



Eszter Ronai* and Laura Stigliano*

Experimentally testing the interpretation of multiple sluicing and multiple questions in Hungarian

<https://doi.org/10.1515/tlr-2022-2092>

Published online July 8, 2022

Abstract: Hungarian multiple sluicing has been claimed to only be allowed in contexts that set up a pair-list, but not a single-pair reading. This has been taken as evidence that multiple sluicing is derived from multiple wh-fronting questions, which only license pair-list, but not single-pair answers. Providing novel experimental evidence, we show that all three relevant constructions in Hungarian – multiple sluicing, single wh-fronting questions, and multiple wh-fronting questions – in fact pattern alike in their answerhood conditions: both pair-list and single-pair readings are allowed, with a modest preference for single-pair readings. Based on our results, we thus argue that answerhood conditions are not sufficient to determine the source of multiple sluicing.

Keywords: ellipsis; experimental syntax; Hungarian; multiple sluicing; multiple wh-questions

1 Introduction

An important theoretical claim in the ellipsis literature is that properties of non-elliptical sentences in a language should predict the properties of elliptical ones (i.a. Tancredi 1992). One domain where elliptical sentences have been claimed to parallel non-elliptical ones is multiple sluicing. Merchant (2001) shows that languages that allow multiple wh-movement also allow multiple sluicing (e.g. Bulgarian, Hungarian, Polish, and Russian) – though while the availability of multiple wh-movement is a sufficient condition for a language to have multiple sluicing, it is not a necessary one. This correlation between elliptical and non-elliptical structures has been claimed to extend beyond the availability of such

*Corresponding authors: **Eszter Ronai and Laura Stigliano**, Department of Linguistics, The University of Chicago, 1115 E. 58th Street, Chicago, IL, 60637, USA, E-mail: ronai@uchicago.edu (E. Ronai), laurastigliano@uchicago.edu (L. Stigliano)

Open Access. © 2022 the author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

configurations, to the possible interpretations that they allow. In turn, correlations in interpretation have been used to diagnose the source of elliptical structures.

For Hungarian in particular, van Craenenbroeck and Lipták (2013) have argued that multiple sluicing is derived from multiple wh-fronting questions, and not from single wh-fronting questions – both of which are available, and are therefore potential sources, in this language. This argument relies on a parallel the authors identify between multiple sluicing and multiple wh-fronting questions (which is allegedly not shared by single wh-fronting questions): namely, that both structures only license pair-list, but not single-pair interpretations.

In this paper, we offer novel experimental evidence testing the relevant claims from the literature. Using two experimental paradigms, we show that all three relevant constructions in Hungarian – multiple sluicing, single wh-fronting questions, and multiple wh-fronting questions – in fact pattern alike in their interpretation, contra prior claims. On the basis of this evidence, we argue that the availability of pair-list versus single-pair interpretations is not sufficient to determine the source of multiple sluicing in Hungarian. More broadly, our study has consequences for syntactic theory, bearing in on what can be the source of ellipsis, and it also provides insight into how to experimentally test interpretations and answerhood conditions.

The structure of this paper is as follows. Section 2 provides a brief background on multiple sluicing (Section 2.1) and Hungarian multiple wh-questions and multiple sluicing (Section 2.2), as well as their possible interpretations (Section 2.3). Section 3 reports on two acceptability judgement experiments: Experiment 1a (Section 3.1) and its replication, Experiment 1b (Section 3.2), which additionally addresses the potential effect of extreme fillers. Section 4 describes a forced choice experiment (Experiment 2). Section 5 discusses the overall findings from the three experiments, and Section 6 concludes.

2 Background

2.1 Brief background on multiple sluicing

‘Sluicing’, first discussed in Ross (1969), is a form of clausal ellipsis in which a wh-question (1-a) undergoes deletion, except for its wh-phrase, as shown in (1-b):

- (1) a. *Mary bought something, but I don't know what she bought.*
- b. *Mary bought something, but I don't know what.*

‘Multiple sluicing’, a term coined by Takahashi (1994), refers to a type of sluicing in which there is more than one wh-phrase that survives ellipsis, as (2-b) shows:

- (2) a. *Someone bought something, but I don't know who bought what.*
 b. *Someone bought something, but I don't know who what.*

Multiple sluicing is regularly treated as arising from multiple wh-questions, and is found in languages with a wide variety of strategies for forming multiple wh-questions. For instance, multiple sluicing is found in languages with single wh-fronting, like German (3);¹ in languages with wh-in-situ questions, like Japanese (4); and in languages with multiple wh-fronting, like Slovenian (5) (Merchant 2001):

- (3) *Jeder Student hat ein Buch gelesen, aber ich weiss nicht mehr*
 every student has a book read, but I know no longer
wer welches.
 who which
 'Every student read a book, but I can't remember which student which book.'
- (4) *John-ga dareka-ga nanika-o katta to it-ta. Mary-wa*
 John-NOM someone-NOM something-ACC bought that said Mary-TOP
dare-ga nani-o ka siri-tagat-te iru.
 who-NOM what-ACC Q know-want is
 'John said someone bought something. Mary wants to know who what.'
- (5) *Vid je rekel, da je Rok predstavil nekomu nekoga, pa ne*
 Vid AUX said that AUX Rok introduce one.DAT one.ACC, but not
vem komu koga.
 know who.DAT who.ACC
 'Vid said that Rok introduced someone to someone, but I don't know who to who.'

Examples adapted from Abels and Dayal 2017, ex. (6)–(7)

The fact that multiple sluicing is found in these different types of languages raises several questions regarding how ellipsis arises and the relationship between elliptical and non-elliptical constructions. Under a move-and-delete approach (see, e.g. Merchant 2001, and subsequent work) all wh-phrases in a multiple sluicing configuration must be fronted and moved out of the ellipsis site. This is the case regardless of the availability of such movement in non-elliptical constructions. In other words, a move-and-delete approach needs to posit an 'exceptional movement'

¹ Examples are glossed in accordance with the Leipzig conventions, and all examples from the literature cited have been adapted accordingly.

to obtain multiple sluicing in languages like German or Japanese, which do not have multiple wh-fronting. On the other hand, there exist alternative approaches that eschew exceptional movement and allow remnants to stay in-situ (see, e.g. Abe 2015; Stigliano 2022), which can account for the existence of multiple sluicing in different languages without the need to propose ad-hoc exceptional operations.

2.2 Multiple wh-questions and multiple sluicing in Hungarian

Hungarian grammar allows both single and multiple wh-fronting questions, as well as multiple sluicing, making it particularly suitable to evaluate different accounts of multiple sluicing. Multiple sluicing is exemplified in (6) below:

- (6) *Valaki meghívott valakit, de nem tudom ki kit.*
 someone invited someone.ACC but not know.I who.NOM who.ACC
 ‘Someone invited someone, but I don’t know who whom.’

As mentioned, multiple wh-questions either display single wh-fronting, as in (7-a), or multiple wh-fronting, as in (7-b):²

- (7) a. *Ki hívott meg kit?* SINGLE WH-FRONTING
 who.NOM invited PRT who.ACC
 Literal: ‘Who invited whom?’
 b. *Ki kit hívott meg?* MULTIPLE WH-FRONTING
 who.NOM who.ACC invited PRT
 Literal: ‘Who whom invited?’

We take as our starting point that an ellipsis site contains structure that is deleted or left unpronounced (Merchant 2018), and furthermore that this structure is a wh-question isomorphic to the antecedent. Given that both single wh-fronting (7-a) and multiple wh-fronting (7-b) questions are available in Hungarian, multiple sluicing could in principle be derived from either of them. These two potential sources are illustrated in Structures A and B below:

- (8) *Valaki meghívott valakit, de nem tudom, ...*
 someone invited someone.ACC but not I.know
 ‘Someone invited someone. But I don’t know ...’

² PRT denotes verbal particles in Hungarian, which belong to the class of verbal modifiers and form a complex predicate with the verb. In sentences with negation or structural focus, the verbal particle occurs postverbally (as in (7-a)–(7-b)); otherwise, it precedes the verb (see (6)). The precise meanings of the different particles are not relevant in this paper, and therefore we gloss them as PRT throughout.

- (10) A: *János kit kinek mutatott be?* (É. Kiss 2002, ex. 69)
 John who.ACC who.to introduced PRT
 ‘Who did John introduce to whom?’
- B: *Pétert Marinak és Évának, Zoltánt Évának és Júliának,*
 Peter.ACC Mary.to and Eva.to Zoltan.ACC Eva.to and Julia.to
Istvánt pedig Júliának és Marinak mutatta be.
 Istvan.ACC and Julia.to and Mary.to introduced PRT
 ‘He introduced Peter to Mary and Eva, Zoltan to Eva and Julia, and Istvan to Julia and Mary.’

Contrasting with É. Kiss’s judgements, Surányi (2006) claims that single wh-fronting questions license both a pair-list and a single-pair answer, as (11) shows. However, he shares É. Kiss’s judgement that multiple wh-fronting questions must have a pair-list answer, as in (12):

- (11) A: *Ki nézett rá kire?* (Surányi 2006, ex. 28)
 who looked PRT who.on
 ‘Who looked at who?’
- B: *János nézett rá Marira, Pali Gabira, ...*
 John looked PRT Mary.on Paul Gaby.on
 ‘John looked at Mary, Paul looked at Gaby, ...’
- B’: *János nézett rá Marira.*
 John looked PRT Mary.on
 ‘John looked at Mary.’
- (12) A: *Ki melyik tárgyat tanítja?* (Surányi 2006, ex. 27)
 who which subject.ACC teaches
 ‘Who teaches which subject?’
- B: *Pál a szintaxist tanítja, Márk a szintaxist és a*
 Paul the syntax.ACC teaches Mark the syntax.ACC and the
morfológiát, ...
 morphology.ACC
 ‘Paul teaches syntax, Mark teaches syntax and morphology, ...’
- B’: *#Pál a szintaxist tanítja.*
 Paul the syntax.ACC teaches
 ‘Paul teaches syntax.’

While Gazdik (2010) only includes multiple questions in her examples, rather than question-answer pairs, she does provide judgements on the available readings of the questions—these judgements align with Surányi’s (2006). In particular, Gazdik (2010) claims that multiple wh-fronting questions only have a pair-list reading

(see her Sections 3.1.1 and 6), but single *wh*-fronting questions allow for both single-pair and pair-list readings, though single-pair is the one usually expected (Section 3.1.2). Similarly, Bîlbîie and Gazdik (2012) say that multiple *wh*-fronting questions ('paratactic *wh*-phrases' in their terminology) exclusively license pair-list answers, and only single *wh*-fronting questions allow single-pair answers (Section 2.2.3); whether the latter type of question may also allow pair-list answers is not discussed by these authors.

Finally, van Craenenbroeck and Lipták (2013) agree with the previous authors: multiple *wh*-fronting questions must have a pair-list answer, as the paraphrases to the question in example (13) illustrate (see also É. Kiss 1993):

- (13) *Ki kinek hagyott egy üzenetet?*
 who who.to left a message.ACC
 'Who left a message for whom?'
 a. Available: Everyone left a message for someone. I wonder who each person left a message for.
 b. Unavailable: A single person left a message for someone. I wonder who the person was and for whom he left a message.
 (van Craenenbroeck and Lipták 2013, ex. 66)

Crucially, according to van Craenenbroeck and Lipták, multiple sluicing is only compatible with a pair-list interpretation (14), promoted by *mindenki* 'everyone' in their examples (see also Nishigauchi 1998 for similar examples in Japanese and Merchant 2001 for similar examples in English):

- (14) a. *Mindenki hagyott egy üzenetet valakinek. Nem tudom,*
 everyone left a message.ACC someone.to not I.know
hogy ki kinek.
 that who who.to
 'Everyone left a message for someone. I don't know who for whom.'
 b. **Valaki hagyott egy üzenetet valakinek. Nem tudom,*
 someone left a message.ACC someone.to not I.know
hogy ki kinek.
 that who who.to
 'Someone left a message for someone. I don't know who for whom.'
 (van Craenenbroeck and Lipták 2013, ex. 67–68)

On the basis of examples such as (13) and (14), following the assumption that there is a strict parallel between elliptical constructions and their non-elliptical counterparts, van Craenenbroeck and Lipták argue that multiple sluicing must be derived from multiple *wh*-fronting questions.

To sum up, there is disagreement in the literature regarding the interpretations of single *wh*-fronting multiple questions, viz. whether they license only a single-pair reading, or both single-pair and pair-list readings. With respect to multiple *wh*-fronting multiple questions, existing literature agrees that they allow only pair-list readings. Finally, multiple sluicing has been claimed to only be allowed with a pair-list interpretation. This was then taken as evidence that the source for multiple sluicing is multiple *wh*-fronting questions. No previous work has, to our knowledge, reported explicit judgements on all three relevant constructions. Table 1 summarizes the existing claims:

Table 1: Summary of existing claims regarding the available interpretations in Hungarian multiple sluicing and multiple questions.

	Multiple <i>wh</i> -fronting	Single <i>wh</i> -fronting	Multiple sluicing
É. Kiss (2002)	PL reading	SP reading	–
Surányi (2006)	PL reading	SP & PL reading	–
Gazdik (2010)	PL reading	SP & PL reading	–
Bílbíe and Gazdik (2012)	PL reading	at least SP reading	–
van Craenenbroeck and Lipták (2013)	PL reading	–	PL reading

It is important to note that none of the reported judgements have been subjected to rigorous experimental testing. Additionally, the examples provided in prior work did not always come in minimal pairs, which may have contributed confounding factors, calling into question the universality of the reported judgements. Some of these possibly confounding factors are (i) the use of D-linked *wh*-phrases (e.g. *which NP*) versus the use of non-D-linked *wh*-phrases (e.g. *who*) (see examples (11)–(12)), (ii) the use of transitive and ditransitive verbs (see examples (9)–(10) on the one hand, and examples (11)–(12) on the other hand), (iii) the presence or absence of a verb in the answer, (iv) the different available positions of the verb in the answer (Verb-Object versus Object-Verb), and (v) the presence or absence of a verbal particle, which in Hungarian indexes focus movement.

3 Experiment 1: acceptability rating

As mentioned, reported judgements about the interpretations (single-pair vs. pair-list) of multiple sluicing and multiple questions in Hungarian have not been subjected to rigorous experimental testing. Additionally, there is also disagreement among authors in prior work. To address this issue, we conducted two

acceptability rating studies, which differ in the kind of controls they use to establish a baseline for unacceptability.³

3.1 Experiment 1a

In Experiment 1a, we report on the first acceptability judgement study that tested the interpretations of Hungarian multiple sluicing and multiple questions.

3.1.1 Participants, task and materials

Fourty five native speakers of Hungarian, recruited on social media, participated in the experiment, which was administered on the Ibex platform (Drummond 2007). Participants were compensated 1,000 HUF or €3.

Participants saw dialogues such as the one in (15), and had to rate on a 1–7 Likert scale how acceptable a single-pair or pair-list answer (i.e., B’s answer) is to the relevant question (i.e., A’s question). This methodology has been used successfully to test the answerhood conditions of questions in English (see i.a. Achimova et al. 2013).

- (15) A: {Valaki / Mindenki} *meghívott valakit. Tudod, hogy ...*
 {Someone / Everyone} PRT.invited someone.ACC you.know that ...
- a. ... *ki kit?*
 who who.ACC
- b. ... *ki hívott meg kit?*
 who invited PRT who.ACC
- c. ... *ki kit hívott meg?*
 who who.ACC invited PRT
- ‘Someone/Everyone invited someone. Do you know who (invited) who?’
- B: {Mari Jánost /
 Mary John.ACC /
Mari Jánost, Péter Zsuzsit, Ádám pedig Évát.}
 Mary John.ACC Peter Susie.ACC Adam and Eva.ACC

³ We conducted an earlier experiment that tested the acceptability/grammaticality of the three relevant constructions, without manipulating answers. For multiple sluicing, in particular, this experiment asked participants ($N = 38$) to rate the acceptability (1–7) of stimuli such as: {*Someone/Everyone*} *invited someone. But I don’t know who whom.* Multiple sluicing received high acceptability ratings (mean = 6.05, SE = 0.12)—we take this to confirm that unlike in other languages like English (see, e.g., Lasnik 2014), this construction is perfectly grammatical in Hungarian.

The experiment had a 3×2 design: we tested three Constructions (multiple sluicing—15a, single wh-fronting questions—15b, multiple wh-fronting questions—15c) in two different Readings (single-pair and pair-list). Readings were promoted by a preceding sentence (*Someone ...* for single-pair and *Everyone ...* for pair-list), as well as, importantly, by a matching explicit single-pair/pair-list answer given in a dialogue context. Experimental items were identical to (15) in the following respects: verbs assigned the accusative case, verbs included a verbal particle (signalling focus movement, É. Kiss (2002)), and answers did not include the verb. The decision not to include the verb was motivated by the following: an SOV answer would have paralleled, and therefore biased toward, a multiple wh-fronting question, while an SVO answer would have paralleled and biased toward a single wh-fronting one.

Before the start of the experiment, 3 practice trials were included to familiarize participants with the task. Each participant saw 18 experimental trials, administered in a Latin Square design, as well as 30 filler trials. Three types of fillers were included: acceptable fillers, where the answer was an unambiguously good one (e.g. Q: *Today's exam was really hard. Did everyone fail?* A: *No, two people passed.*); unacceptable fillers, where the answer clearly did not address the question (e.g. Q: *Every child went skiing in February. Do you know where?* A: *Over Christmas.*); and medium fillers, where the answer given was a partial answer (e.g. Q: *Oh my God, there isn't any cake left! Which girls ate it?* A: *Mary.*).

3.1.2 Predictions

Based on judgements reported by native speaker authors in existing literature (see Table 1), we can make the following predictions. For multiple wh-fronting questions, as well as multiple sluicing, we should find high acceptability ratings for pair-list answers, but low acceptability ratings for single-pair answers. For single wh-fronting questions, on the other hand, we may either find that pair-list and single-pair answers both receive high acceptability ratings (with a potential preference for single-pair), as claimed by Surányi (2006) and Gazdik (2010), or we may find that only single-pair answers receive high acceptability ratings, and pair-list answers receive low ratings, as claimed by É. Kiss (2002).

3.1.3 Results and discussion

Figure 1 shows the results of Experiment 1a: a violin plot of the acceptability ratings for single-pair/pair-list answers as potential responses to the three relevant constructions. For the statistical analysis, a linear mixed effects model (lmer from the lme4 package in R, Bates et al. 2015) was fit, predicting Ratings by Reading (single-

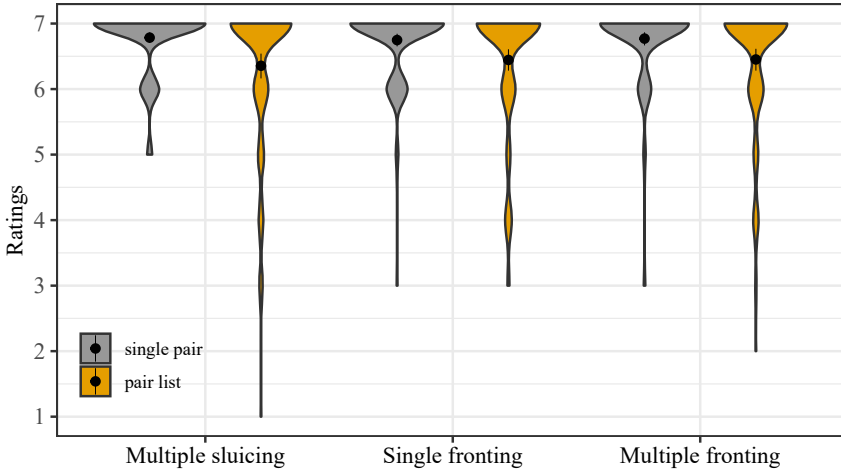


Figure 1: Acceptability rating results in Experiment 1a, including condition means and bootstrapped 95% confidence intervals.

pair vs. pair-list), Construction (multiple sluicing vs. single wh-fronting questions vs. multiple wh-fronting questions) and their interaction. The model included the maximal random effects structure supported by the data (Barr et al. 2013): random intercepts for participants. Because our main hypothesis concerns whether multiple sluicing differs from either of the non-elliptical question types, the fixed effects predictor of Construction was simple coded, with multiple sluicing serving as the reference level. The Reading predictor was sum-coded, with the level single-pair mapping to -0.5 and pair-list to 0.5 . We found a significant effect of Reading ($\beta = -0.35$, $SE = 0.04$, $t = -7.82$, $p < 0.001$), such that single-pair readings were rated higher than pair-list readings. However, there was no effect of Construction: neither single wh-fronting questions ($\beta = 0.03$, $SE = 0.05$, $t = 0.47$, $p = 0.64$), nor multiple wh-fronting questions ($\beta = 0.04$, $SE = 0.05$, $t = 0.74$, $p = 0.46$) differed significantly from multiple sluicing. There were also no significant interactions: the acceptability of pair-list versus single-pair readings did not differ either between sluicing and single wh-fronting ($\beta = 0.13$, $SE = 0.11$, $t = 1.15$, $p = 0.25$) or between sluicing and multiple wh-fronting ($\beta = 0.11$, $SE = 0.11$, $t = 1$, $p = 0.31$).

As can be seen Figure 1, all conditions received high acceptability ratings across the board, despite previous literature uniformly claiming that multiple sluicing and multiple wh-fronting questions do not license single-pair readings. We did, however, observe differences between the acceptability of single-pair and pair-list answers, to wit: single-pair answers were rated higher than pair-list answers for all constructions. It should be noted that the overall high ratings are not

due to participants' inability to do the task: unacceptable fillers received low ratings (mean = 1.6), while acceptable fillers received high ratings (mean = 6.76), and medium fillers fell between these two (mean = 3.11). This suggests that participants were able to determine whether an answer was an acceptable answer to a given question, and nonetheless gave high ratings to our critical stimuli sentences.

Overall, our acceptability judgement findings show that all three constructions pattern alike: single-pair and pair-list interpretations are equally acceptable for multiple sluicing, as well as both types of multiple questions. This poses a significant challenge to the claim that multiple sluicing parallels only multiple wh-fronting questions, and that therefore multiple wh-fronting questions must be the source of ellipsis. Based on possible interpretations, it is possible that multiple sluicing is derived from either single or multiple wh-fronting questions.

Nonetheless, before we completely reject the possibility that answerhood conditions can serve as a diagnostic for the source of multiple sluicing in Hungarian, we conduct two more experiments, which address potential reasons that the results of Experiment 1a did not align with judgements reported in prior work.

3.2 Experiment 1b: acceptability rating with better controls

In this section we explore a possible reason that the findings of Experiment 1a did not align with the judgements reported in prior work, which were made by authors who were trained linguists. For this, we draw inspiration from Achimova et al. (2015), who found (in their Experiment 1) that naive native speaker participants diverged from reported judgements and from experiments conducted with trained linguists. Specifically, they found that naive speakers accepted pair-list answers for questions with both subject and object quantifiers, while linguists only accept them for questions with subject quantifiers. Achimova et al. (2015) hypothesized that this difference across populations is due to naive experimental participants being more susceptible to noise than trained linguists (see i.a. Devitt 2006; Newmeyer 1983, 2007; Spencer 1973), specifically noise introduced by the type of fillers in the experiment. In Achimova et al.'s (2015) Experiment 1, unacceptable fillers represented incoherent discourse, and this was also the case in our Experiment 1a (e.g. Q: *John met a girl. Is it clear to you who?* A: *I think he likes Mary.*). As the authors reason, the presence of such fillers may have created too strong an expectation of what counts as an unacceptable answer, and as a result naive speakers accepted grammatically deviant, but pragmatically coherent pair-list answers in experimental items. In a follow-up experiment that included fillers with violations that better paralleled those in experimental items, naive participants' judgements were aligned with reported expert judgements: they rejected pair-list answers to object

quantifier questions. The authors take this as evidence that the reason the experimental data diverged from reported judgements in the first place was that, in contrast with the strong deviance of fillers, participants were able to accommodate the milder deviance of experimental items.

In our Experiment 1b, we test whether a similar explanation can be given to why Experiment 1a's findings conflict with judgements reported in the literature. To this end, we replace pragmatically incoherent fillers with ones where the violation better matches those found in experimental items.

3.2.1 Participants, task and materials

Thirty nine native speakers of Hungarian, recruited on social media, participated in the experiment, which was administered on the PCIbex platform (Zehr and Schwarz 2018). Participants were compensated via a lottery, where one participant won 25,000 HUF.

Experiment 1b had the same task, experimental and practice items and general design as Experiment 1a. Participants rated on a 1–7 Likert scale how acceptable a given answer is to a question, and we manipulated Construction (multiple sluicing vs. single wh-fronting vs. multiple wh-fronting) via the question, and Reading (single-pair vs. pair-list) via the preceding context and the answer—see (15). Crucially, however, the filler items in Experiment 1b were different from those in Experiment 1a. Instead of using pragmatically incoherent answers as unacceptable fillers, Experiment 1b included questions with downward entailing quantifiers such as *nobody*, *most* and *few*, resulting in question-answer pairs such as (16)–(17).

- (16) A: *A konferencia hotelben svédasztalos reggeli volt.*
 the conference in.hotel buffet breakfast be.PST
Milyen ételt nem evett senki reggelire?
 what.kind.of food.ACC not ate nobody for.breakfast
 ‘The conference hotel had all-you-can-eat breakfast. What food item did nobody eat for breakfast?’
- B: *Dénes nem evett gofrit, Emma palacsintát,*
 Dénes not ate waffle.ACC Emma pancake.ACC
Henrik pedig kiflit.
 Henrik and bun.ACC
 ‘Dénes did not eat waffles, Emma did not eat pancakes, and Henrik did not eat buns.’

- (17) A: *Az iskolai ebédlöben sokféle ital volt. Milyen the school in.cafeteria many.kinds.of drink be.PST what.kind.of üdítöt ivott csak kevés diák a reggelihez? drink.ACC drank only few student the for.breakfast*
 ‘The school cafeteria has many drinks. Which type of juice did few students drink for breakfast?’
- B: *Szonja almalevet ivott. Szonja apple.juice.ACC drank*
 ‘Szonja drank apple juice.’

To create such unacceptable fillers, we adapted and modified Achimova et al. (2015)’s materials e.g. by adding an initial context sentence to the questions, to ensure that they were not systematically different from our experimental items. In total, Experiment 1b included 42 filler items: 18 unacceptable fillers that all had a(n unacceptable) pair-list or single-pair answer (see above); 18 acceptable fillers, 8 of which resembled unacceptable fillers and 10 of which were unrelated to pair-list/single-pair readings; and 6 medium fillers (where the answer given was a partial answer). Similarly to Experiment 1a, there were 18 experimental and 3 practice items.

3.2.2 Predictions

If our Experiment 1a results only diverged from claims in prior work due to naive participants’ sensitivity to fillers items, Experiment 1b should confirm judgements reported in the literature. That is, we should find that for multiple wh-fronting questions and multiple sluicing, only pair-list answers are acceptable, while for single wh-fronting questions, either both answer types are acceptable, or only single-pair is.

3.2.3 Results and discussion

Figure 2 shows the results of Experiment 1b: a violin plot of the acceptability ratings for single-pair/pair-list answers as potential responses to the three relevant constructions. The statistical analysis was identical to the one conducted for Experiment 1a, with the model including random intercepts for both participants and items. The analysis revealed a significant effect of Reading ($\beta = -0.46$, $SE = 0.06$, $t = -8$, $p < 0.001$), such that single-pair readings were rated higher than pair-list readings. We found no effect of Construction: neither single wh-fronting questions ($\beta = 0.02$, $SE = 0.07$, $t = 0.23$, $p = 0.82$), nor multiple wh-fronting questions ($\beta = -0.07$, $SE = 0.07$, $t = -0.98$, $p = 0.33$) differed significantly from multiple

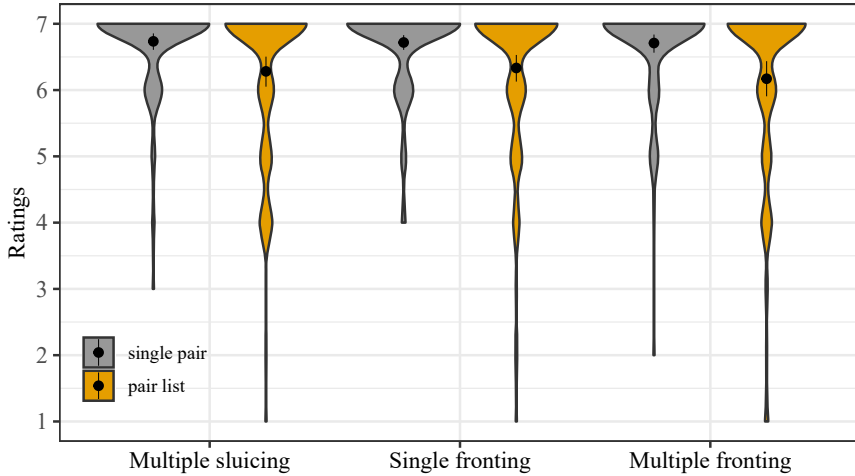


Figure 2: Acceptability rating results in Experiment 1b, including condition means and bootstrapped 95% confidence intervals.

sluicing. There were also no significant interactions: the acceptability of pair-list versus single-pair readings did not differ either between sluicing and single wh-fronting ($\beta = 0.07$, $SE = 0.14$, $t = 0.49$, $p = 0.63$), or between sluicing and multiple wh-fronting ($\beta = -0.1$, $SE = 0.14$, $t = -0.67$, $p = 0.5$).

In other words, we replicate the findings of Experiment 1a (compare Figures 1 and 2): all conditions received high ratings across the board, suggesting that both single-pair and pair-list readings are available in all three relevant constructions, though single-pair readings are slightly preferred. Just like in Experiment 1a, filler results demonstrate that participants were in general able to complete the task and distinguish acceptable from unacceptable answers: unacceptable fillers received low ratings (mean = 2.29), acceptable fillers received high ratings (mean = 6.76), and medium fillers received ratings in between these two (mean = 3). This suggests that the reason we obtained overall high ratings for all experimental conditions was not that participants simply gave a high rating to every trial.

The goal of Experiment 1b was to test whether what caused the observed divergence of naive speakers' judgements from existing theoretical literature was that naive speakers are more sensitive to noise—more specifically, that they may have accommodated the violation of unavailable single-pair/pair-list readings when our experiment also included pragmatically incoherent fillers. However, what we found is that even when the kind of violation included in filler items more closely matches those in experimental items, we still fail to find evidence for claims

reported in the literature: namely, that single-pair answers are unacceptable for multiple wh-fronting questions and multiple sluicing.

4 Experiment 2: forced choice

As we have seen, the findings of Experiment 1a did not align with judgements reported in existing theoretical work, and this finding persisted even after we ruled out the potential effect of extreme fillers in Experiment 1b. In Experiment 2, we explore the possibility that the discrepancy between the results of Experiments 1a–b and the judgments reported in prior literature are due to methodological differences. In order to do this, Experiment 2 utilized a forced choice task instead of an acceptability rating task. Participants had to choose what constitutes a better answer to a given question: a single-pair or a pair-list one. It is possible that such a two-alternative forced choice task resembles the process underlying introspective grammaticality judgments better than an acceptability rating task does. If this is the case, then the findings of Experiment 2 should better align with what has been claimed in the literature.

Additionally, Experiments 1a–b revealed an overall preference for single-pair answers across all constructions. With the forced choice task of Experiment 2, we are able to test the magnitude of this preference, since this paradigm provides a natural baseline (50%).

4.1 Participants, task and materials

Thirty nine native speakers of Hungarian, recruited on social media, participated in the experiment, which was administered on the Ibex platform (Drummond 2007). Participants were compensated 1,000 HUF or €3.

Experiment 2 was a forced choice task: participants again saw dialogues such as the one in (18), but this time their task was to choose between a single-pair and a pair-list answer (i.e., B's answer) in response to a question (i.e., A's question). They were instructed to choose the option that was a better/more acceptable/more appropriate answer to A's question.

- (18) A: *Valaki, vagy valakik meghívtak valakit. Tudod, Someone.SG or someone.PL PRT.invited someone.ACC you.know hogy ... that ...*

- a. ... *ki* *kit?*
 who who.ACC
- b. ... *ki* *hívott* *meg* *kit?*
 who invited PRT who.ACC
- c. ... *ki* *kit* *hívott* *meg?*
 who who.ACC invited PRT
- ‘Someone, or some people invited someone. Do you know who (invited) who?’
- B: {*Mari* *Jánost.* / *Mari* *Jánost,* *Péter* *Zsuzsit,* *Ádám*
 Mary John.ACC / Mary John.ACC Peter Susie.ACC Adam
 pedig *Évát.*}
 and Eva.ACC

As before, we manipulated the type of construction that the question contained: multiple sluicing (18a), single wh-fronting questions (18b), or multiple wh-fronting questions (18c). The preceding context sentence was modified to allow for both single-pair and pair-list readings (“*Someone.SG or Someone.PL invited ...*”). Experimental items were identical to those used in Experiments 1a–b, with the only difference being in the preceding context sentence.

Before the start of the experiment, 3 practice trials were included to familiarize participants with the task. Each participant saw 18 experimental trials, administered in a Latin Square design, as well as 30 filler trials. Three types of fillers were included: questions where one potential answer was acceptable and one was unacceptable (e.g. Q: *There were lots of things in the mail today. Who wrote a letter to Fanni?* A1: *David.* A2: *Yesterday.*); questions where both answers were potentially acceptable answers (e.g. Q: *I had ice cream yesterday. Guess which flavor!* A1: *Maybe vanilla.* A2: *Maybe vanilla and chocolate.*); and questions where both answers were acceptable, but the choice potentially depended on interpretation (e.g. Q: *Oh my God, there isn't any cake left! Which girl or which girls ate it?* A1: *Mary.* A2: *Mary and Susan.*).

4.2 Predictions

If a forced choice task better resembles the process underlying the introspective judgements reported in earlier work, and the discrepancy between that work and our Experiments 1a–b is merely due to methodological difference, then Experiment 2 should reveal the following results. We should find that pair-list answers are chosen most of the time for multiple wh-fronting questions and multiple sluicing.

For single wh-fronting we may find either that single-pair answers are the preferred choice, or that both potential answers are chosen with equal likelihood.

4.3 Results and discussion

Figure 3 shows the percentage of single-pair answers being chosen for each construction type in Experiment 2. For the statistical analysis, a logistic mixed effects model (lmer from the lme4 package in R, Bates et al. 2015) was fit, predicting Response (single-pair vs. pair-list) by Construction (multiple sluicing vs. single wh-fronting questions vs. multiple wh-fronting questions). Similarly to Experiments 1a–b, the fixed effects predictor of Construction was simple coded, with multiple sluicing serving as the reference level. The model included the maximal random effects structure supported by the data (Barr et al. 2013): random intercepts for participants and items. Neither single wh-fronting questions ($\beta = 0.32$, $SE = 0.25$, $z = 1.28$, $p = 0.2$), nor multiple wh-fronting questions ($\beta = -0.39$, $SE = 0.24$, $z = -1.6$, $p = 0.11$) were found to differ significantly from multiple sluicing. However, an additional pair comparison between the two non-elliptical questions revealed a significant difference ($\beta = -0.71$, $SE = 0.25$, $z = -2.85$, $p < 0.01$): single wh-fronting questions led to significantly more single-pair answers than multiple wh-fronting questions.

In Experiment 2, as in Experiments 1a–b, all three constructions patterned alike in showing a preference for single-pair answers; the single-pair answer was

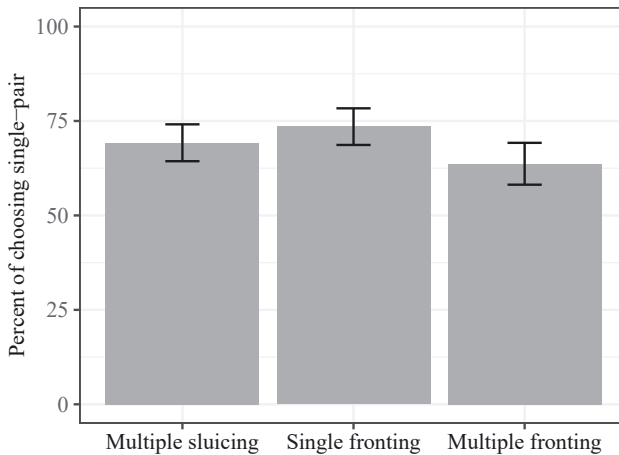


Figure 3: Percent of choosing single-pair answers in Experiment 2. Error bars represent standard error.

chosen at a rate (well) above the baseline 50% for all constructions. Additionally, we see that multiple sluicing does not differ from either type of non-elliptical question in how strong the single-pair preference is. Thus, it seems that even when we utilize a different experimental paradigm, one that perhaps better resembles the way introspective judgments are arrived at, experimental data still does not support the reported judgements from the literature: we did not find that in multiple wh-fronting questions and multiple sluicing, pair-list answers are strongly preferred. Altogether, neither experiment's results are in line with the claims made in existing literature regarding the availability of single-pair and pair-list readings in the three relevant constructions.

5 General discussion

Existing theoretical literature on Hungarian uniformly claims that multiple sluicing and multiple wh-fronting questions are only compatible with pair-list interpretations. This parallel in interpretation has been used as evidence that multiple sluicing is derived from multiple wh-fronting questions (van Craenenbroeck and Lipták 2013). As for single wh-fronting questions, the literature is divided regarding their answerhood conditions: whether only single-pair answers are allowed, or both single-pair and pair-list answers (with a potential preference for single-pair). In this paper, we presented the first experimental investigation of the relevant empirical domain. Altogether, most previously reported judgements regarding the interpretations of multiple sluicing and multiple questions in Hungarian were not confirmed by our findings. We repeat Table 1 below, adding the results of our experiments.

In Experiment 1a we found that Hungarian multiple sluicing, single wh-fronting questions, and multiple wh-fronting questions pattern alike with respect to their answerhood conditions: single-pair answers are slightly preferred over pair-list ones across the board, but both answer types are clearly available. In Experiment 1b, we ruled out the possibility that experimental participants' judgements only deviated from those reported in the literature due to the potential effect of pragmatically incoherent filler items. In Experiment 2, we investigated whether judgements of (un)acceptability are better reflected in a different experimental paradigm (forced choice, instead of acceptability rating), but our findings again did not align with reported claims. As can be seen in Table 2, the only claims from the literature that our experimental findings support are Surányi's (2006) and Gazdik's (2010) regarding single wh-fronting. Importantly, across three experiments, we failed to find evidence of there being a strict parallel between multiple sluicing and one, but not the other, type of multiple question. This poses a

Table 2: Summary of existing claims (and our experimental findings) regarding the available interpretations in Hungarian multiple sluicing and multiple questions.

	Multiple wh-fronting	Single wh-fronting	Multiple sluicing
É. Kiss (2002)	PL reading	SP reading	–
Surányi (2006)	PL reading	SP & PL reading	–
Gazdik (2010)	PL reading	SP & PL reading	–
Bílbíe and Gazdik (2012)	PL reading	at least SP reading	–
van Craenenbroeck and Lipták (2013) this paper	PL reading SP & PL reading	– SP & PL reading	PL reading SP & PL reading

challenge to using interpretations as a diagnostic for determining the source of multiple sluicing in Hungarian. Assuming that properties of non-elliptical sentences predict properties of elliptical ones, there is no reason, in principle, to prefer analyzing multiple sluicing as deriving from either type of multiple wh-question.

Additionally, Experiment 2 revealed that multiple sluicing does not differ from either type of question in the strength of the preference for a single-pair interpretation; rather, it is (at least numerically) ‘in between’ the two types of questions. This raises the possibility that multiple sluicing represents a middle ground when it comes to interpretation, which in turn might suggest that both kinds of questions are available as its source. We illustrate this potential optionality of sources in (19).

- (19) *Valaki/Mindenki meghívott valakit. De nem tudom,*
 someone/everyone invited someone.ACC but not I.know
ki kit.
 who.NOM who.ACC
 ‘Someone/Everyone invited someone. But I don’t know who whom.’
- a. move-and-delete approach
 ... *De nem tudom, ki kit* [C *hívott meg*].
 ... but not I.know who.NOM who.ACC invited PRT
- b. in-situ approach
 ... *De nem tudom, ki* [C *hívott meg* [kit]_F].
 ... but not I.know who.NOM invited PRT who.ACC

Example (19-a) is the structure representing the scenario where multiple sluicing is derived from multiple wh-fronting questions. This can be captured under the move-and-delete approach to ellipsis, which posits that both wh-phrases are moved, and thus both escape deletion, which targets the complement of C (i.a. Merchant 2001; van Craenenbroeck and Lipták 2013). On the other hand, if multiple

sluicing is derived from single *wh*-fronting questions, that structure points to an in-situ approach to ellipsis, where one of the *wh*-phrases escapes deletion without needing to move (i.a. Abe 2015, 2016; Stigliano 2022), as shown in (19-b).

Of course, our findings do not entirely eliminate the possibility that multiple sluicing in Hungarian is definitively derived from either single or multiple *wh*-fronting questions. It is conceivable that other diagnostics might reveal evidence in favor of one underlying source or the other; what our experimental findings show is that the interpretation of the relevant structures cannot constitute such evidence. Future work should also investigate the answerhood conditions of a greater variety of multiple question and multiple sluicing structures, systematically varying e.g. D-linking, transitivity, or the presence of the verb, which were left uncontrolled in prior theoretical work.

Finally, our results highlight the importance of rigorous experimental testing in the development of linguistic theory. Although Phillips (2009) has shown that in many cases ‘informal’ judgments are as robust as rigorous experimentation, the latter type of data can shed light on the answers to more nuanced questions – especially those for which different authors disagree about the empirical facts. Within the domain of multiple sluicing, there has been some experimental work (see, e.g., Chung and Park 2017; Cortés Rodríguez submitted, Cortés Rodríguez this volume; Lasnik 2014), a body of research that this paper contributes to. But besides their empirical value, our results show that experimental methods are a useful tool for reassessing and clarifying conflicting judgments in the literature generally.

6 Conclusion

Claims about the answerhood conditions of Hungarian multiple sluicing and single/multiple *wh*-fronting questions have been made on the basis of heterogeneous examples. Additionally, there has been some disagreement among authors in existing work, and the relevant judgments have not previously been subjected to experimental testing. The novel experimental data reported in this paper provides a challenge to existing judgements, and suggests that all relevant constructions pattern alike in that they license both single-pair and pair-list answers, with a preference for single-pair. Given that no difference is found among the three constructions, answerhood conditions cannot distinguish between the two possible sources (single vs. multiple *wh*-fronting questions) for the ellipsis site in multiple sluicing. Therefore, *contra* previous claims, answerhood conditions do not provide evidence for multiple sluicing deriving from multiple *wh*-fronting.

Acknowledgment: For their helpful comments and suggestions, we thank the three anonymous reviewers, Karlos Arregi, Michael Tabatowski and Ming Xiang, as well as the audiences at the Morphology and Syntax Workshop at the University of Chicago, TLS 2019, WCCFL 38 and ECBAE 3. We are grateful to Asya Achimova for sharing experimental materials with us. We thank the Humanities Division Council at the University of Chicago for their generous financial support. All errors are our own.

References

- Abe, Jun. 2015. *The in-situ approach to sluicing*, 222. Amsterdam: John Benjamins Publishing Company.
- Abe, Jun. 2016. Make short answers shorter: Support for the *in situ* approach. *Syntax* 19. 223–255.
- Abels, Klaus & Veneeta Dayal. 2017. On the syntax of multiple sluicing. In Andrew Lamont & Katerina Tetzloff (eds.), *NELS 47: Proceedings of the 47th annual meeting of the North East Linguistic Society*, 1–20. Amherst, MA: GLSA Publications.
- Achimova, Asya, Peter Staroverov, Viviane Déprez & Julien Musolino. 2015. *Pair-list answers in naïve speakers and professional linguists*, 21–35. Moscow: LRC Publishers.
- Achimova, Asya, Viviane Deprez & Julien Musolino. 2013. What makes pair list answers available: An experimental approach. In Nicholas LaCara, Yelena Fainleib & Yangsook Park (eds.), *NELS 41: Proceedings of the 41st annual meeting of the North East Linguistic Society*, 1–14. Amherst, MA: GLSA.
- 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. 255–278.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48.
- Bílbíle, Gabriela & Anna Gazdik. 2012. Wh-coordination in Hungarian and Romanian multiple questions. In Christopher Piñón (ed.), *Empirical issues in syntax and semantics*, vol. 9, 19–36.
- Chung, Wonil & Myung-Kwan Park. 2017. Multiple sluicing and swiping meet RT-based experimental syntax. *Studies in Linguistics* 42. 115–141.
- Cortés Rodríguez, Alvaro. submitted. Which linguist which structure: An experimental investigation of multiple sluicing in English. In Andreas Konietzko & Susanne Winkler (eds.), *Information structure and discourse in generative grammar. Mechanisms and processes*. Berlin and Boston: De Gruyter Mouton.
- Cortés Rodríguez, Alvaro. this volume. Multiple sluicing and islands: Cross-linguistic experimental investigation on the clausemate condition. *The Linguistic Review*.
- van Craenenbroeck, Jeroen & Anikó Lipták. 2013. What sluicing can do, what it can't and in which language: On the cross-linguistic syntax of ellipsis. In Lisa L.-S. Cheng & Norbert Corver (eds.), *Diagnosing syntax*, 502–536. Oxford: Oxford University Press.
- Devitt, Michael. 2006. *Ignorance of language*. Oxford: Oxford University Press.
- Drummond, Alex. 2007. *Ibex farm*.

- É. Kiss, Katalin. 1993. Wh-movement and specificity. *Natural Language & Linguistic Theory* 11. 85–120.
- É. Kiss, Katalin. 2002. *The syntax of Hungarian*. *Cambridge Syntax Guides*. Cambridge: Cambridge University Press.
- Gazdik, Anna. 2010. Multiple question in French and in Hungarian. An LFG account. In *LFG10 conference*. Stanford, United States: CSLI Publications. Available at: <https://halshs.archives-ouvertes.fr/halshs-00751119>.
- Lasnik, Howard. 2014. Multiple sluicing in English? *Syntax* 17. 1–20.
- Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.
- Merchant, Jason. 2018. Ellipsis: A survey of analytical approaches. In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *The Oxford handbook of ellipsis*, 19–45. Oxford: Oxford University Press.
- Newmeyer, Frederick J. 1983. *Grammatical theory, its limits and possibilities*. Chicago: The University of Chicago Press.
- Newmeyer, Frederick J. 2007. Commentary on Sam Featherston, ‘data in generative grammar: The stick and the carrot’. *Theoretical Linguistics* 33. 395–399.
- Nishigauchi, Taisuke. 1998. ‘Multiple sluicing’ in Japanese and the functional nature of wh-phrases. *Journal of East Asian Linguistics* 7. 121–152.
- Phillips, Colin. 2009. Should we impeach armchair linguists. *Japanese/Korean Linguistics* 17. 49–64.
- Ross, John Robert. 1969. *Guess who?, Papers from the 5th Regional Meeting of Chicago Linguistic Society*. 252–286, Chicago: Chicago Linguistic Society.
- Spencer, Nancy J. 1973. Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability. *Journal of Psycholinguistic Research* 2. 83–98.
- Stigliano, Laura. 2022. P-omission in ellipsis in Spanish: Evidence for syntactic identity. *Natural Language & Linguistic Theory*. <https://doi.org/10.1007/s11049-021-09535-w>.
- Surányi, Balázs. 2006. Mechanisms of wh-saturation and interpretation in multiple wh-movement. In Norbert Corver & Lisa Lai Shen Cheng (eds.), *Wh-movement: Moving on*, chapter 11, 289–318. Cambridge, MA: MIT Press.
- Takahashi, Daiko. 1994. Sluicing in Japanese. *Journal of East Asian Linguistics* 3. 265–300.
- Tancredi, Christopher Damian. 1992. *Deletion, deaccenting, and presupposition*. Cambridge, MA: Massachusetts Institute of Technology Doctoral Dissertation.
- Zehr, Jérémy & Florian Schwarz. 2018. *Penn Controller for internet based experiments (IBEX)*.