2021. *Proc Ling Soc Amer* 6(1). 448–462. https://doi.org/10.3765/plsa.v6i1.4981.

D-linking and the semantics of *wh*-in-situ

Joshua Martin*

Abstract. Theories of pair-list readings of multiple *wh*-questions commonly posit an interpretive asymmetry between the fronted and in-situ *wh*-phrases, where the fronted *wh*-phrase is argued to function as the sortal key, have a requirement to be interpreted exhaustively, or be obligatorily D-linked. To clarify the empirical landscape of such debate, I present three experiments which tease apart the effects of these often-confounded discourse factors on the order and interpretation of multiple *wh*-questions. They are found to have either inconsistent or insignificant effects, arguing against a unique discourse-sensitivity of the fronted *wh*-phrase. Theories of questions which encode such an asymmetry should accordingly be revised.

Keywords. question semantics; multiple wh-questions; d-linking; discourse structure; contrastive topic; experimental semantics

1. Introduction. Multiple *wh*-questions like (1) are canonically argued (Hagstrom 1998; Krifka 2001) to be composed of a series of sub-questions, generated by quantifying over the domain of sub-answers to one or the other *wh*-phrase.

(1) Who bought what?
   a. {What did **Alex** buy?, What did **Blake** buy?, What did **Casey** buy?, ...}
   b. {Who bought **arugula**?, Who bought **broccoli**?, Who bought **carrots**?, ...}
   c. Alex bought arugula, Blake bought broccoli, Casey bought carrots, ...

The pair-list interpretation of (1), then, can be rendered as either (1-a), where the domain of the fronted *wh*-word *who* provides the answers to generate the sub-questions, or as (1-b), where the in-situ *wh*-word *what* does. Note that the complete answer to (1) is, in either case, (1-c). The answer alone is therefore ambiguous between either choice of sub-questions, and symmetrical with respect to which *wh*-phrase is interpreted as generating them.

A number of approaches to deriving pair-list readings, however, have posited asymmetries in the interpretation of multiple *wh*-questions, such that one of the *wh*-phrases obligatorily or preferentially determines the sub-questions. In particular, the *wh*-phrase which undergoes movement to [Spec,CP] is often argued to have a kind of interpretive primacy. Exactly what this asymmetry consists in is much less consistently articulated. The earliest expression of this intuition appears to be in Kuno (1982), who introduces the notion of the 'sortal key' of a multiple *wh*-question to mark the *wh*-phrase which determines the sub-questions, identifying the moved *wh*-phrase as the sortal key, at least in English. But what semantic content does sortality encode? Is it its own primitive, or should we understand it as a reflex of another property?

É Kiss (1993) recasts the intuition in terms of specificity, taken from Enç (1991) who in turn develops it from Heim (1982), defining a specific *wh*-phrase as one which 'quantifies over a set which is familiar to the participants of the discourse' (É Kiss 1993; 87). She imposes a Specificity Filter on the interpretation of sentences containing multiple operators in general, of which multiple *wh*-questions are a special case: any operator which both takes scope over and

* Thanks to Kathryn Davidson, Gunnar Lund, and the rest of the Meaning & Modality Lab at Harvard, and to the reviewers and audience of the 95th Annual Meeting of the Linguistic Society of America, for helpful advice and feedback. Author: Joshua Martin, Harvard University (joshuamartin@g.harvard.edu)

binds a variable in the scope of another operator must be specific. The wider scope operator, for É Kiss, also behaves as a universal quantifier, rather than a true interrogative operator like the lower *wh*-phrase. This idea that the wide scope *wh*-phrase carries universal quantificational force is itself taken from Comorovski (1989).

'Specificity', as É Kiss defines it, is difficult to distinguish informally from the related concept of D-linking, which refers to *wh*-phrases for which 'the range of felicitous answers is limited by a set both speaker and hearer have in mind' (Pesetsky 1987; 108). Comorovski's view on the issue appears to evolve along these lines, and by Comorovski (1996) is phrasing the condition in terms of D-linking. Wide scope *wh*-phrases, she says, are always D-linked:

(2) a. It's nice to have all those times scheduled, but **when** are you doing **what**? (#But **what** are you doing **when**?)
b. It's nice to have all those activities ahead of you, but **what** are you doing **when**? (#But **when** are you doing **what**?)

The examples and their judgments in (2), taken from Bolinger (1978), motivate Comorovski's conclusion that the wider-scope *wh*-phrase is obligatorily the D-linked one.[1] Comorovski also reports that the set quantified over by the higher *wh*-phrase is required to be exhausted, while the lower *wh*-phrase has no such requirement. Whether this requirement is meant to follow from D-linking, D-linking is meant to follow from it, both follow from some third property, or they are enforced independently, however, is up in the air.

D-linking has well-known consequences for *wh*-movement in terms of ameliorating Superiority violations, though there is equally little agreement on how exactly this occurs: Pesetsky (1987) takes it to be the D-linking of the in-situ *wh*-phrase that pacifies Superiority, while Bolinger (1978) argues it is the D-linking of the moved *wh*-phrase that is relevant. It may be that neither answer is clearly correct (Grohmann 1998). The possibility that the moved *wh*-phrase bears the D-linking requirement, however, may be considered surprising in light of work on single *wh*-in-situ questions (AKA declarative syntax questions), which have been argued to subject the in-situ *wh*-phrase to some similar kind of requirement of answers being, roughly, in the common ground (Pires and Taylor 2007; Bobaljik and Wurmbrand 2015).

Separating the notion of exhausted domains from D-linking, though, Dayal (1996) also argues for a wide-scope exhausitivity requirement, based on the following data:

(3) *Speaker A:* We're organizing singles tennis games between men and women. There are three men interested in playing against the women, but there are five women interested in playing against the men.
a. *Speaker B:* #So, **which woman** is playing against **which man**?
b. *Speaker B:* So, **which man** is playing against **which woman**?

Dayal describes (3-a) as infelicitous and argues that this infelicity is due to the higher *wh*-phrase having a higher cardinality than the lower *wh*-phrase: since there are only three men, there can only be three games, and so two women will necessarily be left out of the answer. Thus, it is impossible to exhaust the set denoted by *which woman*, and so it may not occur as the higher *wh*-phrase. The inverse order in (3-b), with the exhaustible set in the higher posi-

[1] These judgments are, however, quite subtle and tenuous, and my own intuition is exactly the reverse. Native speakers I consulted informally shared my judgments, but introspection will be too precarious to construct a theory here.

tion, is described as acceptable. Dayal's restriction is refined by Willis (2008), who argues that it is only applicable to questions with two *which*-phrases:

(4) *Speaker B:* So, **who** is playing against **which man**?

Per Willis, when the higher *wh*-phrase is not a *which*-phrase, the set it denotes can outnumber the set denoted by the lower *which*-phrase. Willis casts this revised requirement in terms of Topichood: a *wh*-Topic population cannot outnumber a *wh*-non-Topic population. Taking principles of Topichood from Krifka (2001), then, '(i) when two phrases are equally specific, a subject makes a better Topic than a direct object does; and (ii) lexically specific phrases make better Topics than non-specific phrases do' (Willis 2008; 5). This information-structural approach is echoed by Šimík (2010), who explicitly adopts this asymmetrical subject-object distinction, and Constant (2014), who calculates sub-questions via contrastive topic marking but only acknowledges that such an asymmetry could be encoded that way, if one wanted.

All of these accounts take significant inspiration from Hagstrom (1998), whose account of pair-list readings involves the explicit merger of a Q morpheme at the lower *wh*-phrase. On this analysis, an asymmetry between the *wh*-phrases is crucial for the derivation of pair-list readings, which are argued to be a consequence of one *wh*-phrase being in the scope of Q and the other outside. Questions without Q, as is possible in Japanese (which instantiates Q overtly), are thus restricted to single-pair readings. The intuition that the higher *wh*-phrase carries universal force - which arguably results in Dayal's restriction on exhaustible domains - is thus a consequence of its place outside the scope of Q. Any account which follows in Hagstrom's footsteps by having a Q-like morpheme compose with one *wh*-phrase, then, can be taken as implicitly making some commitment about the presence or absence of interpretive differences between the *wh*-phrases. On the other hand, Grebenyova (2006) argues that Hagstrom's choice of target for Q ends up being inconsequential for the final denotation.

Given all of this, we can perhaps pose one question in two ways. First, for a particular multiple *wh*-question, what can the choice of word order tell us about the intended sub-questions? Second, given a particular discourse context, what factors influence the choice of word order for a multiple *wh*-question? It should be clear, at least, that these are the same question, asked from the perspective of the addressee and the producer, respectively.

In light of the significant theoretical morass surrounding these questions, the goal of the present paper is to clarify at least the empirical foundation for such debate. Here, I will restrain from doing much of any theorizing, and rather focus on attempting to pull apart these various strands of interrelated discourse factors. In doing so, we will see that many of these commonly deployed introspective judgments are more complex than previously assumed. As a result, I argue, it should not be a desideratum of a semantics of pair-list readings to encode a unified restriction on the interpretation of either the higher or lower *wh*-phrase. Sensitivity to discourse factors should be allowed, but not hard-coded in any asymmetric manner.

2. Experiments. In order to clarify the role that these overlapping discourse factors play in interpreting pair-list readings, this section presents three experiments aimed at evaluating the effect of sortality, exhaustivity, and D-linking. All three experiments follow the same general structure. First, participants read a context which established that one domain of answers bore the relevant discourse feature (was sortal, could be exhausted, was D-linked). Then, they evaluated two possible multiple *wh*-questions for their appropriateness in the context, where the

questions differed only in the choice of *wh*-phrase fronted. All questions in the experiments were designed to avoid Superiority violations; while factors like D-linking interact in interesting ways with movement conditions, I set that aside here to simplify the picture.

2.1. EXPERIMENT 1: SORTAL KEYS. The first experiment evaluates the role that discourse context plays in determining which *wh*-phrase is used as the 'sortal key', in the sense of Kuno (1982), and whether there is a preference for using the higher or lower *wh*-phrase as said key. It is a further development of an experiment piloted by Constant (2014) to test these intuitions:

> 'subjects... were asked to pick the best title for a list of frequently asked questions on a college admissions website. In condition A, the FAQ consisted of questions addressing where to send the green form, the blue form and the pink form. In condition B, the questions addressed what to send to the admissions office, to financial aid, and to the department of proposed major. Presented with a forced binary choice, the majority (69%) of condition A subjects (n=108) chose the title "Where to Send What", whereas most (71%) condition B subjects (n=111) chose the title "What to Send Where".' (Constant 2014; footnote 49)

Constant's preliminary results suggest the opposite pattern from the intuition of Kuno and others: participants preferred the titles where the in-situ *wh*-phrase functioned as the sortal key. Here, I expand on Constant's design to include a wider variety of contexts and *wh*-words.

2.1.1. METHODS. *Participants:* Sixty participants, self-identified native speakers of English and located in the United States, were recruited through Amazon's Mechanical Turk system.

Stimuli: For each question, participants would read a context in the form of text for a hypothetical webpage, and be asked to choose the more appropriate title for the page. Each context consisted of a short preamble describing the purpose of the page, and then a series of bullet-pointed information sorted in one of two ways, each corresponding to one of the *wh*-words which would appear in the title choices.

(5) *Example stimulus context for Experiment 1:*

'***A university wants to put this page on their website:***

It can be easy to lose track of all of the components of your application, which may seem complicated because different offices in our department need access to different pieces of information. To help prevent this confusion, this page details exactly what you need to do to make sure your application is complete and can be processed correctly. Luckily, the pages of the application packet are color-coded to make sorting easier.

- The blue form: These pages contain your demographic information. You need to send a copy of the blue form to the admissions office and to the financial aid office.
- The green form: These pages contain all of the financial information. You need to send a copy of the green form to the financial aid office.
- The pink form: These pages contain your academic history and your personal statement. You need to send a copy of the pink form to the admissions office and to the department of your intended major.

What should this page be titled?'

In the other condition, the preamble to the webpage would be identical, but the bulleted information would be organized differently, to sort by the other list of answers. For example, for the admissions question, participants in the other condition would instead see:

(6) *Example stimulus context for Experiment 1, other sorting option:*

> •Admissions: The admissions office uses your demographic information, your academic history, and your personal statement to determine whether you will be admitted. Therefore, you need to send this office a copy of the blue form and of the pink form.
>
> •Financial aid: The financial aid office considers your financial information and demographic information, to determine if you will be offered any scholarships. Therefore, you need to send this office a copy of the green form and of the blue form.
>
> •Major department: The department of your intended major evaluates your academic history and personal statement in order to gather information about which kinds of topics students are most interested in. Therefore, you need to send this office a copy of the pink form.

In both conditions, participants would be given two options for a title, both of which were multiple *wh*-question fragments, with the only difference being the swapped order of the *wh*-words. So, for this example, participants would select between *What to Send Where* and *Where to Send What*, as in Constant's pilot. Three other target contexts were constructed, with the *wh*-word pairs *what/who*, *when/where*, and *who/when*, and one filler context to act as a catch question, where only one option could be used to correctly describe the context. Crucially, these descriptions only established one *wh*-word as the preferential sortal key, but did not asymmetrically D-link them: for both *wh*-words, the full membership of the answer domain was explicitly mentioned.

Procedure: Participants were given these instructions at the beginning of the survey:

> "In this survey, we are interested in the appropriate way to title webpages with different information. In each question, you will be given an example of some content that would appear on a webpage to read, and then asked to choose between two options for the appropriate title of that webpage. For example, a college's website may have a Frequently Asked Questions section about their application forms, and we would like to know what the right label for that page should be."

Participants were randomly assigned to either *wh*-sorting condition for each question independently. With five *wh*-pairs and two contexts for each, sixty participants answered five of the ten possible questions each, for a total of thirty responses per question.

2.1.2. RESULTS. The key prediction was whether participants would show a significant trend towards selecting answers with the sortal *wh*-word in the fronted position or in-situ. Participants selected the answer where the sortal *wh*-word was fronted in 59.0% of questions. A chi-squared goodness of fit test was performed to determine whether this was significant, and found that it was, $X^2(1, N = 236) = 7.4746,\ p < 0.01$.

The forced-choice design for a word-order effect raises a difficulty for statistical modeling: binomial logistic regression assumes a binary output variable which can be coded as a

'success' or 'failure'. Here, the dependent variable is the choice of *wh*-order, for which there is no clear way to code one of the choices as a 'success'. A logistic model which tries to predict whether the fronted-sortal option was chosen, from whether the fronted option was sortal, is not well-formed. So, following Lohmann and Takada (2014), a logistic regression was performed where the response variable was treated as a 'dummy' variable always set to 1, and the independent variable (whether the sortal *wh*-word was fronted) was coded as either 1 or -1. This regression also found a significant main effect of the position of the sortal key ($\beta = 0.36, p < 0.01$), with participants more likely to select choices where the sortal key was fronted. This would confirm Kuno's hypothesis, contrasting the results from Constant's pilot.

However, as Figure 1 illustrates, this tendency did not hold across all questions. Certain questions showed significant effects in the opposite direction.

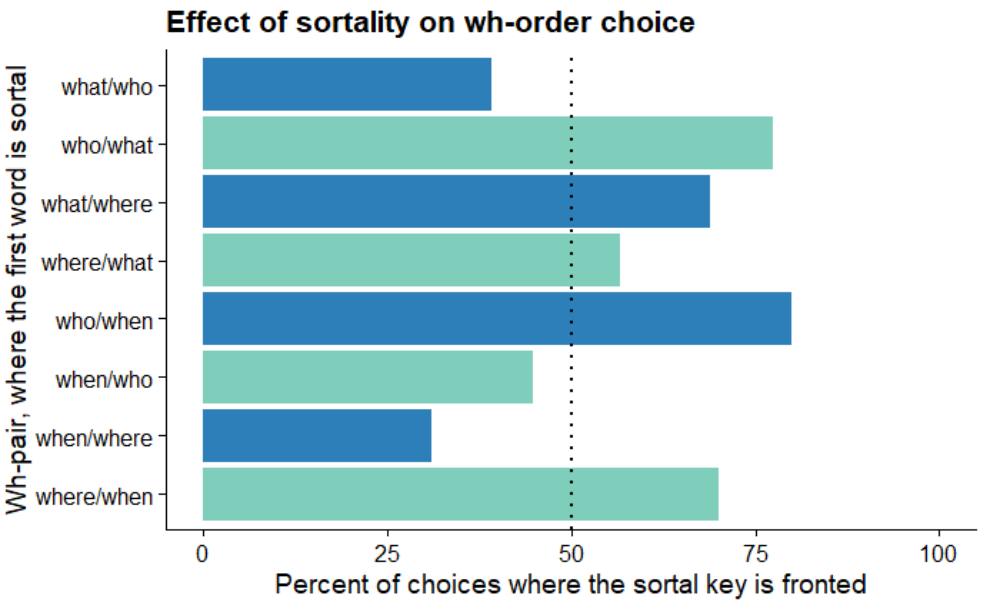


Figure 1. Experiment 1 results, by *wh*-word pair. Each pair appears twice, reflecting the two contexts which marked either word as the sortal key.

Thus, it is difficult to conclude from these results that the fronted *wh*-phrase uniformly functions as the sortal key. There does appear to be pressure in that direction due to some, potentially noisy, discourse features, but this interacts with the relative positional preferences of particular *wh*-words.

2.2. EXPERIMENT 2: EXHAUSTIVITY. The second experiment probes the role of exhaustivity and universal quantification in pair-list readings. In particular, it tests Dayal's hypothesis that the fronted *wh*-phrase must have its domain exhausted over, such that a multiple *wh*-question is infelicitous if the domain of the fronted *wh*-phrase has fewer members than the domain of the in-situ *wh*-phrase. In this (and the following) experiment, I also modify the design of Constant (2014) by adopting the paired continuous-scale rating task from Marty et al. (2020), who argue that this particular task is well-suited to eliciting contrasting judgments from participants. They find that, on top of pairing theoretically related judgments as Experiment 1 already employed, the use of continuous sliders for responses increases effect sensitivity and removes unneces-

sary reasoning interference on the part of the participant. Thus, if an effect of exhaustivity is present, this task is more likely to detect it.

2.2.1. METHODS. *Participants:* One hundred and twenty participants, self-identified native speakers of English and located in the United States, were recruited through Amazon's Mechanical Turk system.

Stimuli: As in Experiment 1, for each question, participants would first read a short context which established two groups that differed in cardinality. Then, they would rate two multiple *wh*-questions, which again differed only in swapping the position of the *wh*-phrases. Here, full questions were used rather than the fragment titles from Experiment 1. To avoid Superiority effects, the *wh*-phrases always realized the direct and indirect object, or direct object and an adjunct, rather than the subject, which was always the second-person pronoun *you*. Thus, no moved *wh*-phrase crossed over a *wh*-subject. Figure 2 shows an example of a question, as a participant would see.

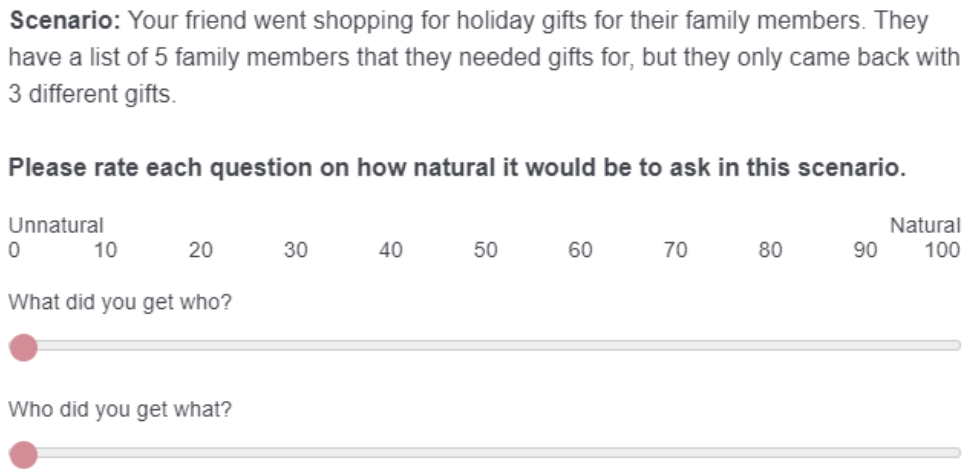


Figure 2. Example question from Experiment 2; in this context *what* has the smaller domain.

The *wh*-phrases used in this experiment were only *what*, *who*, and *which N*, the domains of *wh*-adjuncts being inherently more difficult to quantify. Two contexts were generated for each pair: *what/who*, *what/which*, *who/which*, *which/which*, and a filler catch question.

Procedure: Participants were given these instructions at the beginning of the survey:

> "In this survey, you will be given short scenarios to read, and then asked to rate pairs of questions that you might ask in that scenario. Please read the scenarios carefully. For each question, please use the sliding scale to indicate how natural that question would be to ask in the scenario, from 0 (completely unnatural) to 100 (completely natural)."

Participants then completed three training questions familiarizing them with the scale. They then saw five questions in random order, and randomly assigned to one of two scenarios for each question. Within each scenario, they were randomly assigned a version of the context

that established one or the other *wh*-word as the smaller domain. With five *wh*-pairs (four target, one filler) and four contexts for each, one hundred and twenty participants answered five of twenty possible questions each, for a total of thirty responses per question.

2.2.2. RESULTS. The key prediction is whether participants assign higher ratings of naturalness to multiple *wh*-questions where the fronted *wh*-phrase has a smaller domain (and is thus exhaustively interpreted) than its in-situ counterpart. This prediction is not borne out: there is no significant difference in ratings for questions where the exhaustive *wh*-phrase is raised ($M = 59, SD = 35$) than questions where the exhaustive *wh*-phrase is in-situ ($M = 61, SD = 35$), $t(942) = 1.0, p = 0.31$. A linear mixed model with question rating as the dependent variable, and fronted vs. in-situ exhaustivity and the particular *wh*-pair as independent variables, finds no effect of exhaustivity ($\beta = 2.3, p = 0.3$). Figure 3 shows the distribution of ratings.

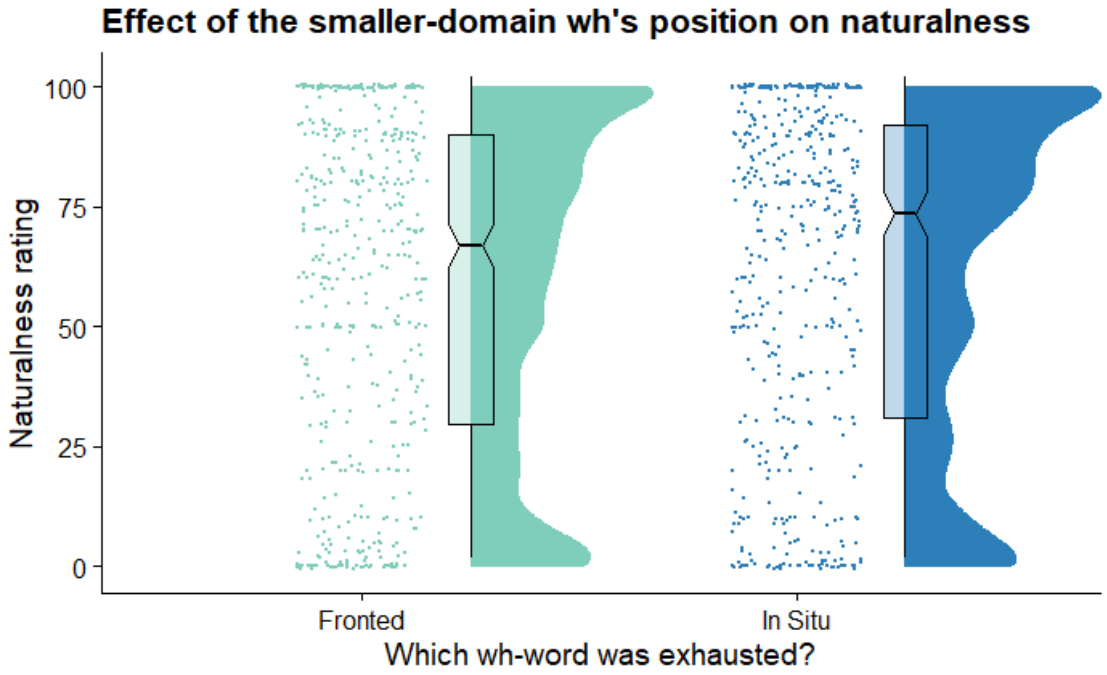


Figure 3. Experiment 2 results, across *wh*-words. Constructed using the Raincloud Plots R toolkit (Allen et al. 2021).

As Figure 3 shows, the distributions when all *wh*-words are grouped together are largely indistinguishable, reflecting the statistical insignificance of the exhaustivity factor. What contributes, then, to the variance observed? Looking at the data broken down by *wh*-word pairs suggests that potentially inherent preferences for fronting some *wh*-words over others may be the source of variance in ratings, and that these preferences are largely insensitive to the exhaustivity manipulation. Figures 4, 5, and 6 show the ratings broken down by each pair of *wh*-words. In Figure 4, there is a clear preference for fronting *who* over *what*, irrespective of the cardinalities of their domains. In Figure 5, the same effect holds for fronting *which N* over *what*. Figure 6, with *who* and *which N*, does appear to show an effect, but with a preference for fronting *who* only when it has the larger domain, contrary to Dayal's prediction. I will return to a potential explanation for this in the discussion.

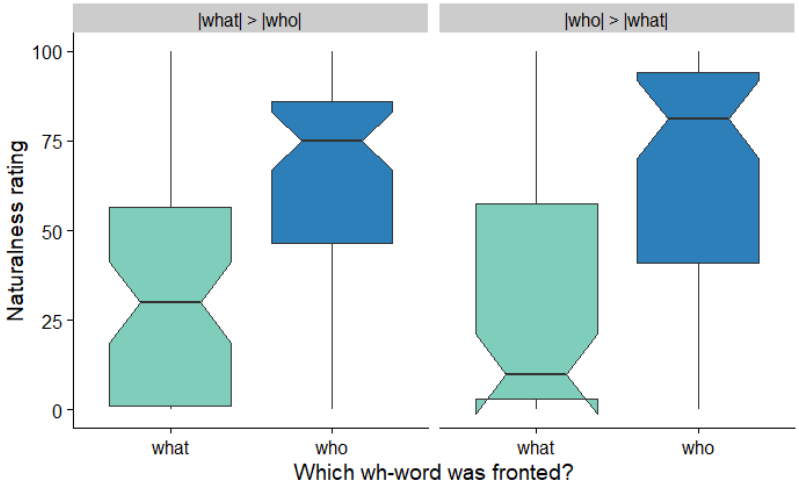


Figure 4. Ratings of sentences with *what* and *who*

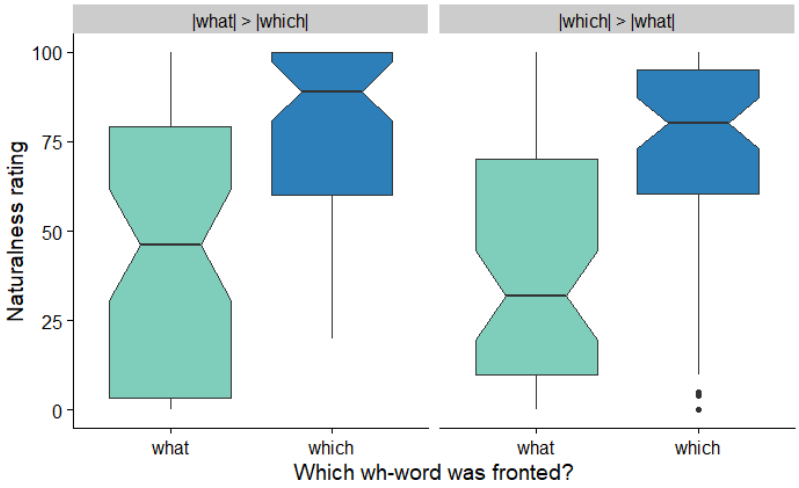


Figure 5. Ratings of sentences with *what* and *which N*

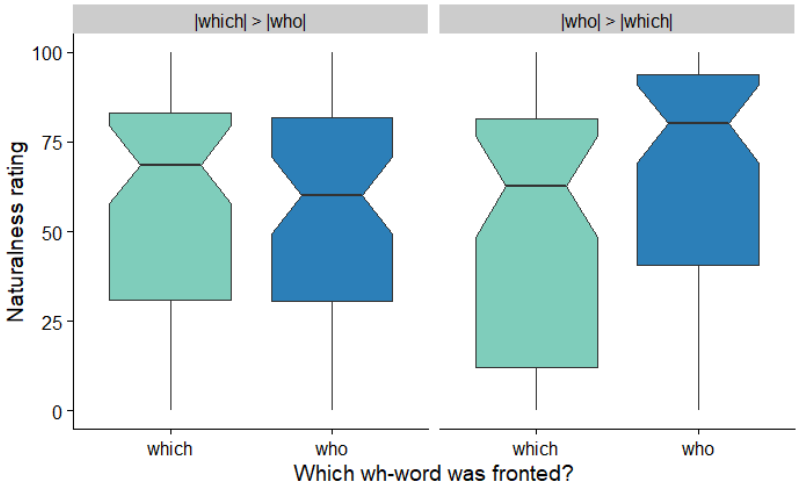


Figure 6. Ratings of sentences with *who* and *which N*

Recall also that Willis (2008) argues for a more restricted version of Dayal's hypothesis, on which the domain-size requirement only applies to questions with *which*-subjects. Restricting the analysis to that subset of the data, however, similarly produces no effect of domain size, $t(470) = 0.5, p = 0.6$. Figure 7 plots the ratings of questions with two *which N*-phrases.

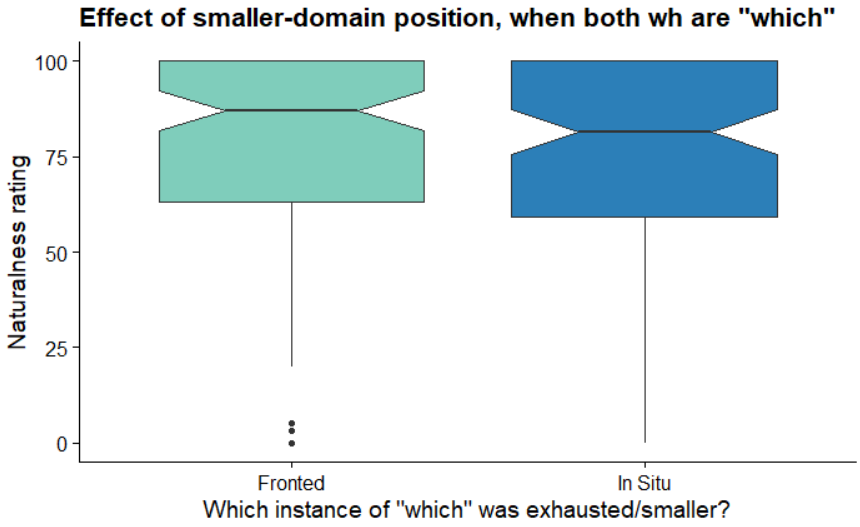


Figure 7. Ratings of sentences with two *which N*-phrases

2.3. EXPERIMENT 3: D-LINKING. The third experiment investigates the role of D-linking in pair-list readings. In particular, it tests the hypothesis from Comorovski (1996) - and, to the extent that Enç's notion of 'specificity' can be identified with D-linking, also from É Kiss (1993) - that the fronted *wh*-phrase is obligatorily D-linked. Here, D-linking is represented by providing contexts in which a participant is informed of their knowledge of the relevant sets, with some illustrative examples, rather than explicitly providing the full membership of the sets. This is done for two reasons. First is simply streamlining of presentation and reduced reading load for participants. Second is the difficulty of listing out individuated set members for the domains of some *wh*-words, in particular *wh*-adjuncts.

2.3.1. METHODS. *Participants:* Sixty participants, self-identified native speakers of English and located in the United States, were recruited through Amazon's Mechanical Turk system.

Stimuli: Stimuli were nearly identical to those of Experiment 2, with two differences. First, rather than the scenario establishing the cardinality of the sets, it establishes familiarity of one of the sets in the discourse. Second, there is a wider variety of *wh*-words used, all except *which*: being inherently D-linked, it was not appropriate for this contextual manipulation. Figure 8 shows an example of a question.

Procedure: Participants saw the same instructions and completed the same three training questions to familiarize them with the scale as in Experiment 2. They then answered eighteen questions, fifteen target and three filler catch questions (where one option was unambiguously ungrammatical), in random order, and for each question were randomly assigned to a context which D-linked one or the other *wh*-word. With thirty-six possible questions and sixty participants answering eighteen questions each, each question had a total of thirty responses.

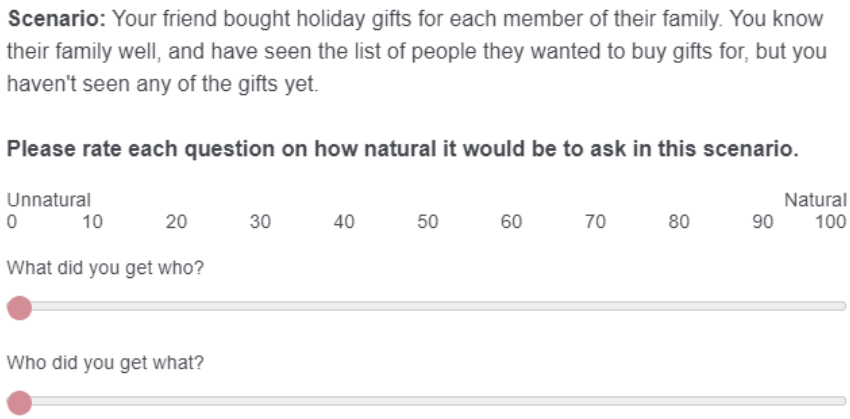


Figure 8. Example question from Experiment 3; in this context *who* is D-linked.

2.3.2. RESULTS. The key prediction is whether participants assign higher ratings of naturalness to multiple *wh*-questions where the fronted *wh*-phrase is D-linked. This prediction is not borne out: there is a significant difference in ratings, but in the opposite direction. Questions where the fronted *wh*-word is D-linked ($M = 41, SD = 36$) are rated significantly lower than questions where the in-situ *wh*-word is D-linked ($M = 51, SD = 37$), $t(1764) = 6.0, p < 0.0001$. A linear mixed model with question rating as the dependent variable and independent variables for the position of the D-linked *wh*-word and the identity of both *wh*-words found a significant main effect of D-linking ($\beta = 19.8, p < 0.0001$), along with significant main effects for which *wh*-word was in-situ for all *wh*-words. Figure 9 shows the distribution of ratings.

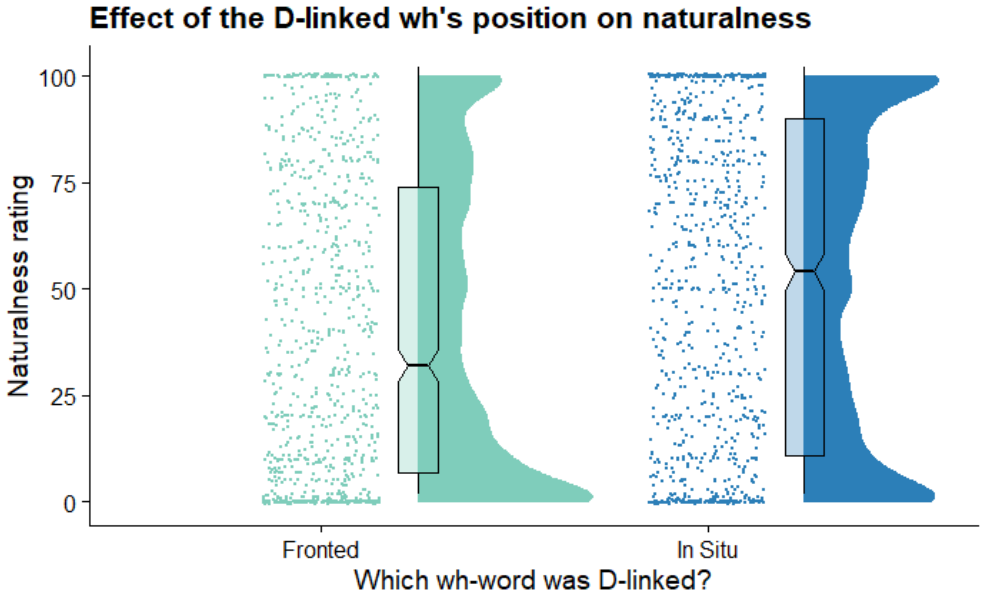
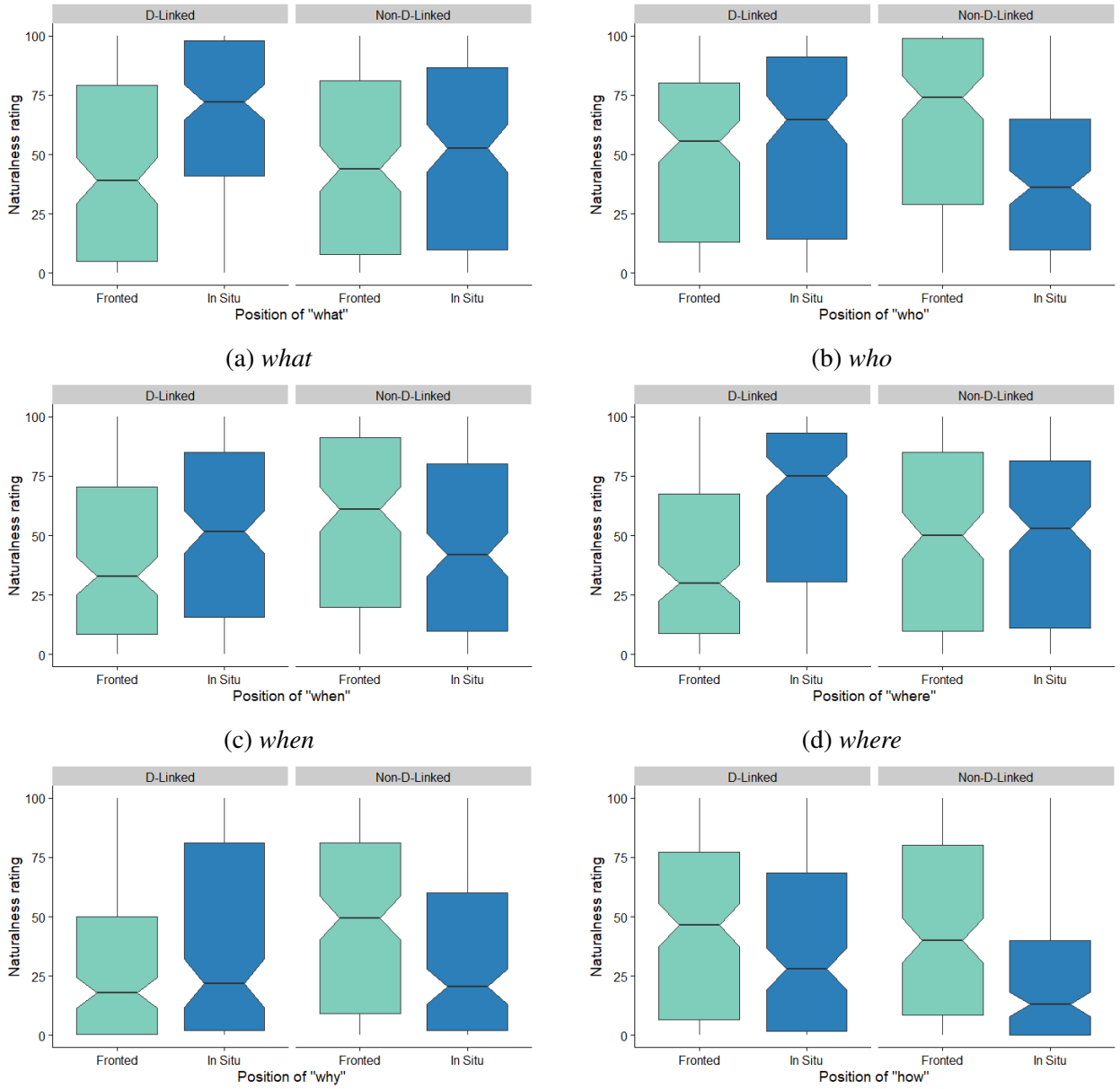


Figure 9. Experiment 3 results, across all *wh*-words. Constructed using the Raincloud Plots R toolkit (Allen et al. 2021).

Here, the effect of D-linking was more consistent across *wh*-words than in Experiment 2.

Figures 10a-10f break the data down by *wh*-word. We can see that, in general, each word is rated more highly when D-linked if it is in situ rather than fronted. The only obvious exception is *how*, which is rated significantly higher when fronted than in situ, in both the D-linked ($t(290) = -2.5, p < 0.05$) and the non-D-linked ($t(283) = -4.6, p < 0.001$) conditions. This is possibly because it is so degraded in situ regardless of the attempts to D-link it; we can see that its adjunct counterpart *why* is similarly degraded in its in situ uses relative to the other *wh*-words. However, despite their general unnaturalness, there is still a significant effect of D-linking improving the ratings for in situ *how* ($t(288) = -2.7, p < 0.01$), though it does not reach significance for *why* ($t(287) = 1.2, p = 0.2$).



(a) *what* (b) *who*

(c) *when* (d) *where*

(e) *why* (f) *how*

Figure 10. Effect of D-linking on ratings of each *wh*-word fronted and in situ.

3. Discussion. Let me first give a brief summary of the results. Experiment 1 tested the role of sortal information structure, and found that which *wh*-word functioned as the sortal key had a significant effect on participants' choice of position for the *wh*-words: sortal keys were preferred fronted. This was in the direction of Kuno's (1982) hypothesis, but the effect did not show up uniformly across *wh*-words, and in fact appeared as the inverse in some cases.

Experiment 2 tested the hypotheses of Dayal (1996) and Willis (2008) that fronted *wh*-phrases need to have a smaller domain than their in-situ counterparts, because they are required to be exhausted over (perhaps due to being interpreted as universal quantifiers). It found no effect of the position of the smaller-domain *wh*-phrase. Instead, the identity of the specific *wh*-words used in the question appeared to have a larger effect on which one to move.

Experiment 3 tested the role of D-linking, and specifically Comorovski's (1996) hypothesis that the fronted *wh*-phrase is preferentially D-linked over the in-situ *wh*-phrase. It found a significant effect, but in the opposite direction: participants preferentially D-linked the in-situ *wh*-phrase. In contrast to Experiment 1, this effect was largely consistent across *wh*-words, though unsurprisingly different *wh*-words also resisted being left in-situ to different extents.

What should we conclude from this? First of all, the type of conclusion I don't think we should reach is, for example, that sortality is an inherent semantic property of fronted *wh*-questions and that D-linking is an inherent semantic property of in-situ *wh*-questions, and that we should construct theories of the semantics of questions with the goal of deriving these properties. While there are significant effects of these discourse factors (in these particular experimental setups), they are too poorly understood at the moment to commit to that kind of conclusion, and the effects are in no way categorical. The fact thatparticipants, in a forced choice task, selected the option with the sortal key fronted 59% of the time might indicate that there is a kind of noisy information-structural process creating such a bias; it doesn't seem like it indicates an effect of the grammatical process of interpreting multiple *wh*-questions.

There are, however, some takeaways. I would argue, from this data, that these three oft-cited discourse factors are not all reflexes of the same underlying semantic property. Given their similarity and potential overlap, it would be very natural and theoretically tempting to tell a Hagstrom-like story, like this: fronted *wh*-words are interpreted as universal quantifiers, and so the complete list of sub-questions needs to exhaust the *wh*-word's domain, which also requires that the members of that domain are discourse-familiar and results in the sub-questions being sorted in a particular way. What this data suggests is that this is not the case: sortality, cardinality/exhaustivity, and D-linking have distinct and occasionally diverging effects. If they are to be derived from a single underlying logical property, it will need to be through some significant filtering that allows them to be realized inconsistently on the surface.

So, we should not consider it a desideratum for a theory of multiple *wh*-questions that it logically derives these discourse properties in an asymmetric way. The other, positive takeaway from this data is that we should expect a successful theory of multiple *wh*-questions to have an account for the interaction that these discourse factors have with individual *wh*-words. The inconsistent application of these properties across different *wh*-words suggests that individual *wh*-words bring their own prior expectations with respect to things like being a sortal key or being D-linked. This isn't surprising - if it's actually the case that English *wh*-in-situ have some sort of D-linking preference, we should expect that interact with what we might call the inherent 'D-linkability' of *wh*-adjuncts like *why* or *how*. Comorovski (1996) suggests that D-linking 'requires that the participants... be able to partition identically the set that the *wh*-

phrase ranges over'; domains containing manners or reasons seem, inherently, more difficult to partition than domains containing individuals. So, it may still be the case that, for example, a semantic account of why *wh*-adjuncts rarely appear in-situ is desirable, but such an account will need to synthesize the role of adjunct's own domains, the influence of the discourse context, and whatever our eventual semantic derivation of such questions will be.

In that sense, most of what this paper has accomplished has been to muddy the waters. Where we previously assumed there to be some clear associations between syntactic position and semantic interpretation, now again we have a noisier picture. This complication has also highlighted the need for a deeper understanding of what, exactly, we mean when we say things like 'D-linking', in order to distinguish it from other clearly related but empirically distinct ways of describing discourse structure. We have, hopefully, a better sense of what we want our theories to *not* do and what assumptions they should *not* make. The next step for future work in this area is to flesh out the empirical landscape further, until we can start to see more clearly the patterns that our analysis *should* capture.

References

Allen, Micah, Davide Poggiali, Kirstie Whitaker, Tom Rhys Marshall, Jordy van Langen & Rogier A. Kievit. 2021. Raincloud plots: a multi-platform tool for robust data visualization [version 2; peer review: 2 approved]. *Wellcome Open Research* 4(63). https://doi.org/10.12688/wellcomeopenres.15191.2.

Bobaljik, Jonathan & Susi Wurmbrand. 2015. Questions with declarative syntax tell us what about selection? In Ángel J. Gallego & Dennis Ott (eds.), *50 years later: Reflections on Chomsky's Aspects* (MIT Working Papers in Linguistics 77), 13–31. Cambridge, MA: MIT.

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

Comorovski, Ileana. 1989. *Discourse and the syntax of multiple constituent questions*. Ithaca, NY: Cornell dissertation.

Comorovski, Ileana. 1996. *Interrogative phrases and the syntax-semantics interface* (Studies in Language & Philosophy 59). Dordrecht: Springer. https://doi.org/10.1007/978-94-015-8688-7.

Constant, Noah. 2014. *Contrastive topic: Meanings and realizations*. Amherst, MA: University of Massachusetts dissertation. https://doi.org/10.7275/wdwx-7p85.

Dayal, Veneeta. 1996. *Locality in wh quantification*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-4808-5.

É Kiss, Katalin. 1993. Wh-movement and specificity. *Natural Language & Linguistic Theory* 11. 85–120. https://doi.org/10.1007/bf00993022.

Enç, Mürvet. 1991. The semantics of specificity. *Linguistic Inquiry* 22. 1–25.

Grebenyova, Lydia. 2006. *Multiple interrogatives: Syntax, semantics, and learnability*. College Park, MD: University of Maryland dissertation.

Grohmann, Kleanthes K. 1998. Syntactic inquiries into discourse restrictions on multiple interrogatives. *Groninger Arbeiten zur germanistischen Linguistik* 42. 1–60.

Hagstrom, Paul Alan. 1998. *Decomposing questions*. Cambridge, MA: MIT dissertation.

Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. Amherst, MA: University of Massachussetts dissertation.
Krifka, Manfred. 2001. Quantifying into question acts. *Natural Language Semantics* 9. 1–40. https://doi.org/10.1023/A:1017903702063.
Kuno, Susumu. 1982. The focus of the question and the focus of the answer. In Robinson Schneider, Kevin Tuite & Robert Chametzky (eds.), *Papers from the Parasession on Non-Declarative Sentences* (CLS 18), 134–157. Chicago: Chicago Linguistic Society
Lohmann, Arne & Tayo Takada. 2014. Order in NP conjuncts in spoken English and Japanese. *Lingua* 152. 48–64. https://doi.org/10.1016/j.lingua.2014.09.011.
Marty, Paul, Emmanuel Chemla & Jon Sprouse. 2020. The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: A Journal of General Linguistics* 5(1). 1–23. https://doi.org/10.5334/gjgl.980.
Pesetsky, David. 1987. Wh-in-situ: movement and unselective binding. In Eric J. Reuland & Alice G.B. Ter Meulen (eds.), *The representation of (in)definiteness* (Current Studies in Linguistics 14), 98–129. Cambridge, MA: MIT Press.
Pires, Acrisio & Heather Lee Taylor. 2007. The syntax of wh-in-situ and common ground. *Chicago Linguistic Society (CLS)* 43(2). 201–215.
Šimík, Radek. 2010. Interpretation of multiple interrogatives: An information structure sensitive account. In Wayles Browne, Adam Cooper, Alison Fisher, Esra Kesici, Nikola Predolac & Draga Zec (eds.), *Formal approaches to Slavic linguistics 18: The second Cornell meeting*, 486–501. Ann Arbor, MI: Michigan Slavic Publications.
Willis, Paul M. 2008. The role of topic-hood in multiple-wh question semantics. *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL)* 27. 87–95. http://phonetics.linguistics.ucla.edu/wpl/issues/wccfl27/papers/willis.pdf.