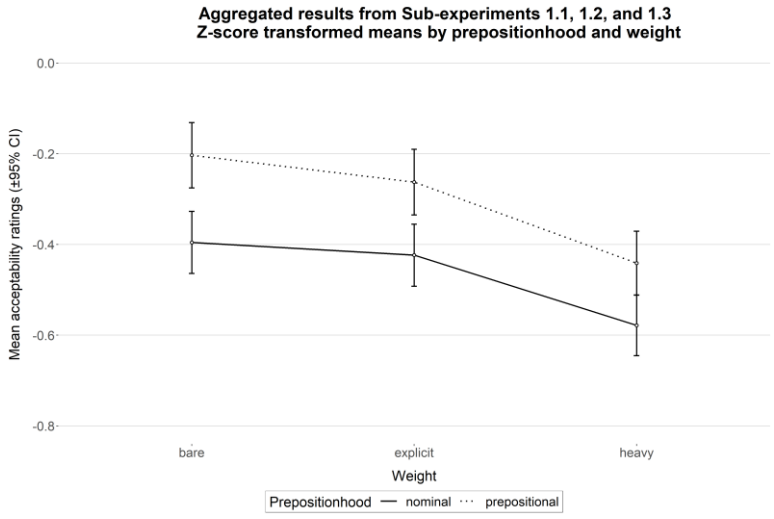

Figure 4: Combined mean acceptability *z*-score for all three sub-experiments (n = 93). Error bars show 95% confidence interval.

Table 8: Linear Mixed Effect Model fitted by maximum likelihood for the aggregated data across the three sub-experiments

	Estimate	Std. error	df	t value	Pr(>|t|)
(Intercept)	-0.45996	0.05897	92.97472	-7.800	
PREPOSITIONHOOD*(prepositional)*	0.16496	0.02933	92.65150	5.624	1.96e-07 ***
WEIGHT*(explicit-bare)*	-0.04362	0.02703	144.53699	-1.613	0.109
WEIGHT*(heavy-explicit)*	-0.16774	0.03038	110.74041	-5.522	2.24e-07 ***

Formula: *z-score ~ prepositionhood + weight + (prepositionhood + weight | subject)*

3.1.3 Discussion

The results reported in the previous subsection show that varying the properties of swh2 affects the acceptability of English MS.

WEIGHT

Notice also that the best-fitting ordinal logistic regression model for each sub-experiment is one without an interaction between PREPOSITIONHOOD and WEIGHT. Taken together, these results show that the prediction of Lasnik's (2014) analysis – namely, prediction [Wi] – is entirely incorrect. Not only does having a *heavy* swh2 decrease the acceptability (as opposed to increasing it, as Lasnik's analysis predicts), but also this effect is observed with DP and PP swh2s (as opposed to only DP swh2s, as predicted in Lasnik's analysis).

These results provide a new reason to doubt the plausibility of Lasnik's proposal that English MS is derived via rightward extraposition of swh2. These results, therefore, align with observations from the previous literature that Lasnik's proposal undergenerates English MS configurations and, thus, cannot be correct. For instance, Park & Kang (2007) point out that although ECM verbs do not permit rightward extraposition of the embedded subject (21), such phrases can function as swh2 in MS configurations (22). Similarly, Boone (2014) and Thoms (2016) show that although ditransitive verbs do not permit their indirect objects to undergo rightward extraposition (23), such phrases can function as swh2 in MS configurations (24).

(21) a. I believe [the politician with high profile in international affairs] to be dishonest.
b. * I believe to be dishonest [the politician with high profile in international affairs].
(adapted from Park & Kang 2007:398)

(22) One of the boys [believes behind one of the trees] to be the best place to hide, but I don't know which [behind which tree]. (adapted from Park & Kang 2007:398)

(23) * John gave a lot of money [the people that deserved it most]. (adapted from Thoms 2016:289)

(24) Some student gave some professor a lot of money, but I don't know which student [which professor]
(adapted from Boone 2014:110)

It should also be noted that the validity of Lasnik's argument regarding weight is questionable from the outset. When Lasnik notes that increasing the weight of a DP facilitates its rightward extraposition in non-elliptic environments, he is talking about increasing weight by adding information-structurally new material (e.g., in (25), the additional material *signed by Chomsky in 1975* is new). However, increasing the weight of swh2 in MS involves adding information structurally *given* (Schwarzschild 1999). To my knowledge, there are currently no experimental results about whether adding given material facilities rightward extraposition in non-elliptic environments. A native speaker of English I consulted informally reports that adding given material has no facilitatory effect (compare the B-examples in (26). I speculate that this situation arises because adding such material violates Williams' (1997) *Don't Overlook Anaphoric Possibilities!* constraint, which penalizes overlooked opportunities to employ pronominal items and hence rules out a rightward extraposition configuration for (26B). In short, Lasnik's argument about weight is dubious from the beginning because he compares unlike phrases, namely DPs composed of new information (in non-elliptic environments) and DPs composed of mostly given information (in MS environments).

(25) a. John saw yesterday in the shop window [DP books].
b. John saw yesterday in the shop window [DP books signed by Chomsky in 1975].

(26) a. A: I heard that the local shop has received a shipment of books.
B: * Yes, John saw yesterday in the shop window [DP books].

b. A: I heard that the local shop has received a shipment of books signed by Chomsky in 1975.
B: * Yes, John saw yesterday in the shop window [DP books signed by Chomsky in 1975].

Notice that, in my experiment, increased weight was equivalent to an increase in given material. Therefore, an alternative, processing-based explanation of the weight results is that repeating given material yields a penalty that translates into lower acceptability ratings (i.e., the more repeated material, the lower the acceptability rating). This explanation seems particularly plausible when one considers that a similar constraint – namely, Sag & Nykiel's (2011) *repeated name penalty* (see also Gordon, Grosz & Gilliom 1993) – has been postulated for single sluicing. Additional support for this proposal comes from the observation that the weight of swh1 also contributes to lowering acceptability. For instance, one may argue that the combined weight of swh1 and swh2 in the *heavy* condition in Sub-experiment 3 (i.e., when both swh1 and swh2 are both *heavy*) nullifies the potential ameliorating effect of swh2 being a PP, yielding a situation in which the MS configuration receives the same low rating regardless of whether swh2 is a PP or DP.

If the weight results do indeed reflect a repeated-word penalty, why was no significant difference observed between the WEIGHT levels *bare* and *explicit* in sub-experiments 2 and 3? One possible explanation is that a confounding factor – namely, the presence or absence of CONGRUENCE between swh1 and swh2 – affected acceptability ratings such that no significant effect of weight was observed. I discuss the effect of CONGRUENCE on acceptability in detail below.

PREPOSITIONHOOD

One observes that English MS utterances with PP swh2s are judged as significantly more acceptable than MS utterances with a DP swh2, which experimentally confirms the reports from the prior literature mentioned in Section 2.1.1, which are based on informally collected data. Therefore, prediction [Pi] is fulfilled. With Lasnik's (2014) rightward extraposition analysis discarded as a plausible explanation for why PREPOSITIONHOOD yields a main effect, Richards' (2010) *Distinctness*-based explanation, according to which MS configurations in which both sluiced wh-phrases are DPs are disfavored because they are indistinct (i.e., they bear the same syntactic category label and are linearly adjacent within the same Spell-Out domain), becomes a leading contender as a possible explanation for why prepositionhood yields a main effect on acceptability.

Unfortunately, there are independent reasons to doubt the plausibility of Richards' explanation. The principal reason is that his analysis does not extend to similar ellipsis configurations. Richards (1997) proposes that English gapping is syntactically generated in the same manner as English MS, by two phrases undergoing A′-movement to the left-periphery of the same Spell-Out domain. This Spell-Out domain is the clausal left-periphery in MS, and also the clausal left-periphery if gapping involves clausal coordination (Ross 1967; Jayaseelan 1990; among others) or the left-periphery of *v*P (or some similar functional thematic projection) if gapping involves *v*P coordination (e.g., Coppock 2001; López & Winkler 2003; Johnson 2009; see Potter, Frazier & Yoshida 2017 for a detailed discussion). Therefore, if gapping involves two linearly adjacent DPs in the same Spell-Out domain, this should yield a violation of Distinctness, just as MS purportedly does. Thus, one predicts that DP-DP gapping (such as (27)) is less acceptable than DP-PP gapping (such as (28)), in which no violation of Distinctness occurs. This prediction is not borne out, however. The 16 gapping fillers included in each of the three sub-experiments consistently show no significant difference in acceptability between the *nominal gapping* and *prepositional gapping* conditions.[14] An advocate of Richards' analysis could propose that Distinctness only affects wh-DPs, but such a proposal lacks any explanatory force. Additionally, this amended version of the Distinctness makes the incorrect prediction that *stripping*, which is an ellipsis configuration involving a non-wh elliptic phrase, should behave similarly to gapping in showing no sensitivity to the presence or absence of a preposition in the elliptic phrase. In reality, stripping has been claimed to show the same sensitivity as MS, see (29).

(27) On Halloween, Steve will wear a hat and William a mask. [gapping nominal]
(28) In July, David will vote for the greens and Lisa for the liberals. [gapping prepositional]

(29) a. ? Bill visited Jim yesterday after dinner, and John.
b. Bill talked to Jim yesterday after dinner, and **to** John. (adapted from Stolcke 1990:345)

Additionally, Richards' Distinctness analysis, which only affects functional projections, predicts that a DP followed by an adverbial should produce an acceptable outcome since adverbials do not fall under the phrases affected by

[14] The combined mean for the gapping conditions across the three sub-experiments is as follows: gapping nominal ($\bar{x}$=5.69) and gapping prepositional ($\bar{x}$=5.64). As mentioned in the main text, this difference is not statistically significant.

Distinctness. However, in a separate study,[15] I compared MS constructions containing DP-DP and DP-ADV combinations and found no significant results between both conditions, see (30).

(30) a. Everyone was reading something, but I just don't know who what. [mean DP-DP conditions=2.96]
b. Everyone was reading somewhere, but I just don't know who where. [mean DP-ADV conditions=3.00]

I view this as independent evidence that Richards' Distinctness condition overgeneralizes and should therefore be discarded as a possible explanation for why DP-DP English MS configurations are rated lower than their DP-PP counterparts.

An alternative, processing-based explanation for the prepositionhood results obtained in my experiment is that the presence of a preposition facilitates processing by providing hearers with a cue for how to retrieve the propositional meaning associated with the sluiced wh-phrases. Specifically, I propose that prepositions function as a cue for identifying the argument-structural status / thematic role of a sluiced wh-phrase within the inferred proposition, i.e., whether the wh-phrase is understood as associated with a subject, or direct or indirect object (or also other phrases associated with the verbal event of the inferred proposition). Moreover, it appears that only cues that help to discern the argument structure status of a sluiced wh-phrase facilitate processing and, therefore, positively affect acceptability judgments. By adding additional material within the sluiced wh-phrase (such as an NP restrictor with or without a PP modifier, e.g., *which student* (*of Chemistry*)), one increases the number of cues that the hearer may potentially use to determine with which correlate in the antecedent clause a sluiced wh-phrase should be associated, yet these cues are not actually utilized to facilitate processing. If they were, increasing the weight of a sluiced wh-phrase would improve acceptability ratings, contrary to observation.

This cue-retrieval analysis also has additional advantages. First, if the benefit of providing a cue outweighs the cost of increasing the weight of the sluiced wh-phrase, then this explains why, *ceteris paribus*, English MS configurations with PP swh2s are rated as more acceptable than English MS configurations with DP swh2s, despite swh2 in the former construction type being heavier, as it includes a *given* preposition. Second, this analysis provides a straightforward explanation for the effect of *P-omission* in English MS. Here, P-omission refers to when a sluiced DP wh-phrase is interpreted as associated with the DP complement of an overt PP in the antecedent clause. In such cases, it appears that a preposition is 'omitted' from the sluiced wh-phrase (see (31)). In English single sluicing configurations, P-omission usually does not negatively affect acceptability and, in some contexts, actually improves it (Merchant 2001; see Nykiel & Hawkins 2020 for details). However, in English MS, P-omission has a negative effect on acceptability, as in (32).

(31) John met [$_{PP}$ with someone], but I don't know [$_{DP}$ who].

(32) a. ? John talked to someone about something, but I just don't know to whom about what.
b. ?? John talked to someone about something, but I just don't know who(m) about what.
c. ?* John talked to someone about something, but I just don't know to who(m) what.
d. * John talked to someone about something, but I just don't know who(m) what.

I conducted an independent judgment study on this dataset and found that a significant difference in acceptability obtains between each construction type exemplified in (32), i.e., (32a) > (32b) > (32c) > (32d), with the following means on a 7-point Likert scale: a= 5.40, b= 4.14, c= 3.26, d= 2.98. Assuming that my results are accurate, this pattern of acceptability is straightforwardly explained if the processing burden is increased with every instance of P-omission due to one fewer cue being present each time. My results are also incompatible with Lasnik's (2014) analysis, which predicts that P-omission with swh1 should have no effect on acceptability, whereas P-omission with swh2 should have a substantial negative effect on acceptability. This incorrect prediction arises from his claim that swh1 undergoes standard wh-movement, which can ordinarily strand prepositions in English and therefore strands them in the ellipsis site in elliptic utterances (thus yielding their omission), while swh2 undergoes rightward extraposition, which cannot strand prepositions in non-elliptic contexts (see (33)). Note that the results of my independent study on the dataset in (32) are also incompatible with Richards' Distinctness-based analysis, which incorrectly predicts that no significant difference in acceptability obtains between (32a), (32b), and (32c), but that (32d) is significantly worse than (32a), (32b) and (32c). This prediction arises because (32a-c) obey the Distinctness condition, whereas (32d) does not. Furthermore, the significant difference

[15] This acceptability study followed a similar procedure to the other acceptability studies presented here. A total of 27 native speakers of English evaluated on a 7-point Likert scale 24 experimental items in an experiment comprised by a total of 96 items.

observed between (32b) and (32c) also aligns with a cue-based interpretation where cues for identifying the argument-structural status / thematic role modulate the acceptability since the second preposition plays a bigger role in re-establishing the argument structure, as it can be used to sub-set from for the information presented in swh1.

(33) a. * A linguist spoke about yesterday a paper on sluicing.
b. A linguist criticized yesterday a paper on sluicing. (Lasnik 2014:9)

INTERSPEAKER VARIATION

Recall that if prediction [Pii] is correct, one should observe substantial interspeaker variation regarding the acceptability of DP-DP and DP-PP English MS configurations, with some speakers judging DP-DP configurations as fully acceptable. In my experiment, out of the 94 speakers that took part across all three sub-experiments, only one participant can be viewed as considering MS "unimpeachable," with a score of ~6.5 for both DP-DP and DP-PP configurations. This corresponds to the standard item point A. Four speakers rated both configurations similarly with a score of ~6, and three more speakers rated the DP-PP configurations reasonably highly (~6) – which corresponds to the standard item point B but rated the DP-DP condition much lower, between 4.5 and 5 – corresponding to a rating between standard item point B and C. Seven more participants rated either one or both conditions higher than 5. Hence, in my study, only one speaker considered MS fully acceptable, while a maximum of 15% of the cohort considered (some instance of) MS acceptable but marginally degraded. Based on these figures, it appears that, although some speakers do judge DP-DP configurations as fully acceptable, such speakers belong to a very small minority. An anonymous reviewer suggests that one's attention should be focused on providing theoretical analyses of the grammar of these speakers, as researchers should concentrate on the most permissible grammars of a given language community. I respectfully disagree that researchers should focus exclusively on judgment data representative of only ~1% of a language community. Firstly, there is no *a priori* causal link between consistently rating a datapoint exceptionally highly and possessing a more permissible grammar. Such speakers might instead possess the same grammar as the remaining ~99% of their cohort but possess exceptional online processing abilities, or conversely, they potentially even misunderstood the task. If the factors influencing the acceptability of English MS configurations have their roots in processing (a position I will endorse in Section 4), then this alternative scenario appears quite plausible for the current results. Nevertheless, even if it were true that exceptionally higher ratings always equated with a more permissible grammar, I disagree that researchers should concentrate on the most permissible grammars if such grammars only represent ~1% of the language community in question. Doing so leaves ~99% of that community's grammar unanalyzed and is therefore tantamount to undertaking idiolectal research.

The data of these experiments also show a tremendous interspeaker variation. While in its majority, speakers show a preference for the instances of MS where the non-initial wh-element is a prepositional phrase, it is not the case that two types of speakers can be differentiated: the ones that accept MS and the ones that do not. Additionally, the number of speakers whose judgment for <DP, DP> multiple sluices could be considered grammatical is minimal, and those speakers still show a preference for the <DP, PP> combination (see Kotek & Barros 2018 for a discussion of interspeaker variation).

Regarding whether there is a substantial amount of interspeaker variation for English MS judgments more generally, one must compare English with another language to obtain a baseline. Answering this question, therefore, provides additional motivation for introducing a cross-linguistic comparative aspect to this study by re-running (a version of) Experiment 1 in German. I return to discuss interspeaker variation after reporting the results of the German experiment in Section 3.2.3.[16]

CONGRUENCE

Suppose that the factors under investigation in this experiment interfere with each other regarding the extent to which they can negatively affect acceptability, with WEIGHT as the greatest hinderer. With these assumptions in place, one can make sense of the effect of CONGRUENCE in the English experiments. Because having a *heavy* swh1 already lowers acceptability, the presence of congruence in sub-experiment 3 (where the *heavy* conditions are *congruent*) lessens the detrimental effects of weight. Hence, we observe very similar ratings for the *heavy* condition across all three sub-experiments.

[16] For the plots with the descriptive statistics for the interspeaker variation observed in each of the experiments reported in this article, see https://osf.io/5p4ax/?view_only=200c5df19bba42f3888496b2c4ab9b67

Conversely, because having a *bare* swh1 does not lower acceptability (relative to having a *heavy* swh1, at least), the absence of congruence in sub-experiment 1 has a more noticeable effect on ratings. Indeed, one observes that when congruence obtains for the *bare* condition (i.e., in Sub-experiment 1), the *bare* condition is rated as significantly more acceptable than the next-weightier condition, namely, the *explicit* condition. But when congruence fails to obtain for the *bare* condition (i.e., in sub-experiment 2 and 3), the *bare* condition drops in acceptability such that a significant difference no longer obtains between it and the *explicit* condition. This provides a possible explanation for why no significant difference for weight was observed between the *bare* and *explicit* conditions in sub-experiments 2 and 3, despite such a difference being predicted if increased weight is tantamount to more violations of a repeated-words penalty.

A similar but statistically insignificant pattern holds for the *explicit* condition, which is rated as more acceptable when congruent (Sub-experiments 2) than when incongruent (Sub-experiments 1 and 3). If this interpretation of the results is on the right track, then congruence does indeed influence the acceptability of English MS, thus showing preliminary support for prediction [Wii]. Alternatively, the difference observed across sub-experiments could be triggered by having D(iscourse)-linked wh-words.[17] In opposition to the *congruent* conditions in Sub-experiment 1, the *congruent* conditions in the other sub-experiments involve D-linked wh-words, and this could be responsible for provoking additional effects that mask or nullify the congruence effect.

One might wonder whether the effect of CONGRUENCE is actually just an effect of WEIGHT in disguise. Could it be the case that a negative correlation obtains between overall weight – i.e., the combined weight of swh1 and swh2 – and acceptability? The results for the *explicit* condition speak against this idea. Despite the overall weight being greater for the *explicit* condition in Sub-experiment 2 (where both sluiced wh-phrases were *explicit*, e.g., *which doctor which nurse*) than in Sub-experiment 1 (where swh1 was *bare* and swh2 was *explicit*, e.g., *who which nurse*), acceptability was slightly higher for the *explicit* condition in Sub-experiment 2 than in Sub-experiment 1 (follow also the link Footnote 15 for null differences obtained between congruent *bare* and *explicit* conditions). Thus, although CONGRUENCE is related to weight insofar as it involves varying the weight of swh1, it appears that CONGRUENCE should actually be considered as a factor independent of WEIGHT. Therefore, I tentatively propose that an absence of congruence lowers acceptability because of the processing burden associated with establishing a pair-list or single-pair dependency between a set of specific items (e.g., the set of doctors mentioned in the antecedent) and a set of non-specific items (e.g., all humans).

Finally, it should be reported that the experimental results presented here show that the overall mean ratings for English MS configurations are in the 4 range, which translates to a rating equating to standard item C (see Gerbrich, Schreier & Featherston 2019). These results indicate that MS in English is indeed a marked construction, regardless of the individual amelioration/degradation effects that some factors contribute.

To summarize: the results of the experiment on English MS count strongly against Lasnik's (2014) rightward extraposition analysis. Although the results regarding PREPOSITIONHOOD are compatible with Richard's *Distinctness*-based explanation, there are independent reasons to reject this analysis. I opted to advocate a processing-based explanation, according to which the presence of a preposition facilitates processing because prepositions function as a cue for determining the argument structure / thematic role of the sluiced wh-phrases within an inferred proposition.

Because increased weight is equivalent to an increase of given material, the results WEIGHT are compatible with the idea that English MS is subject to a repeated-word penalty condition, whereby each repetition of given material in the sluiced wh-phrases incurs a penalty that is reflected in lowered acceptability judgments. The results on the effect of congruence are compatible with the idea that an absence of congruence lowers acceptability judgments relative to the negative effects induced by other factors, such as increased weight. The results also show that CONGRUENCE is not merely a reflection of increased weight but an independent factor. I suggested that the absence of congruence might incur a

---

[17] Discourse-linked in the sense of (Pesetsky 1987, 2000). Therefore, the experimental conditions *explicit* and *heavy* wh-phrases fall under this category. D-linked wh-phrases are claimed to ameliorate some other syntactic constructions (e.g., Culicover 2013; Kroch 1998; Rizzi 1990). I wish to thank one of the reviewers for drawing my attention to these ameliorations. One such amelioration happens under *superiority effects* (i.e., when the wh-phrase cross their paths). My present study has not been concerned with those cases. However, MS has been claimed to respect superiority (Abels & Dayal 2017, to appear). In any case, I will venture to say that even if MS must obey superiority, when forcing a superiority-violating configuration, D-linked wh-phrases do seem to mitigate the unacceptability, see (i). Those judgments are hard to tell since the sentences are already very degraded. I leave this matter for future research.

(i) a. ** Everyone complained about something, but I don't know about what who.

(ii) b. ?* Every student complained about some book, but I don't know about which book which student.

processing burden, which arises from the increased processing load required to derive a pair-list or single-pair dependency between a specific wh-phrase and non-specific wh-phrase.

The results of this experiment demonstrate conclusively that MS is a marked configuration for all but a very small minority of native speakers, and those results have suggested that there exists a substantial interspeaker variation observed in the judgments. To determine whether this variation spreading across the whole acceptability scale – namely, means ranging from ~1.1 to ~6.6 – is particular to English MS, the experiment should be re-run on another language, so a cross-linguistic comparison of interspeaker variation can be undertaken. For practical and conceptual reasons, I chose to re-run the experiment in German. Conceptually, German has similar wh-movement properties to English and derives matrix multiple wh-questions like English. Additionally, German has a rich case-marking morphology, meaning that we can isolate the potential amelioration caused by the presence of a preposition in the non-initial wh-remnant from the betterment presented by inherently case-marked nominal wh-phrases. Moreover, MS in German has been reported as overall better than MS English, but as we will learn, it also is not flawless. Crucially, an investigation on German MS will let us test to see if we can obtain cross-linguistic support for the cue-retrieval analysis.

3.2 Experiment: Multiple sluicing in German

Recall that the acceptability of English MS drops when both sluiced wh-phrases are nominal expressions. This situation does not obtain in German, in which DP+DP MS configurations are reported as acceptable (34) (Merchant 2006; Richards 2010; Winkler 2013). Similarly, there are no previous reports (known to me) that the weight of swh2 affects the acceptability of German MS.

(34) Jemand hat [et]was gesehen, aber ich weiß nicht, wer was. (Merchant 2006:285)
someone has something seen but I know not who what
'Someone has seen something, but I don't know who what.'

3.2.1 Methods

Design and Materials

This experiment is similar to the sub-experiments on English described in Section 3. This experiment compared the effect on the acceptability of the presence or absence of a preposition in swh2 and the weight (in terms of the number of heads following the wh-element) of swh2. I constructed 24 items containing multiple sluicing using a 2×2 design with PREPOSITIONHOOD and WEIGHT as within-item and within-subject factors. Both factors modify swh2. PREPOSITIONHOOD included two levels; *+P* and *–P*; WEIGHT was also controlled to represent two levels: *bare* and *explicit*. In contrast with the English experiment reported in Section 3, I decided to exclude *heavy* swh2s from this experiment. This level was excluded because weight is no longer a primary focus of my investigations, and it was consistently shown in English that *heavy* conditions have a detrimental effect on the acceptability of MS. Congruence was instead controlled for by ensuring that congruence was obtained between swh1 and swh2 in every test utterance. Furthermore, to test how differently case-marked swh2s behave, half of the items included a combination of animate subject and inanimate object, i.e., *wer was* ('who.NOM what.ACC'), while the other half comprised the combination of animate subject and animate object, e.g., *wer wen* ('who.NOM who.ACC'). I will refer to the former as having a swh2 with *covert* case marking and the latter with *overt* case marking. A sample item is provided in Table 9.

Table 9: Sample items for German multiple sluicing study

Sentences	Condition
Am Montagmorgen hat jeder etwas gelesen, aber ich weiß nicht, wer was. 'On Monday morning, everyone read something, but I don't know who what.'	[-P/bare]
Am Montagmorgen hat jeder an etwas gedacht, aber ich weiß nicht, wer an was. 'On Monday morning, everyone thought about something, but I don't know who about what.'	[+P/bare]
Am Montagmorgen hat jede Studentin ein Buch gelesen, aber ich weiß nicht, welche Studentin welches Buch. 'On Monday morning, every student read some book, but I don't know which student which book.'	[-P/ explicit]

Sentences	Condition
Am Montagmorgen hat jede Studentin an ein Buch gedacht, aber ich weiß nicht, welche Studentin an welches Buch. 'On Monday morning, every student thought about some book, but I don't know which student about which book.'	[+P/explicit]

All participants saw a total of six items in each condition. In addition to the 24 critical items, 72 fillers were included. Fifteen out of those filler sentences corresponded to the 5-degree standardized items from Featherston (2009), which range from A-type fillers (most natural) to E-type fillers (least natural). Items were distributed across four lists according to the Latin square design.

Participants and Procedure

The procedure used in this experiment is the same as the for the experiment series for MS in English presented above, except that all the instructions in the experiment were presented in German. Forty native German speakers were recruited via the *Prolific* platform, and they were all naive with respect to the purpose of the experiment. Participants were asked to rate the naturalness sentences on a 7-point scale, from 1 (very unnatural) to 7 (very natural), and they were also instructed that there is no "right" or "wrong" answer and asked to follow their intuitions. Participants received monetary compensation of £3.50 for their participation in this study, which lasted approximately 15 minutes. Based on the judgments given by the participants to the standardized items (Featherston 2009), ten participants were excluded from the statistical analysis. Finally, before starting to judge the experimental items, every participant had a practice round with five sentences. Overall, a total of 96 items were presented in each trial, and within a session, critical items and filler sentences were ordered randomly.

Predictions

The predictions of the German experiment are based on the analysis developed in the previous section. Recall that this analysis states that processing MS is facilitated by the presence of cues in a sluiced wh-phrase concerned with the argument structural / thematic status of the sluiced wh-phrase in the inferred proposition. Such cues include prepositions and morphological case marking. The simplest cue-based model is an additive one, which assumes that every relevant cue present in an MS configuration will increase acceptability by some observable amount. For now, I adopt this assumption. My current analysis also states that increasing weight lowers acceptability, as increasing weight involves introducing *given* material into a sluiced wh-phrase, which is penalized. However, the presence of congruence can lessen the detrimental effect of weight.[18] Based on this analysis, the predictions for my experiment on German MS are as follows:

(35) **Predictions regarding PREPOSITIONHOOD**

[Pi] German MS configurations with PP swh2s are significantly more acceptable than German MS configurations with DP swh2.

[Pii] Both DP-DP and DP-PP German MS configurations are significantly more acceptable than their English counterparts.

(36) **Prediction regarding WEIGHT**

[Wi] When controlled for congruence, German MS configurations with *bare* swh2s do not show significant differences in acceptability compared to configurations with *explicit* swh2s.

(37) **Prediction regarding CASE**

[Ci] German configurations whose swh2 displays *overt* case are significantly more acceptable than German MS configurations whose swh2 displays *covert* case.

[Pi] arises because German PP swh2s contain two relevant cues (namely, a preposition and morphological case marking on the wh-word). In contrast, German DP swh2s contain only one (namely, morphological case marking on the wh-word). For every condition, [Pii] is predicted since there are always more cues in German than English: in the DP swh2

[18] In retrospective, including the *heavy* condition in the German experiment would have allowed for a more complete picture of the data. I leave this matter for further research and tentatively argue that the results for *heavy* German MS would be parallel to those found in English.

cases, there are no cues for English and one cue for German, whereas, in the PP swh2 cases, there are two cues for German and one cue for English. [Ci] arises because sluiced wh-phrases with covert case (i.e., the same morphological realization of abstract case for NOM, ACC, and DAT) do not provide the hearer with an explicit cue to the wh-phrase's argument structural status in the implied proposition, whereas sluiced wh-phrases with overt case do. Note that, because all conditions in this experiment are *congruent*, WEIGHT can be isolated from other factors without fear of introducing confounds.

3.2.2 Data analysis and results

I analyzed the judgment data using the same procedure as described in Section 3.1.2. Here, I report the model with the best maximal fixed and random effect structure supported by the data. The corresponding formula is also included in Table 10. Finally, the *p*-values are estimated by the Wald test.

Figure 5 shows the mean acceptability ratings obtained for this experiment. I present the results of its statistical analysis in Table 10. The model yielded no main effect for PREPOSITIONHOOD and no main effect for WEIGHT.[19] Additionally, I computed a pairwise comparison using the more conservative posthoc *Tukey* test, which also did not reveal any significant result: difference between the levels *nominal bare* and *prepositional bare* (*z*-ratio = -1.706; *p*-value = 0.3205) and between levels *prepositional bare* and *prepositional explicit* (*z*-ratio = 2.487; *p*-value = 0.0619).

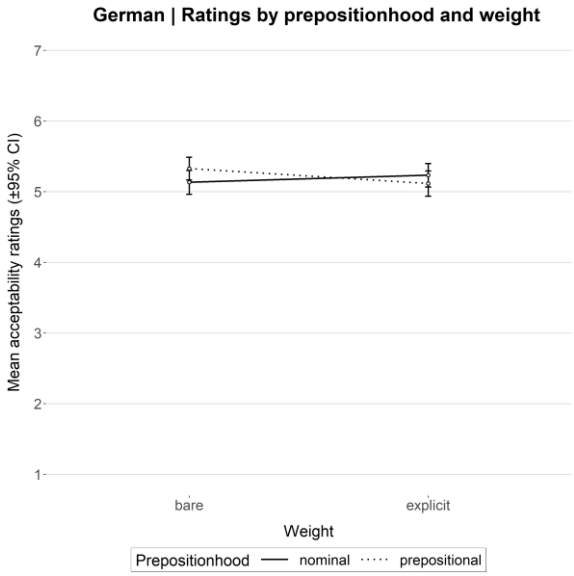

Figure 5: Mean acceptability rating (n = 30). Error bars show standard error.

Table 10: Cumulative Link Mixed Model fitted with the Laplace approximation

	Estimate	Std. error	z value	Pr(>|z|)
PREPOSITIONHOOD*(prepositonal)*	0.4742	0.2508	1.890	0.0587
WEIGHT*(explicit)*	0.2307	0.2014	1.145	0.2521
PREPOSITIONHOOD:WEIGHT	-0.7029	0.2855	-2.462	0.0138 *

Formula: *rating ~ prepositionhood * weight + (1 | subject) + (prepositionhood | item)*

[19] The interactions between those two factors did not reach significance. However, it is important to mention that both the main effect for PREPOSITIONHOOD and the interaction were marginally significant. Further research should decide if those results would consolidate with more statistical power or, on the contrary, if, as reported here, do not represent differences for the acceptability.

3.2.3 Discussion

PREPOSITIONHOOD / CASE

Because the results yielded no main effect for PREPOSITIONHOOD, prediction [Pi] was not borne out. Therefore, the simple additive version of the cue-based theory cannot be maintained. This null difference is instead compatible with a "diminishing returns" cue-based theory, according to which diminishing processing benefits are conferred by each additional cue of the same type. Therefore, relative to the benefits granted by the presence of the first cue, the presence of a second cue (in the case of the *prepositional* conditions) provides only a small additional processing advantage, if any. This "diminishing-returns" version of the cue-based theory, therefore, explains why the presence of a preposition raised the acceptability of German MS, but only by an insignificant amount: having two cues (the preposition and the wh-word's case-marking) as opposed to one (only the wh-word's case marking) confers only a slightly increased processing advantage, rather than double the processing advantage.

This "diminishing-returns" cue-based approach yields the prediction that German MS configurations with swh2s with *overt* case are significantly more acceptable than their counterparts with *covert* case, but only when swh2 is DP. This prediction arises because a relevant cue is always present in the PP condition (namely, a preposition), thus rendering the presence or absence of an additional similar-functioning cue irrelevant, yet a relevant cue is not always present in the DP condition: instead, the case-marked wh-word of swh2 functions as the only relevant cue in the *overt* condition but, in the *covert* condition, no relevant cue is present whatsoever. However, the experiment results showed that CASE yields no main effect and no interaction with PREPOSITIONHOOD: covertly marked sentences ($\overline{x}$ = 5.51) were judged to be more acceptable than overtly case-marked sentences ($\overline{x}$ = 5.19), but this difference was insignificant (z-value = 1.142; p = .254). This means that the above-described [Ci] prediction is not borne out.

A possible explanation for this result is that independent cues are available, which offset the decrease in acceptability one expected to observe when the wh-word's case is wholly syncretic and, therefore, a poor cue to the wh-word's thematic role). For instance, one might entertain the idea that linearity functions as a processing cue in MS configurations. According to this idea, the fact that the swh2 *was* ('what') linearly follows *wer* ('who.NOM') is itself a cue to *was*'s argument structure status (i.e., *was* must either be the direct or indirect object). However, for this analysis to be feasible, one requires a supplementary explanation for why, in English – a language in which wh-phrases are similarly case-poor – linearity fails to mitigate the adverse effects of P-omission in DP-DP MS configurations. I appeal to the diverging cross-linguistic possibilities for P-omission under sluicing to explain this difference. In a language that generally disallows P-omission under sluicing, such as German, the bare wh-DP *was* in an MS configuration such as '… *wer was'* can only be interpreted as either a direct or indirect object. However, in a language that generally allows P-omission under sluicing, such as English, the bare wh-DP *what* in an MS configuration such as '… *who what'* could be interpreted as a direct object, indirect object, or the complement of an omitted preposition inside PP argument or a PP modifier. In the English case, linearity fails to sufficiently reduce the hypothesis space for resolving a DP swh2 to a particular thematic role.

In addition, German also has single-word interrogative pronominal adverbs (so-called *Pronominaladverbien,* or more specifically *Interrogativeadverbien* (Negele 2012)) such as "womit", "wofür", "worauf", etc. The combination of a wh-phrase with a pronominal adverb, e.g., *wer wofür* seems to produce a better MS outcome than two wh-phrases *wer für was*.[20,21] If this is true, I speculate that those interrogative pronominal adverbs present more cues than either case-marked nominal or prepositional swh2s, whereby the ratings for those conditions would call for a larger difference (potentially solidifying the marginal effect observed for PREPOSITIONHOOD).

Overall, the mean ratings for German MS utterances are slightly higher than 5, which correlates with a score between standard items B and C (see Featherston 2009). This is one point / one standard item category more acceptable than English MS configurations, a result that confirms [Pii] and which, therefore, supports the cue-based analysis being advocated here.

---

[20] I would like to thank Gisbert Fanselow for this comment and the audience that shared their intuitions about this observation during my presentation in the *SynSem Colloquium* at the University Potsdam.

[21] The same seems to hold true for Afrikaans. Many thanks to Sulene Pilon for pointing this out and to staff in the Afrikaans department at the University of Pretoria for corroborating those intuitions.

WEIGHT

The results between *bare* and *explicit* conditions did not yield a significant difference, indicating that WEIGHT per se does not improve acceptability. Thus prediction [Wi] is not borne out. As mentioned above, this could be due to the fact that *explicit* nominals do not introduce any relevant cue beyond case-marking since the additional information presented by the explicitness does not part-take in the argument structure assignment. In the English experiment, we have observed that weight does indeed negatively affect acceptability. However, the presence of congruence lessens the detrimental effects of weight.

Furthermore, thanks to the case-marking, German could be used to test explicitness without additional weight. This can be achieved by applying N-deletion to the *explicit* nominal wh-phrases. While in English such N-deletion incurs a homonymic clash (Bolinger 1978) "*which which*", German allows such combinations, e.g., "*welcher über welchen*" ('which.NOM about which.ACC') or "*welcher mit welchem*" ('which.NOM with which.DAT'). If indeed weight has a detrimental effect that goes beyond congruence, then a difference is expected in German MS between the *explicit* cases presented in the experiment and *explicit* cases that undergo an additional N-deletion. I call for future research on this issue.

INTERSPEAKER VARIATION

Following the prediction and subsequent findings in English MS, one would expect German to disclose a similar substantial interspeaker variation. As mentioned above, on average, both German DP-DP and DP-PP instances of MS received higher ratings than their English counterparts. Overall, conditions in German MS received a score of ~5 – which corresponds to a rating between the standard item points B and C – this is in sharp contrast with the ratings obtained for English MS, whose ratings averaged between the standard item points C and D. From the 30 participants admitted to the analysis, seven (23%) rated MS sentences highly (mean of 6 or above), and another 19 participants (63%) gave ratings between 5 and 6. These results show that, as predicted by the literature, MS in German is more readily accepted than in English. Lastly, recall that the mean results of English MS range from ~1.1 to ~6.6. In opposition, the range of mean for German MS is considerably smaller: ~3.4 to ~6.7. Therefore, we can conclude that English shows a much larger interspeaker variation than German.

In a nutshell, the results of the experiment on German MS regarding PREPOSITIONHOOD are compatible with Richards' (2010) *Distinctness*-based explanation. Despite Distinctness allowing DP-DP combination for German, one would expect that more distinct phrases, i.e., DP-PP, or even DP phrases with *overt*/*covert* case-marking or *animate*/*inanimate* combination (cf. Gallego 2017) would contribute to Distinctness and consequently improve the acceptability. Since such amelioration is not observed and given the independent argument presented against Distinctness in Section 3.1.3, it is safe to disregard this analysis. I believe the processing-based explanation using cue-retrieval information can also capture the results obtained for German MS. Relevant to this analysis is the fact that the presented cues help determine the argument structure / thematic role of the sluiced wh-phrases within an inferred proposition. In particular, I advocate for a "diminishing-returns" cue-based approach, where the presence of an additional similar-functioning cue makes no relevant contribution to the processing.

I have viewed the null effect observed for WEIGHT to be modulated by CONGRUENCE and compatible with a repeated-word penalty condition. Although under this latter condition, a decrease in acceptability is expected for *explicit* cases, the fact that *explicit* conditions in this experiment are *congruent* and that, as discussed above, D-linked wh-expression allow for otherwise unacceptable collocations lessens the penalty introduced by repeating given material. I have conjectured that *heavy* conditions would indeed incur a visible repeated-word penalty affecting the acceptability, as well as hypothesizing that explicitness without additional weight – obtained by applying N-deletion – produces more acceptable MS instances in German.

The experimental results show that MS in German receives consistently higher ratings than English MS. Additionally, the results suggest that German speakers display a much smaller interspeaker variation. Moreover, it should be reported that the experimental results presented here show that the overall mean ratings for German MS configurations are in the 5 range, which translates to a rating equating to standard item C (see Featherston 2009). While this is one whole point higher than English, MS in German, although acceptable in general terms, should also be viewed as a marked construction. Furthermore, I believe it is worth mentioning that German, unlike English, allows for wh-subject and wh-

object words to be adjacent to each other in contexts such as verb-final embedded clauses. The fact that this collocation is possible could increase the observed overall acceptability for multiple sluicing in German.

Last, a major contribution of the present paper is the cross-linguistic comparison. In order to provide further evidence to the cue-based approach presented in the discussion of the English data, analyzing a language with inherent cues in wh-nominal seemed clearly motivated. Thus, German multiple sluicing made the perfect case. We have observed that German differs from English with respect to prepositionhood. This is indeed the expected behavior, given that German nominal wh-phrases already present cues for determining the argument structure / thematic role. Furthermore, the other major difference between the two languages is that German speakers rate MS generally as more acceptable, and we have observed that the range of judgments is more homogeneous in German than in English: hence, the former disclosing a smaller interspeaker variation.

4 Conclusion

The literature on multiple sluicing in English has been characterized for identifying different factors that modulate the acceptability of MS in this language. One crucial observation has been the distinction between multiple sluices that contain two nominal expressions and those encompassing a nominal and a prepositional phrase. Despite recognizing the marginality of MS in English and the great interspeaker variation, most authors have acknowledged the amelioration in acceptability under the presence of a preposition in the non-initial sluiced wh-phrase. The current study has extended the analysis of multiple sluicing, including an experimental approach to provide empirical evidence that prepositionhood is indeed an ameliorating factor in multiple sluicing in English. I have argued that Richards' (2010) Distinctness rule, a priori, seems to capture the acceptability differences between <DP, DP> and <DP, PP> kinds of multiple sluicing. However, this rule cannot account for the acceptability contrast observed for wh-phrases of different weights. Moreover, the null difference observed in the gapping data with two DP remnants in English suggests that the Distinctness rule both under- and over-generates. I have proposed that one way of understanding the properties observed for the investigated factor in the two analyzed languages resides in the number of cues the sluiced wh-phrase provides to retrieve the information from the antecedent (Martin & Mcelree 2008, 2011; Nykiel, Kim & Sim 2022). Additionally, I have suggested that those cues must enable the unambiguous assignment of syntactic and morphological properties to the sluiced wh-phrases concerning, for example, argument structure, theta-roles, and case-marking. This assignment of the relevant features to the poorly case-marked English language could be the reason for a strong dispreference when generating two adjacent DPs in a multiple sluicing configuration due to the lack of retrieval-cues. Conversely, the German data suggests that the cue retrievability approach makes the right predictions since, thanks to the overt case-morphology, wh-words in German contain enough cues to enable the necessary retrieval of information from the antecedent. Nevertheless, I have presented the "diminishing-returns" cue-based approach idea to capture the null difference between nominal and prepositional MS cases in German, arguing that two cues providing the same information about how to assign the argument structure do not have an additive meaning.

The other main characteristic of MS analyzed in this paper is weight, as counted by the number of heads succeeding the *wh*-phrase, and whether *heavy* swhs produce more acceptable MS sentences than *bare* swhs, as suggested by Lasnik (2014). The three sub-experiments presented for English show the opposite result, i.e., heavier *wh*-phrases produce less acceptable sentences. Lasnik's argument that MS in English is fed by standard wh-fronting of the initial swhs and rightward extraposition of the non-initial swhs is based on the alleged similar behavior the non-initial swhs displays compared to rightward extraposed constituents, viz. being a PP or a heavy DP. Besides the arguments reviewed from the literature against analyzing MS as involving rightward movement of the second remnants, the results presented here show that the kind of *heavy* swhs that Lasnik claims to ameliorate MS have the opposite effect and that *heavy* swhs deteriorate the overall acceptability of this construction. This deterioration has been argued to be caused by the repetition of given material carrying nuclear stress in the PP modifier of the *heavy* swhs. No such degradation in acceptability was observed between the levels *bare* and *explicit*. For one, because the nuclear stress falls in the nominal head of the wh-phrase and not in the nominal head of the modifier. Additionally, I have argued that the null difference observed between those two levels is a direct effect of congruence. The effect of congruences can also be observed in the German results, where weight is not an improving factor between *bare* and *explicit* conditions. These results show that nominal expression in a wh-phrase, even if explicit, do not provide enough relevant cue to enable retrieval beyond the cue information introduced by the case morphology.

The direct cross-linguistic comparison presented here has shown that English MS configurations receive consistently lower acceptability ratings than their German counterparts, thus making German MS a more productive construction than in English. However, neither language has rated MS sentences as 'very acceptable' (i.e., on par with standard items A), suggesting that those constructions are still marked. An important difference between the two languages analyzed here is the fact that English speakers disclose a much greater interspeaker variation than German speakers.

This paper has presented data on multiple sluicing from an experimental point of view. Some of the recurrent issues in the field have been discussed here; however, the intricate puzzle of multiple sluicing still requires much further research.

Acknowledgments: I wish to thank two anonymous reviewers for their insightful comments and recommendations. Many thanks to Larissa Specht and Edith Scheifele for helping me with the data analysis. Also, I would like to thank Alena Hieber and Anja Haubelt for their help in creating the experimental material. Special thanks to James Griffiths for helpful comments and discussion on previous stages of this research. All remaining errors are naturally my own. This research was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 75650358 – SFB 833.

References

Abels, Klaus. 2019a. Movement and Islands. In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford Handbook of Ellipsis*, 389–424. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198712398.013.17.

Abels, Klaus. 2019b. On "sluicing" with apparent massive pied-piping. *Natural Language & Linguistic Theory* 37(4). 1205–1271. https://doi.org/10.1007/s11049-018-9432-1.

Abels, Klaus & Veneeta Dayal. On the Syntax of Multiple Sluicing and What It Tells Us about Wh -Scope Taking. (Advance publication). 1–49. https://doi.org/10.1162/ling_a_00448.

Abels, Klaus & Veneeta Dayal. 2017. On the Syntax of multiple sluicing. In Andrew Lamont & Katerina A. Tetzloff (eds.), *North East Linguistics Society (NELS)* (47), 1–20. Amherst, MA: GLSA publications.

Aelbrecht, Lobke. 2010. *The syntactic licensing of ellipsis* (Linguistic Aktuell/Linguistics Today 149). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/la.149.

Bai, Xue & Daiko Takahashi. 2022. Pair-list Interpretation in Multiple Sluicing in Mandarin Chinese. Ms. Tohoku University. 1–17.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Barros, Matthew & Robert Frank. Attention and Locality: On Clause-Boundedness and Its Exceptions in Multiple Sluicing. (Advance publication). 1–36. https://doi.org/10.1162/ling_a_00458.

Barros, Matthew & Robert Frank. 2016. Discourse Domains and Syntactic Phases: A Constraint on Long-Distance Multiple Sluicing. Handout presented at NYU Syntax Brown Bag.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Bolinger, Dwight. 1978. Asking more than one thing at a time. In Henry Hiz (ed.), *Questions*, 107–150. Dordrecht: Springer Netherlands.

Boone, Enrico. 2014. *The syntax and licensing of Gapping and Fragments*. University of Leiden dissertation.

Chang, Da-eun & Jeong-Seok Kim. 2013. Multiple sluicing in disguise. *Studies in Modern Grammar* 75. 123–144. https://doi.org/10.14342/smog.2013.75.123.

Chiu, Liching Livy. 2007. A Focus-Movement Account on Chinese Multiple Sluicing. *Nanzan Linguistics* 1(1). 23–31.

Chomsky, Noam. 1975. *The logical structure of linguistic theory*. New York: Springer.

Christensen, Rune Haubo B. 2019. ordinal – Regression Regression Models for Ordinal Data. R package version 2019.12-10. https://cran.r-project.org/web/packages/ordinal/ordinal.pdf.

Chung, Sandra. 2013. Syntactic Identity in Sluicing: How Much and Why. *Linguistic Inquiry* 44(1). 1–44. https://doi.org/10.1162/ling_a_00118.

Chung, Sandra, William A. Ladusaw & James McCloskey. 1995. Sluicing and logical form. *Natural Language Semantics* 3(3). 239–282. https://doi.org/10.1007/BF01248819.

Chung, Wonil & Myung-Kwan Park. 2017. Multiple Sluicing and SWIPING Meet RT-based Experimental Syntax. *Studies in Linguistics* 42. 115–141.

Coppock, Elizabeth. 2001. Gapping: In defense of deletion. In Mary Andronis, Christopher Ball, Heidi Elston & Sylvain Neuvel (eds.), *Chicago Linguistic Society (CLS)* (37), vol. 1, 133–148. Chicago, IL: The Chicago Linguistic Society.

Cortés Rodríguez, Álvaro. 2021. Multiple adjacent wh-interrogatives in Spanish. Paper presented Presentation at the II. Encuentro de Lingüística Formal en México. Benemérita Universidad Autónoma de Puebla. Paper presented Presentation at the II. Encuentro de Lingüística Formal en México. Benemérita Universidad Autónoma de Puebla, Puebla, Mexico (Online). https://doi.org/10.13140/RG.2.2.29854.41283.

Cortés Rodríguez, Álvaro. 2022. Multiple sluicing and islands: a cross-linguistic experimental investigation of the clausemate condition. *The Linguistic Review* 39(3). 425–455. https://doi.org/10.1515/tlr-2022-2093.

Corver, Norbert & Henk Van Riemsdijk. 2001. Semi-lexical categories. In Norbert Corver & Henk Van Riemsdijk (eds.), *Semi-Lexical Categories: The Function of Content Words and the Content of Function Words*, 1–20. Berlin/New York: de Gruyter Mouton.

Culicover, Peter W. 2013. The role of linear order in the computation of referential dependencies. *Lingua* 136. 125–144. https://doi.org/10.1016/j.lingua.2013.07.013.

Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler Syntax*. Oxford University Press.

Drummond, Alex, Norbert Hornstein & Howard Lasnik. 2010. A Puzzle About P-Stranding and a Possible Solution. *Linguistic Inquiry* 41(4). 689–692. https://doi.org/10.1162/LING_a_00018.

Embick, David & Rolf Noyer. 2007. *Distributed Morphology and the Syntax—Morphology Interface*. (Ed.) Gillian Ramchand & Charles Reiss. *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199247455.013.0010.

Farkas, Donka F. & Anastasia Giannakidou. 1996. Anastasia Giannakidou. In Teresa Galloway & Justin Spence (eds.), *Semantics and Linguistic Theory (SALT)*, 35–52. Ithaca, NY: Cornell University.

Featherston, Sam. 2009. A scale for measuring well-formedness: Why syntax needs boiling and freezing points. In Sam Featherston & Susanne Winkler (eds.), *The Fruits of Empirical Linguistics*, vol. 1, 47–74. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110216141.47.

Fox, Danny & David Pesetsky. 2003. Cyclic Linearization and the Typology of Movement. Ms, MIT.

Gallego, Ángel J. 2017. Multiple Wh-Movement in European Spanish. In Olga Fernández-Soriano, Elena Castroviejo & Isabel Pérez-Jimenez (eds.), *Boundaries, Phases and Interfaces. Case studies in honor of Violeta Demonte*, 196–222. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/la.239.10gal.

Gerbrich, Hannah, Vivian Schreier & Sam Featherston. 2019. Standard items for English judgment studies: Syntax and semantics. In Sam Featherston, Robin Hörnig, Sophie von Wietersheim & Susanne Winkler (eds.), *Experiments in focus: Information structure and semantic processing*, 305–328. Berlin: de Gruyter.

Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*. Stanford, CA: CSLI Publications.

Gordon, Peter C., Barbara J. Grosz & Laura A. Gilliom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science* 17(3). 311–347.

Grano, Thomas & Howard Lasnik. 2018. How to neutralize a finite clause boundary: Phase theory and the grammar of bound pronouns. *Linguistic Inquiry* 49(3). 465–499. https://doi.org/10.1162/ling_a_00279.

Grebenyova, Lydia. 2009. Sluicing and Multiple Wh-fronting. In Nguyen Chi Duy Khuong & Richa Samar Sinha (eds.), *Asian Generative Linguistics of the Old World (GLOW in Asia)* (5), 219–242. New Dehli: Central Institute of Indian Languages.

Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the Pieces of Inflection. In Kenneth Hale & Jay S. Keyser (eds.), *The view from Building 20*, 111–176. Cambridge, MA: MIT Press.

Hiraiwa, Ken & Shinichiro Ishihara. 2002. Missing Links: Cleft, Sluicing, and "No da" Construction. *MIT Working Papers in Linguistics* 43. 35–54.

Hoyt, Frederick & Alexandra Teodorescu. 2012. How many kinds of sluicing, and why? Single and multiple sluicing in Romanian, English, and Japanese. In Jason Merchant & Andrew Simpson (eds.), *Sluicing: Cross-Linguistic Perspectives*, 83–103. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199645763.003.0005.

Jayaseelan, Karattuparambil. 1990. Incomplete VP deletion and gapping. *Linguistic Analysis* 20(2). 64–81.

Johnson, Kyle. 2009. Gapping is not (VP-) ellipsis. *Linguistic Inquiry* 40(2). 289–328. https://doi.org/10.1162/ling.2009.40.2.289.

Kotek, Hadas & Matthew Barros. 2018. Multiple Sluicing, Scope, and Superiority: Consequences for Ellipsis Identity. *Linguistic Inquiry* 49(4). 781–812. https://doi.org/10.1162/ling_a_00289.

Kroch, Anthony. 1998. Amount quantification, referentiality, and long wh-movement. In Alexis Dimitriadis, Hikyoung Lee, Christine Moisset & Alexander Williams (eds.), *Penn Working Papers in Linguistics*, vol.

5.2. Philadelphia: University of Pennsylvania.
Kuznetsova, Alexandra, Per B. Brockhoff & Rune Haubo B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13). 1–26. https://doi.org/10.18637/jss.v082.i13.
Lasnik, Howard. 1999. On Feature Strength: Three Minimalist Approaches to Overt Movement. *Linguistic Inquiry* 30(2). 197–217. https://doi.org/10.1162/002438999554039.
Lasnik, Howard. 2014. Multiple Sluicing in English? *Syntax* 17(1). 1–20. https://doi.org/10.1111/synt.12009.
Levin, Lori. 1982. Sluicing: a Lexical Interpretation Procedure. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, 590–654. Cambridge, MA: MIT Press.
López, Luis & Susanne Winkler. 2003. Variation at the syntax-semantics interface. In Kerstin Schwabe & Susanne Winkler (eds.), *The interfaces: Deriving and interpreting omitted structures* (Linguistik Aktuelle/Linguistics Today 61), 227–248. John Benjamins Publishing Company. https://doi.org/10.1075/la.61.11lop.
Martin, Andrea E. & Brian McElree. 2008. A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language* 58. 879–906. https://doi.org/10.1016/j.jml.2007.06.010.
Martin, Andrea E. & Brian McElree. 2011. Direct-access retrieval during sentence comprehension: Evidence from Sluicing. *Journal of Memory and Language* 64(4). 327–343. https://doi.org/10.1016/j.jml.2010.12.006.
Marušič, Franc & Rok Žaucer. 2013. A Note on Sluicing and Island Repair. In Steven Franks (ed.), *Annual Workshop on Formal Approaches to Slavic Linguistics: The Third Indiana Meeting 2012*, 176–189. Ann Arbor, MI: Michigan Slavic Publications.
Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford University Press.
Merchant, Jason. 2002. Swiping in Germanic. In Jan-Wouter C. Zwart & Werner Abraham (eds.), *Studies in Comparative Germanic Syntax*, 295–321. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Merchant, Jason. 2006. Sluicing. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, 271–291. Blackwell Publishing Ltd. https://doi.org/10.1002/9780470996591.ch60.
Negele, Michaela. 2012. *Varianten der Pronominaladverbien im Neuhochdeutschen: Grammatische und soziolinguistische Untersuchungen* (Studia Linguistica Germanica 108). Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110273281.
Nishigauchi, Taisuke. 1998. "Multiple sluicing" in Japanese and the functional nature of the wh-phrase. *Journal of East Asian Linguistics* 7(2). 121–152.
Nykiel, Joanna. 2013. Wh-phrases in sluicing : an interaction of the remnant and the antecedent. In Philip Hofmeister & Elisabeth Norcliffe (eds.), *The core and the periphery. Data-driven perspectives on syntax inspired by I. A. Sag*, 253–274. Stanford, CA: CSLI Publications.
NYKIEL, JOANNA & JOHN A. HAWKINS. 2020. English fragments, Minimize Domains, and Minimize Forms. *Language and Cognition* 12(3). 411–443. https://doi.org/10.1017/langcog.2020.6.
Nykiel, Joanna, Jong-Bok Kim & Rok Sim. 2023. Case-matching effects under clausal ellipsis and the cue-based theory of sentence processing. *Journal of Linguistics* 59. 327–360. https://doi.org/10.1017/S0022226722000068.
Ortega-Santos, Iván. 2016. *Focus-related Operations at the Right Edge in Spanish* (Issues in Hispanic and Lusophone Linguistics 7). Vol. 7. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/ihll.7.
Park, Jong Un. 2015. Multiple Sluicing in English and Its Implication for the Unconstrained Pseudosluicing Hypothesis. *Studies in Modern Grammar* (86). 1–30. https://doi.org/10.14342/smog.2015.86.1.
Park, Myung-Kwan & Jung-Min Kang. 2007. Multiple Sluicing in English. In *Proceedings of the 21st Pacific Asia Conference on Language, Information and Computation*, 394–404.
Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman & Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51(1). 195–203. https://doi.org/10.3758/s13428-018-01193-y.
Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In Eric Reuland & Alice G.B. ter Meulen (eds.), *The representation of (in)definiteness*, 98–129. Cambridge, MA: MIT Press.
Pesetsky, David. 2000. *Phrasal Movement and its Kin*. Cambridge, MA: MIT Press.
Potter, David, Michael Frazier & Masaya Yoshida. 2017. A two-source hypothesis for Gapping. *Natural Language and Linguistic Theory* 35. 1123–1160. http://dx.doi.org/10.1007/s11049-017-9359-y.
R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.
Richards, Norvin. 2001. *Movement in language: Interactions and Architectures*. Oxford: Oxford University Press.
Richards, Norvin. 2010. *Uttering Trees*. Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/9780262013765.001.0001.
Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.
Rodrigues, Cilene, Andrew Ira Nevins & Luis Vicente. 2009. Cleaving the interactions between sluicing and P-

stranding. In Danièle Torck & W. Leo Wetzels (eds.), *Romance Languages and Linguistic Theory 2006. Selected papers from "Going Romance", Amsterdam, 7-9 December 2006*, 175–198. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.303.11rod.

Ross, John Robert. 1967. *Constraints on variables in syntax*. MIT.

Ross, John Robert. 1969. Guess who? In Robert I. Blinnick, Alice Davison, Georgia M. Green & Jerry L. Morgan (eds.), *Papers from the fifth regional meeting of the Chicago Linguistic Society*, 252–286. Chicago: University of Chicago: Chicago Linguistic Society.

Sag, Ivan A. & Joanna Nykiel. 2011. Remarks on sluicing. In Stefan Müller (ed.), *18th International Conference on Head-Driven Phrase Structure Grammar*, 188–208. Standford, CA: CLSI Publications.

Schwarzschild, Roger. 1999. Givenness, avoidF and other constraints on the placement of accent. *Natural Language Semantics* 7. 141–177. https://doi.org/10.1023/A:1008370902407.

Stjepanović, Sandra. 1999. Multiple sluicing and superiority in Serbo-Croatian. In Pius Tamanji, Masako Hirotani & Nancy Hall (eds.), *North East Linguistics Society (NELS)* (29), 145–159.

Stjepanović, Sandra. 2003. Multiple wh-fronting in Serbo-Croatian matrix questions and the matrix sluicing construction. In Cedric Boeckx & Kleanthes K. Grohmann (eds.), *Multiple Wh-Fronting* (Linguistik Aktuell/Linguistics Today 64), 255–284. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/la.64.14stj.

Stolcke, Andreas. 1990. Gapping and Frame Semantics : A fresh look from a cognitive perspective. In *13th International Conference on Computational Linguistics, volume 2*, 341–346.

Takahashi, Daiko. 1994. Sluicing in Japanese. *Journal of East Asian Linguistics* 3(3). 265–300.

Takahashi, Daiko & Sichao Lin. 2012. Two notes on multiple sluicing in Chinese and Japanese. *Nazan Linguistics* 8. 129–145.

Thoms, Gary. 2016. Pseudogapping, Parallelism, and the Scope of Focus. *Syntax* 19(3). 286–307. https://doi.org/10.1111/synt.12122.

Vries, Mark de. 2020. An exploration of multiple sluicing in Dutch. Paper presented at the Workshop in honor of the defense of Anastasiia Ionova, Leiden University.

Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change* 9(1). 81–105. https://doi.org/10.1017/S0954394500001800.

Williams, Edwin. 1997. Blocking and Anaphora. *Linguistic Inquiry* 28(4). 577–628.

Winkler, Susanne. 2013. Syntactic Diagnostics for Extraction of Focus from Ellipsis Site. In Lisa Lai-Shen Cheng & Norbert Corver (eds.), *Diagnosing Syntax* (Oxford Studies in Theoretical Linguistics), 463–484. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199602490.003.0023.

Zwarts, Joost. 2011. Lexical and Functional Properties of Prepositions. In Dagmar Haumann & Stefan J. Schierholz (eds.), *Lexikalische und grammatische Eigenschaften präpositionaler Elemente*, 1–18. Berlin/New York: Max Niemeyer Verlag. https://doi.org/10.1515/9783110932768.1.