Robin Lemke, Lisa Schäfer & Ingo Reich

2022

Can Identity Conditions on Ellipsis be Explained by Processing Principles?

in

Robin Hörnig, Sophie von Wietersheim, Andreas Konietzko & Sam Featherston (eds.),

Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data

pp. 541–561

Tübingen: University of Tübingen

https://publikationen.uni-tuebingen.de/xmlui/handle/10900/119301

In R. Hörnig, S. von Wietersheim, A. Konietzko & S. Featherston (eds.) (2022), *Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data*. Tübingen: University of Tübingen. https://publikationen.uni-tuebingen.de/xmlui/handle/10900/119301

Can Identity Conditions on Ellipsis be Explained by Processing Principles?

Robin Lemke, Lisa Schäfer & Ingo Reich
Saarland University
robin.lemke@uni-saarland.de

1 The Problem: Mismatches under Sluicing and Sprouting

In the theoretical literature, the unacceptability of (some) *structural mismatches* between the antecedent and the target of ellipsis have been taken to indicate that ellipsis is subject to syntactic identity conditions. Such constraints have been defended for verb phrase ellipsis (VPE) (Arregui et al., 2006; Merchant, 2013) and sluicing (Chung, 2006, 2013).[1] The assumption of syntactic identity conditions increases the complexity of the grammar, because conditions which are specific to particular ellipses must be added to a system of more general rules. If the data that apparently support syntactic identity conditions could be explained by independently motivated principles, this would consequently reduce the complexity of the syntactic system. In this article we investigate syntactic identity conditions proposed by Chung (2006, 2013) for sluicing, i.e. the ellipsis of the TP in a *wh*-question, which is survived only by the *wh*-phrase (1a) (Ross, 1969). Our study shows that apparent grammaticality contrasts can be accounted for by a probabilistic processing account, which is supported by an acceptability rating, a production and a self-paced reading experiment. In contrast, Chung's constraints lead to predictions which are not supported by our data.

The acceptability pattern that we investigate is exemplified in (1) and (2). The introspective observation by Chung (2006, 2013) ist that sluicing with an overt antecedent for the omitted TP (1a) is more permissive with respect to antecedent-target mismatches than *sprouting* (Chung et al., 1995), a variety of sluicing that lacks an overt antecedent (1b). Specifically, Chung notes that the omission of the preposition in the sluice[2] is acceptable under sluicing with an overt antecedent (2a), but not under sprouting (2b). In order to delimit it from sprouting, in what follows we restrict the term *sluicing* to instances with an overt antecedent.

(1) a. John danced with somebody, but I don't know with whom ⟨John danced⟩. (sluicing)
b. John danced, but I don't know with whom ⟨John danced⟩. (sprouting)

(2) a. John danced with somebody, but I don't know who ⟨John danced with⟩. (sluicing)
b. *John danced, but I don't know who ⟨John danced with⟩. (sprouting)

Chung explains this pattern with syntactic identity conditions that are specific to sluicing. In Chung (2006), she proposes the *Numeration Condition* (NC), which requires all words that are omitted in the sluice to be included in the numeration of the antecedent. This rules out (2b) since, unlike in (2a), the preposition *with* is not contained in the antecedent *John danced*. In Chung (2013), she observes that the NC incorrectly predicts (3a) to be ungrammatical, as the verb *met* is not contained in the numeration of the antecedent. Therefore, she replaces the NC

[1] Cf. Merchant (2001), who proposes e-givenness as a semantic identity condition.

[2] Since sprouting is a particular variety of sluicing, in what follows we refer to the elliptical phrase as the sluice independently of whether it has an overt antecedent in the first conjunct.

by two constraints, the *Argument Structure Condition* (ASC) and the *Case Condition* (CC). The ASC concerns only argument sluices and requires the argument structure of the target to match that of the antecedent. This constraint is motivated by the observation that the argument structure mismatch between antecedent and target is unacceptable if the sluice is an argument (3b), but not when it is an adjunct like in (3a).

(3) a. I remember meeting him, but I don't remember when ⟨I met him⟩. (Chung, 2013: 31)
b. *The vase was stolen, but we don't know who ⟨stole the vase⟩. (Chung, 2013: 31)

The ASC alone does not explain why (2b) is degraded, since the argument structure in the target is identical to that in the antecedent. For this reason, Chung (2013) proposes the CC, which requires sluiced DPs to be case-marked by a head that is identical to a head in the antecedent. Since she assumes that the sluiced DP *who* is case-marked by the omitted preposition *with*, which is not contained in the antecedent, (2b) is ruled out by the CC.[3]

Identity conditions like the two constraints proposed by Chung (2013) might be descriptively appropriate, but it is unclear how they are related to and derived from the theory of ellipsis and the syntactic system in general. For instance, Chung's constraints are specific to sluicing, so it would be necessary to somehow ensure that they do not constrain other ellipses.[4] In order to keep the theory of ellipsis as simple and uniform as possible, it is therefore desirable to explain the empirically observable pattern without resorting to identity constraints and to maintain only those that improve explanatory adequacy above independently motivated principles.

In this article we present a processing account in order to explain the acceptability of mismatches under sluicing and sprouting.[5] Our account is based on the general observation that unlikely expressions are harder to process (Hale, 2001; Levy, 2008). If mismatches are overall less likely, they are also harder to process. From this perspective, we expect that mismatches under sluicing are more acceptable because the explicit mention of e.g. the dancing partner in (2a) increases the likelihood of a sluice referring to this person. Our account provides an explanation for the data in (2) without assuming syntactic identity constraints, and additionally makes more fine-grained predictions on predictability effects that Chung's constraints cannot capture. We outline our account in Section 2 and present a series of acceptability rating, production and self-paced reading studies that test its predictions in Sections 3–7. Our experiments show that our processing account is capable of explaining the pattern observed by Chung.

2 A Processing Account of Sluicing and Sprouting Mismatch Acceptability

Most theoretical accounts of ellipses attribute the unacceptability of utterances to ungrammaticality, i.e. to the violation of syntactic rules. However, even utterances which fully conform to grammatical rules are also perceived as degraded if they are hard to process (see e.g. Sag et al.,

[3] Note that if only the elliptical versions of (2b) and (3b) are considered, it is impossible to determine which of the two constraints in Chung (2013) they violate. If the target is analyzed as containing the parallel argument structure and a stranded preposition (2b), it conforms to the ASC but violates the CC, whereas the opposite holds in the case of (3b). As we discuss in Section 3, we investigate the phenomenon in German, where only one way of resolving ellipsis is available.

[4] Some syntactic identity conditions follow from such general assumptions of the syntactic system. For instance Merchant (2013) argues that voice mismatches are acceptable under verb phrase ellipsis (VPE) but not under sluicing, because voice is encoded in a VoiceP above VP that is deleted under sluicing (TP deletion) but not under VPE. Hence, in (ia) there is an identical antecedent, whereas in (ib) there is none.

(i) a. This problem should have been looked into, but nobody did ⟨look into this problem⟩. (VPE)
b. *This problem should have been looked into, but I don't know who ⟨*shouldhavelookedintothisproblem*⟩. (sluicing)

[5] Poppels & Kehler (2019) make a similar proposal based on the likelihood of Questions under Discussion (Roberts, 1996).

2007). Processing effort in turn has been related to predictability: Expressions that are more likely to appear in context are more easily processed, as is reflected in reading times (see e.g. Hale, 2001; Levy & Jaeger, 2007; Levy, 2008; Demberg & Keller, 2008) or a decrease in the N400 amplitude (see e.g. Delogu et al. (2017) and the references therein).

Since Levy & Jaeger (2007), processing accounts based on this insight have been shown to explain the omission of pronouns (Levy & Jaeger, 2007), complementizers (Jaeger, 2010), articles (de Lange, 2008; Lemke et al., 2017), content words (Lemke, 2021) and topics (Schäfer, 2021). Taken together, this research shows that speakers use optional omissions to adapt their utterance to the limited processing resources that are available to the hearer, and they do so in two ways. First, predictable material is more likely to be omitted in order to avoid underutilization of the hearer's processing resources. Second, additional redundancy is inserted in order to counterbalance processing load by distributing it over more time.

With respect to the above sluicing examples, this predicts that the preference for ellipsis is stronger the more predictable the phrase potentially reduced by ellipsis is in the context of the preceding material. If the phrase is highly predictable, its reduction will increase the efficiency of communication. In contrast, the reduction of an unpredictable phrase might exceed the hearer's processing resources, because processing a *wh*-phrase like *who* in (4) requires not only to interpret the *wh*-phrase but also the reconstruction of the omitted material, which we expect to be particularly difficult when it is unpredictable. In that case, realizing the TP overtly distributes processing effort across more words, i.e. time, so speakers should prefer a full form.

(4) a. John danced with somebody, but I don't know who ⟨John danced with⟩. (sluicing)
b. *John danced, but I don't know who ⟨John danced with⟩. (sprouting)

From this perspective, the acceptability of ellipsis varies as a function of the likelihood of the reduced phrase. We hypothesize that this might explain at least some of the sluicing data discussed above: We expect that structural mismatches between antecedent and target under sluicing are relatively unlikely (see Section 6 for empirical evidence), but that not all mismatches are equally unlikely and consequently unacceptable. An explicit antecedent like *with somebody* in (4a) might increase the likelihood of a question asking *who John danced with* as compared to potential competitors (e.g. *at which club he danced* or *whether he enjoyed it*) more strongly than in (4b), where the antecedent of the sluiced phrase is implicit. If this is correct, we expect a stronger preference for ellipsis in (4a) than in (4b), where the nonelliptical utterance distributes the processing effort over more time. Note that the matching sluices (*with whom*), which are judged as acceptable in Chung (2006, 2013), also distribute processing effort over more words, and the preposition *with* makes the presence of the dancing partner more explicit. Therefore, we also predict a stronger preference for ellipsis than in mismatches with a potential DP sluice. So far, our processing account provides a nonsyntactic explanation for the data, which is based on independently motivated processing constraints instead of identity conditions specific to sluicing: Neither of the structures in (4) is strictly ungrammatical, but (4b) is harder to process.

Since our account is based on the likelihood of the sluiced phrase, it predicts not only a discrete contrast between sluicing and sprouting, or matches and mismatches, but that *any* cue that boosts the likelihood of the potentially sluiced phrase increases the preference for ellipsis. Specifically, we predict and investigate a gradual effect of the likelihood of a partner involved in the action described by the verb on the acceptability of ellipsis: In (4), it might be likely (though not necessary) that John has a dancing partner, but *to chat* in (5a) *requires* an interlocutor, who is left implicit under sprouting. , in (5b) it is possible that John has a studying partner, but this is intuitively less likely than the presence of a dancing partner in (4).

(5) a. John was chatting, but I don't know who.
b. John was studying, but I don't know who.

If these differences in likelihood are empirically confirmed, we expect them to constrain the acceptability of the corresponding sluices: The more likely an implicit second participant involved in the action described by the verb is, the more strongly is ellipsis preferred. These results would provide further evidence for a processing account of sluicing mismatches, since purely syntactic accounts (Chung, 2006, 2013) cannot explain predictability effects.

3 Experimental Rationale

3.1 Experimental Methods

Testing our processing account requires (i) to verify that matching continuations are more likely than mismatching ones, (ii) to determine how acceptable the respective ellipses and full forms are, and (iii) to measure their processing effort. We investigate this (i) with a production task, (ii) with an acceptability rating study and (iii) with a self-paced reading experiment. The production study assesses the likelihood of the continuations, and the rating and reading experiments test whether this likelihood is reflected in acceptability and processing effort.

3.2 Materials and Conditions

We conduct our studies in German because English DP sluices can be reconstructed in two ways: First, the DP can be a nominative subject DP (6a), and second, it can be the object of a stranded (and omitted) preposition (6b). While in English Chung's ASC would rule out (6a) because of the argument structure mismatch between antecedent and target, (6b) is predicted to be fine. Sluicing mismatches can therefore be grammatically derived from the structure with the stranded proposition. However, it is impossible to determine why sprouting (6c) is ungrammatical: If the ellipsis is resolved as in (6a), with an argument structure mismatch, it violates the ASC. If it is resolved as in (6b), as involving P-stranding, it violates the CC: The preposition which case-marks the sluiced DP is not included in the antecedent.

Since German has no P-stranding (Merchant, 2001), the only possible interpretation of the DP sluice is that of a subject (7), like in the English (6a).[6]

(6) a. John was texting with somebody, but I don't know who ⟨was texting with John⟩.
b. John was texting with somebody, but I don't know who ⟨John was texting with⟩.
c. John was texting, but I don't know who.

(7) *Hans hat gechattet, aber ich weiß nicht, wer ⟨mit Hans gechattet hat⟩.*
Hans has texted but I know not who with Hans texted has

We investigated sluicing by crossing the three variables CONSTRUCTION (sluicing (SL)/sprouting (SP)), SLUICE (PP/DP), and MATCH (Match (MA)/Mismatch (MM)) between the argument structure of the antecedent and the target (8). We also tested the corresponding full forms in order to tease apart effects of ellipsis from more general preferences. Otherwise, degraded ratings for the mismatch conditions could occur because structural mismatches are marked in general and hence not traced back to ellipsis.[7]

[6] In German, the nominative case morphology on the sluiced *wh*-phrase marks it unambiguously as the subject, whereas the English *who* can in principle also be analyzed as a direct or oblique object.

[7] In principle, it would be desirable to test a full 2×2×2 design, but we did not test the two conditions in (i), because the ellipsis would not be reconstructed in the same way as in (7). In (ia) the *wh*-phrase and the indefinite *jemand* can be interpreted as not being coreferent, unlike in (7). (ib) is not necessarily analyzed as sprouting if no implicit argument is assumed, which is specifically likely with verbs like *to study*, which do not imply a second participant.

(i) a. *Jemand hat gechattet, aber ich weiß nicht, mit wem.* (SP, PP, MM)
somebody has texted but I know not with whom

(8) a. *Hans hat mit jemandem gechattet, aber ich weiß nicht, mit wem.* (SL, PP, MA)
Hans has with somebody texted but I know not with whom

b. *Hans hat mit jemandem gechattet, aber ich weiß nicht, wer.* (SL, PP, MM)
Hans has with somebody texted but I know not who

c. *Jemand hat mit Hans gechattet, aber ich weiß nicht, wer.* (SL, DP, MA)
somebody has with Hans texted but I know not who

d. *Jemand hat mit Hans gechattet, aber ich weiß nicht, mit wem.* (SL, DP, MM)
somebody has with Hans texted but I know not with whom

e. *Hans hat gechattet, aber ich weiß nicht, mit wem.* (SP, PP, MA)
Hans has texted but I know not with whom

f. *Hans hat gechattet, aber ich weiß nicht, wer.* (SP, DP, MM)
Hans has texted but I know not who

We investigated predictability effects based on the likelihood of a second participant by assessing this probability with a pre-test. In all experiments, verbs for which a second participant is likely and verbs for which it is not were balanced across conditions.

3.3 Predictions of our Account and Chung's Constraints

As we anticipated in Section 2, the predictions of the constraints in Chung (2006, 2013) and our processing account partially overlap and partially differ, as Table 1 illustrates.

Table 1. Predictions of Chung (2006), Chung (2013) and the processing account

Prediction	Chung (2006)	Chung (2013)	Processing
All mismatches are degraded as compared to matches	✘	✓	✓
Sprouting mismatches are specifically degraded	✓	✘	✓
Argument structure mismatches improve for adjunct sluices	✘	✓	✘
Sluices improve when a second participant is likely	–	–	✓
DP sluices are overall degraded as compared to PP sluices	–	–	✓

The *Numeration Condition* (NC) in Chung (2006) predicts that all match conditions (8a), (8c) and (8e), as well as the sluicing mismatch conditions (8b) and (8d) are grammatical, since the omitted words *mit* (only in the case of (8b,c)), *Hans*, *gechattet*, *hat* are contained in the numeration of the antecedent. The sprouting mismatch condition (8f) however should be ungrammatical, because the preposition is not included in the antecedent.

According to Chung (2013), all match conditions, i.e. (8a), (8c), and (8e) should be well-formed, because the parallel argument structure satisfies the *Argument Structure Condition* (ASC) and the DP sluice in (8c) can undergo case checking with a verbal head identical to that in the antecedent, as the *Case Condition* (CC) requires. The sluicing and sprouting mismatches with DP sluices (7b,f) satisfy the CC, because the omitted main verb is identical to that in the antecedent, but violates the ASC, because the subject of the antecedent is the object of the sluice. Finally, for mismatches with PP sluices (8d) Chung (2013) predicts an argument-adjunct asymmetry because the ASC concerns only argument sluices. If the PP is a prepositional object of verbs like *to text* or *to chat*, the ASC must be satisfied, so the mismatch would be rendered

b. *Jemand hat gechattet, aber ich weiß nicht, wer.* (SP, DP, MA)
somebody has texted but I know not who

ungrammatical because the argument structure is not parallel. However, if the PP is an adjunct, like with verbs such as *to study* or *to program*, the ASC may be ignored and the resulting structure should be grammatical.

The predictions of our processing account are conditioned on the likelihood of (i) mismatches and (ii) second participants given the verb. If mismatches are overall more unlikely, we expect all mismatch conditions, i.e. (8b), (8d), and (8f) to be degraded. If the explicit antecedent in the sluicing mismatch conditions increases the likelihood of a corresponding sluice as compared to sprouting, we expect that the sprouting conditions (8e) and (8f) are overall degraded as compared to the corresponding sluicing conditions (8a) and (8d). The processing account predicts no categorial differences between adjunct and argument sluices, but an effect of the likelihood of a second participant: The acceptability of sluices, including mismatches, increases as a function of the likelihood of a second participant given the verb. This is obviously the case when the participant is implied by the argument structure of the verb, but also if it is likely given world knowledge. For instance, if it is likely that people do not dance, bowl, or barbecue alone, this will also affect the likelihood of the corresponding sluices. Finally, the processing account predicts a preference for mismatches with PP sluices. PP sluices distribute the processing effort caused by the unexpected mismatch across two words, hence the effort for processing each individual word in the sluice is lower than in DP sluices.

The most immediate support for the processing account consists in evidence that degraded ratings are related to the low probability of a continuation, which we investigate with the production task (see Sections 5 and 6), and higher processing effort, which is evidenced by reading times and which can also reduce acceptability (see Section 7).

4 Pre-Test: How Likely is a Second Participant?

We assessed the likelihood of a second participant given a specific verb with a pre-test. We presented 61 simple statements like (9) to 33 undergraduate students of German Studies at Saarland University, who rated the likelihood of a second participant to be involved in the action described on a 5-point Likert scale (5 = very likely). All statements were presented in present perfect and contained only a subject and a two-word predicate.

(9) *Hans hat gechattet.*
Hans has texted

Based on the mean likelihood scores for each item, we selected 24 verbs for our experiments: a group of 12 verbs from the upper end of the likelihood scale ($mean = 4.4, sd = 0.28$) and a group of 12 verbs from the lower end ($mean = 1.83, sd = 0.26$).

5 Acceptability Rating Study

5.1 Motivation

The acceptability rating study had the goal to test the predictions of Chung (2006), Chung (2013) and our processing account that we summarized in Table 1.[8] Following Chung (2006), sprouting mismatches (which have DP sluices in our materials) should be particularly degraded as compared to sluicing mismatches, since the omitted preposition is not present in the antecedent's numeration. In contrast, Chung (2013) predicts a dispreference for mismatches with argument sluices, i.e. all DP sluices and the argument PP sluices, as compared to adjunct PP sluices.

[8] Runner & Dozat (2011) also collected acceptability judgments for mismatching sluices and utterances that violate the NC of Chung (2006). Their study however was conducted in English and contained voice mismatches instead of argument structure mismatches. They found that both utterances with voice mismatches and utterances that violated the NC were comparably degraded and that these effects were additive.

Since the ASC is restricted to arguments, Chung (2013) predicts mismatches with adjunct PP sluices (which also conform to the CC) to be fine. Our processing account makes the general prediction that a sentence with sluicing is more acceptable when the sluice is relatively likely. This concerns both matching and mismatching sluices and is the reason for why we expect an additional effect of the likelihood of a second participant. Syntactic accounts of course do not predict that all presumably grammatical encodings are equally acceptable, since processing difficulties can result in reduced acceptability. However, if they were correct, we would expect that conditions predicted to be ungrammatical are degraded as compared to the corresponding grammatical conditions, no matter how they are rated on the scale in absolute terms and whether there are additional effects that do not result from grammatical constraints.

5.2 Materials

In the rating study, we tested not only the six conditions in (8), but also the corresponding nonelliptical full forms like e.g. (10), what results in a total of 12 conditions:

(10) *Hans hat gechattet, aber ich weiß nicht, mit wem Hans gechattet hat.*
Hans has texted but I know not with whom Hans chatted has

This additional binary predictor ELLIPSIS allows us to tease apart effects of argument structure mismatches, which might seem to be overall pragmatically odd, and effects which are specific to ellipsis. ELLIPSIS was varied between subjects in order to avoid a possible floor effect on mismatches under ellipsis.

5.3 Procedure

We conducted our experiment on the Internet using the IBEX survey presentation software (Drummond, 2019). We recruited 96 self-reported native speakers of German on the crowdsourcing platform Clickworker, 48 for each of the two ELLIPSIS groups. All participants received €2.25 for their participation. Subjects were asked to rate the naturalness of the materials on a 7-point Likert scale (7 = fully natural). Materials were first divided into the two ELLIPSIS groups and then distributed across 6 lists respectively with a Latin square design so that each subject saw each item once and only in one condition. Each subject rated 24 items that were mixed with 60 fillers, which included grammatical distractor sentences with causal complement clauses, instances of gapping and their nonelliptical counterparts and garden path sentences which might be hard to process. The materials were presented in pseudo-randomized order assuring that no two sluicing/sprouting stimuli immediately followed each other.

5.4 Results

To test the predictions of Chung's two accounts and our processing account (Table 1) we performed three analyses using Cumulative Link Mixed Models for ordinal data (CLMMs) (Christensen, 2019) in `R`. In all analyses reported this paper we use a backward model selection procedure: Starting from the full (most complex) model, we subsequently exclude those predictors that do not significantly improve model fit, as evidenced by likelihood ratio tests calculated with the `anova` function in `R` (R Core Team, 2019). We also use likelihood ratio tests to obtain p-values by comparing the model fit of a model containing the specific predictor to that of a model without this predictor.

5.4.1 Are Mismatches Overall Degraded, and How do the Likelihood of the Verb and the Form of the Sluice Impact Acceptability?

In our first analysis we investigated whether mismatches are overall rated worse than the corresponding matches, whether the ratings for sluices improve when a second participant is likely given the verb and whether DP sluices are generally degraded as compared to PP sluices. The full model contained the ratings as dependent variable (DV), fixed effects of the sum-coded binary

predictors ELLIPSIS, MATCH, CONSTRUCTION and SLUICE, the z-transformed pre-test score of the VERB and the scaled POSITION of the trial in the experiment.[9] We included all two-way interactions between the predictors except the ELLIPSIS:POSITION interaction, which would not be meaningful because ELLIPSIS was tested between-subjects. We also considered all three-way interactions with ELLIPSIS (except those including POSITION), since they allow us to distinguish general preferences from ellipsis-specific ones. Our model contained random intercepts for items and subjects, by-item random slopes for ELLIPSIS, MATCH, CONSTRUCTION and SLUICE and by-subject random slopes for MATCH, CONSTRUCTION SLUICE and VERB.

Table 2. Fixed effects in the final CLMM for the first analysis of the rating study

Predictor	**Estimate**	**SE**	χ^2	**p-value**	
ELLIPSIS	1.81	0.45	14.91	< 0.001	***
MATCH	2.5	0.31	51.87	< 0.001	***
CONSTRUCTION	0.76	0.13	29.61	< 0.001	***
SLUICE	0.5	0.15	10.02	< 0.001	***
VERB	0.35	0.09	13.9	< 0.001	***
POSITION	–0.11	0.04	6.24	< 0.05	*
ELLIPSIS:MATCH	–4.34	0.6	42.42	< 0.001	***
ELLIPSIS:CONSTRUCTION	–0.87	0.25	11.66	< 0.001	***
ELLIPSIS:SLUICE	–0.95	0.26	11.99	< 0.001	***
ELLIPSIS:VERB	–0.08	0.11	0.54	> 0.4	
MATCH:VERB	0.36	0.11	9.02	< 0.01	**
MATCH:POSITION	0.26	0.09	8.47	< 0.01	**
CONSTRUCTION:VERB	–0.27	0.1	7.66	< 0.01	**
ELLIPSIS:MATCH:VERB	–0.4	0.18	4.95	< 0.05	*

The final model is summarized in Table 2. The main effect of ELLIPSIS ($\chi^2 = 14.91, p < 0.001$) shows that overall participants prefer syntactically complete utterances.[10] The main effect of MATCH ($\chi^2 = 51.87, p < 0.001$) indicates that matches are rated as better than mismatches. There is an interaction between both predictors ($\chi^2 = 42.42, p < 0.001$) that suggests that mismatches are particularly degraded under ellipsis (see Figure 1). According to the main effect of CONSTRUCTION ($\chi^2 = 29.61, p < 0.001$) and the interaction between ELLIPSIS and CONSTRUCTION ($\chi^2 = 11.66, p < 0.001$), constructions with an explicit antecedent are preferred over constructions with an implicit antecedent both generally and in particular for ellipsis. The main effect of SLUICE ($\chi^2 = 10.02, p < 0.001$) and the interaction between ELLIPSIS and SLUICE ($\chi^2 = 11.99, p < 0.001$) show both an ellipsis-independent and an ellipsis-specific effect, too: Utterances with a PP continuation are preferred over ones with a

[9] We also included a predictor for the relative FREQUENCY of the continuation in the production task (see Section 6). It indicates the ratio of produced sluicing or sluiceable continuations that refer to the antecedent per token set and condition and could indicate how likely an antecedent will be picked up again in a sluice. As it was neither significant nor part of a significant interaction of CLOZE FREQUENCY in this analysis, we do not discuss it in more detail here.

[10] This effect might seem to conflict with our prediction that speakers prefer ellipsis when the potentially sluiced phrase is highly predictable. However, and despite the main effect of ELLIPSIS, our data do actually show a preference for ellipsis under specific circumstances, which are reflected in the interactions in our model and which are illustrated by the figures in this section.

DP contiuation, but this preference is particularly strong for elliptical utterances (see Figure 2). The main effect of VERB ($\chi^2 = 13.9, p < 0.001$) indicates that subjects rated utterances as more acceptable the more likely a second participant is given the verb. The VERB effect interacts significantly both with MATCH ($\chi^2 = 9.02, p < 0.01$) and with CONSTRUCTION ($\chi^2 = 7.66, p < 0.01$): Matching utterances and utterances without an explicit antecedent are generally more acceptable when a second participant is more likely. There is a three-way interaction between ELLIPSIS, MATCH and the VERB ($\chi^2 = 4.95, p < 0.05$): The ratings for matching ellipses improve when a second participant is likely (see Figure 3). The theoretically uninteresting main effect of POSITION ($\chi^2 = 6.24, p < 0.05$) and the interaction between MATCH and POSITION ($\chi^2 = 8.47, p < 0.01$) show that the ratings became overall worse in the course of the experiment and in particular for mismatches.

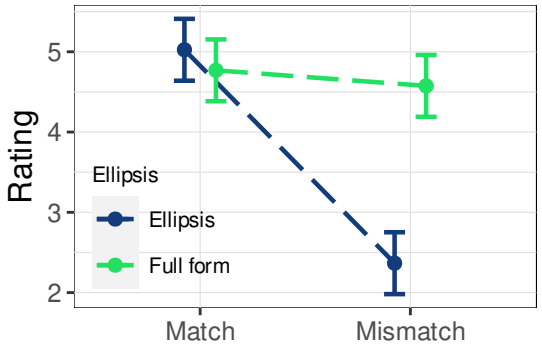

Figure 1. Aggregated mean ratings and 95 % confidence intervals for matches vs. mismatches as a function of ELLIPSIS in the complete data set

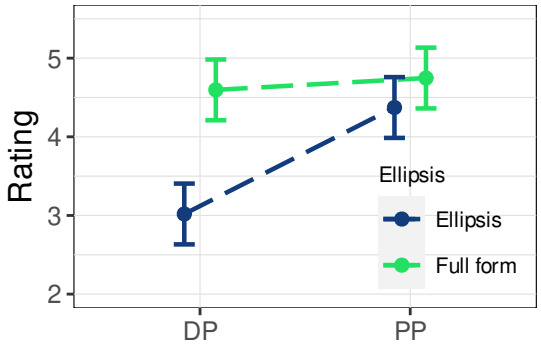

Figure 2. Aggregated mean ratings and 95 % confidence intervals for DP vs. PP continuations as a function of ELLIPSIS in the complete data set

5.4.2 Are Sprouting Mismatches Particularly Degraded?

As we discussed above, our data set is not fully balanced since it contains only sprouting mismatches with a DP sluice and sprouting matches with a PP sluice. Therefore, we performed a second analysis on only a subset of the data to investigate whether sprouting mismatches are particularly degraded, as is predicted by the NC (Chung, 2006) and our processing account. This subset contained the sprouting matches and mismatches and only the corresponding sluicing data, i.e. sluicing DP mismatches and sluicing PP matches, both as full forms and ellipses.

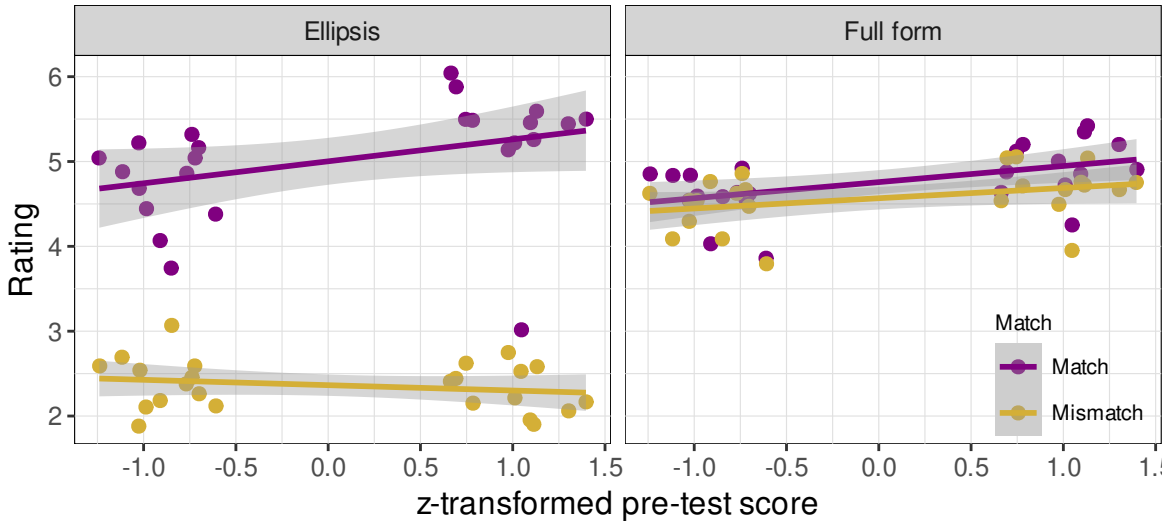

Figure 3. Mean ratings as a function of the z-transformed pre-test scores grouped by MATCH and ELLIPSIS in the complete data set

The full model included the sum-coded predictors ELLIPSIS, CONSTRUCTION and MATCH[11], as well as the pre-test score for VERB and all two-way and three-way interactions between these predictors. The random effects structure consisted of random intercepts for items and subjects, of by-item random slopes for CONSTRUCTION, MATCH and VERB and of by-subject slopes for ELLIPSIS, CONSTRUCTION and MATCH.

The final model contains a three-way interaction between ELLIPSIS, CONSTRUCTION and MATCH ($\chi^2 = 4.55, p < 0.05$) that indicates that elliptical sprouting mismatches are particularly degraded as compared to the corresponding full form (see Figure 4). This is in line with the processing account and with Chung (2006). Furthermore, there were main effects of ELLIPSIS ($\chi^2 = 17.07, p < 0.001$), MATCH ($\chi^2 = 61.05, p < 0.001$), CONSTRUCTION ($\chi^2 = 25.88, p < 0.001$), and VERB ($\chi^2 = 7.95, p < 0.01$) which replicate the results of the full analysis just like the interactions between ELLIPSIS and MATCH ($\chi^2 = 58.81, p < 0.001$), between ELLIPSIS and CONSTRUCTION ($\chi^2 = 10.23, p < 0.01$) and between CONSTRUCTION and VERB ($\chi^2 = 4.92, p < 0.05$).

5.4.3 Do Argument Structure Mismatches Improve for Adjunct Sluices?

In a third analysis we tested whether there is an asymmetry between argument and adjunct sluices, as is predicted by Chung (2013). We created a subset of the data that contains all sluicing matches and mismatches with a PP sluice. We restrict ourselves to PP sluices because only they can either be an argument or an adjunct, whereas a DP sluice appears in nominative case and hence is always an argument (the subject of the continuation). For each verb, we annotated whether it requires an argument or an adjunct as ATYPE by testing whether there is necessarily a second participant involved in the action described by the verb. In 5 of our materials the PP sluice was classified as a prepositional object (i.e. an argument) and in the remaining 19 as an adjunct.[12]

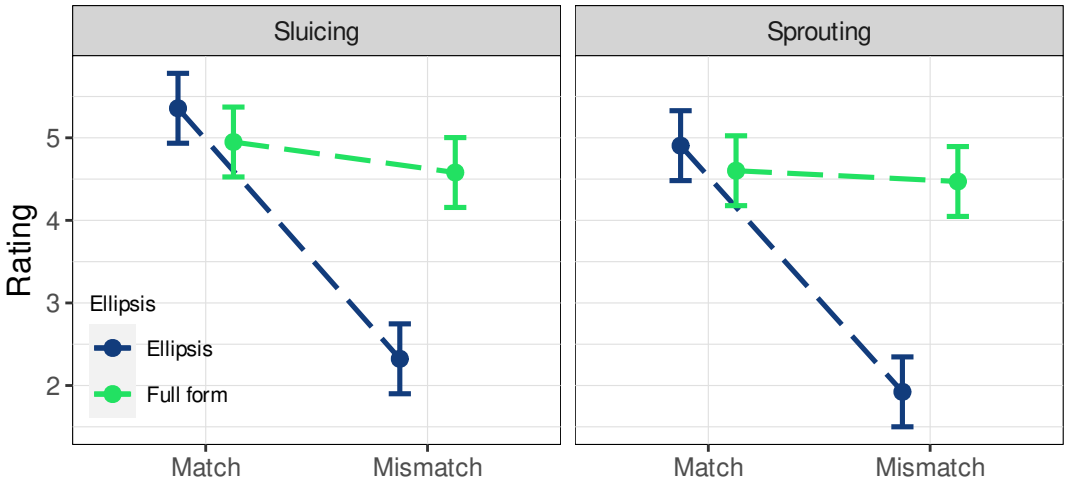

Figure 4. Mean ratings and 95 % confidence intervals for sluicing and sprouting PP matches and DP mismatches as a function of MATCH and CONSTRUCTION in the subset of analysis 2

The full model contained the sum-coded binary predictors ELLIPSIS, MATCH and ATYPE and the pre-test score for the VERB as well as all two-way and three-way interactions between these predictors. We included random intercepts for items and subjects as well as by-item random

[11] Since MATCH and SLUICE correlate perfectly in this data set, only one predictor can be included in the analysis.
[12] The argument verbs were *chatten* ("to text"), *kuscheln* ("to snuggle"), *knutschen* ("to smooch"), *telefonieren* ("to be on the phone with") and *quatschen* ("to chat"). Of course it would have been desirable to balance the verbs between requiring an argument and an adjunct but we had chosen them with respect to the likelihood of a second participant since this was necessary in order to test a central prediction of our processing account.

slopes for ELLIPSIS, MATCH and ATYPE and by-subject random slopes for MATCH, ATYPE and VERB. In the final model there is neither a significant three-way interaction between ATYPE, MATCH and ELLIPSIS ($\chi^2 = 0.04, p > 0.8$) nor any other significant effect which ATYPE is part of. Unlike the ASC by Chung (2013) predicts, we find no evidence for an argument-adjunct asymmetry. Furthermore, the model contained effects for ELLIPSIS ($\chi^2 = 4.26, p < 0.05$), MATCH ($\chi^2 = 50.62, p < 0.001$), their interaction ($\chi^2 = 37.02, p < 0.001$), the interaction between ELLIPSIS and VERB ($\chi^2 = 2.73, p < 0.1$), the interaction between MATCH and VERB ($\chi^2 = 5.9, p < 0.05$), and the three-way interaction ELLIPSIS:MATCH:VERB ($\chi^2 = 7.03, p < 0.01$).

5.5 Discussion

The acceptability rating study tested the predictions of our processing account and the different identity constraints proposed by Chung (2006, 2013). The results of our analyses are summarized in Table 3: Our results support the predictions by Chung (2006, 2013) to a certain extent. However, Chung's constraints cannot account for the complete pattern. Analysis 1 showed that mismatches are in general degraded as compared to matches, that the ratings for sluices improve when a second participant is likely given the verb and that DP sluices are overall rated as worse than PP sluices. By performing analysis 2 on a subset of the data we found that sprouting mismatches are particularly degraded as compared to sluicing mismatches. Analysis 3 did not reveal an argument-adjunct asymmetry for mismatching PP sluices.

Table 3. Summary of the empirical predictions and the corresponding results of the rating study

Prediction	Chung (2006)	Chung (2013)	Processing	Rating study
Penalty for mismatches	✘	✓	✓	✓ (analysis 1)
Penalty for sprouting mismatches	✓	✘	✓	✓ (analysis 2)
Argument-adjunct asymmetry	✘	✓	✘	✘ (analysis 3)
Verb-driven predictability effect	–	–	✓	✓ (analysis 1)
Preference for PP sluices	–	–	✓	✓ (analysis 1)

The data are not fully in line with the NC (Chung, 2006), since sprouting mismatches are particularly degraded: The omitted preposition is not part of the antecedent's numeration, and this violates the NC. However, our analysis 2 shows that mismatches are degraded across the board. In contrast, the NC predicts mismatching sluices to be grammatical since all omitted words are contained in the numeration of the antecedent. Chung (2013) can account for the fact that mismatches are generally degraded because they violate the ASC in having a non-parallel argument structure. However, the ASC predicts an argument-adjunct asymmetry for PP sluices, because the ASC should only be effective for PP arguments but not for PP adjuncts. The fact that we did not find such an asymmetry in the data challenges Chung's more recent proposal.

From the perspective of our processing account, we explain the preference for matching sluices, and specifically mismatches under sprouting with their reduced likelihood: We expect that mismatches are overall less likely, and that explicit mentions of the second participant in our sluicing conditions increase the likelihood of referring to him/her in the continuation. So far, we have not presented empirical evidence for differences in likelihood, but the production study presented in Section 6 will provide such evidence.

Besides these effects, which are predicted both by syntactic identity accounts and our processing account, we found effects of the likelihood of a second participant given the verb on acceptability. Syntactic accounts cannot capture such predictability effects, even though they might partially overlap with the argument-adjunct asymmetry predicted by the ASC in Chung (2013), which we disconfirmed. In contrast to syntactic accounts, under our account

predictability effects are expected. First, we find that the more likely a second participant is, the more acceptable are sluices referring to this participant as compared to the corresponding full forms. This is in line with our expectation that ellipsis is more strongly preferred, the more likely a continuation is. Second, this effect is particularly strong for both full forms and ellipses without an explicit antecedent. We also expected this finding, since mentioning the second participant explicitly could override expectations with respect to its likelihood based on the verb. The ratings for matching utterances improved when a second participant is likely, but mismatches seem to be degraded to such a degree that they remain unaffected by this likelihood manipulation. This effect is present for both full forms and ellipsis but is specifically strong for the latter.

Finally, our account also explains the preference for utterances with PP sluices over utterances with DP sluices in general and particularly for ellipsis. PP sluices like *mit wem* distribute the processing effort across two words compared to only one word like *wer* for DP sluices and thus reduce the effort on each individual word. The observation that this effect seems to be independent of the form of the antecedent however is somewhat unexpected. In principle we would have expected that reducing processing effort is particularly important for mismatches as they should be more likely to cause excessively high processing effort. It could however be the case that, like we argued above for the predictability manipulation, mismatches are degraded to such an extent that the difference between PP and DP sluices is of no consequence. In Section 6 we provide experimental evidence from a production study that participants hardly ever produce mismatching ellipses. This suggests that mismatches need special circumstances to occur, which are not present in our study.

Taken together, our processing account captures the result of the rating study in a more comprehensive way than the two accounts by Chung. In addition to explaining the observed pattern for mismatches in general and sprouting mismatches it accounts also for effects of the form of the sluice and of the the likelihood of a second participant.

6 Production Study

6.1 Motivation

The processing account relates the acceptability of sluices to their likelihood, hence we measured this likelihood with a production experiment. In this production task, we cut off our materials after *nicht* (11) and asked subjects to provide a natural continuation for the utterance. The first goal of the production task was to estimate the likelihood of (mis)matching continuations. By *continuations* we understand both ellipses and full forms starting with *who* or *with who(m)*. The predictions of our account crucially hinge on this likelihood, since the account predicts that the tendency of reducing a continuation to sluicing is stronger the more likely this continuation is. We test the three assumptions with respect to the likelihood of sluices that we discussed in Section 3: (i) mismatching continuations are overall less likely, (ii) continuations referring to a second participant are more likely when (s)he is explicitly mentioned in the antecedent, and (iii) a continuation referring to the second participant is more likely when a second participant is likely given the verb, because a second participant is more likely to exist. We expect this effect to be particularly strong when the participant is not explicitly mentioned in the antecedent (i.e. when ellipsis yields sprouting), because in this case its existence can only be inferred from the verb. The second goal of the production study was to test our account's central prediction that the likelihood of a specific continuation predicts the likelihood of reducing this continuation by ellipsis. Recall that unlike syntactic identity accounts of sluicing such as Chung (2006, 2013) we expect this to concern matching sluices too, and both arguments and adjuncts.

6.2 Materials

The materials were based on the stimuli tested in the rating study. We presented the sentence containing the potentially sluiced phrase as an independent sentence introduced by *Leider weißich nicht* ... 'Unfortunately, I don't know ...' which was cut off after *nicht*, as is shown in (11). We varied the FORM of the potential antecedent for sluicing in an 1×3 design: The utterance either contains a potential PP antecedent (11a), a potential DP antecedent (11b) or no explicit antecedent at all (11c).

(11) a. *Hans hat mit jemandem gechattet. Leider weiß ich nicht,* _____ (PP)
Hans has with somebody texted. Unfortunately know I not

b. *Jemand hat mit Hans gechattet. Leider weiß ich nicht,* _____ (DP)
somebody has with Hans texted. Unfortunately know I not

c. *Hans hat gechattet. Leider weiß ich nicht,* _____ (sprouting)
Hans has texted. Unfortunately know I not

6.3 Procedure

We recruited 120 participants on the crowd-sourcing platform Clickworker for the survey which was conducted on the Internet using the LimeSurvey questionnaire software. Subjects were paid €1.80 for participating in the study. They were asked to complete the stimuli with the continuation that they considered to be most likely. The materials were first divided into two groups with 12 items each that were balanced for the likelihood of a second participant, i.e. there were 6 verbs with a high pretest score and 6 verbs with a low one respectively. Each group of materials was then distributed across three lists with a 1×3 Latin square design: Each subject saw each token set in a group once and only in one condition. Each subject provided continuations for 12 items (4 with a PP antecedent, 4 with a DP antecedent and 4 without an antecedent), which were mixed with 24 fillers and presented in pseudo-randomized order. The fillers resembled the items in consisting of two sentences each. In the fillers, 50 % of the target sentences were introduced by *leider* and 50% by *allerdings* 'however'.

6.4 Annotation

We first excluded responses that were not meaningful continuations, like random key presses. This resulted in the loss of 0.99 % of the responses. For the remaining data we annotated (i) whether the continuation produced was a polar or a *wh*-question, in case of the latter, (ii) whether the *wh*-phrase referred to the antecedent, and, if so, (iii) whether it matched its form. For *wh*-questions we also annotated whether subjects produced a full form or a sluice. Overall, 83.64 % of the remaining data contained *wh*-questions, and in 51.52 % of these *wh*-questions the *wh*-phrase referred to the potential antecedent in the stimulus. Subjects produced sluices in 43.61 % of the *wh*-questions, otherwise full forms.

6.5 Results

Figure 5 provides an overview of the annotated responses as a function of the antecedent. The plot shows that subjects produced argument structure mismatches. However, Figure 6 shows that mismatches occurred almost only in the full forms, except for a small amount of mismatches with DP antecedents and PP sluices ($n = 18$). From a theoretical perspective, it is an interesting observation that subjects produced a relatively high ratio of (nonelliptical) *wer das war* 'who that was' continuations, in particular in the DP antecedent condition. Merchant (2004) suggests that deictic pronouns and copulas are salient enough to license ellipsis even in discourse-initial contexts, but since subjects produced no DP sluices in context of PP antecedents our data suggest that the *wer das war* continuations do not license ellipsis in our experiment.

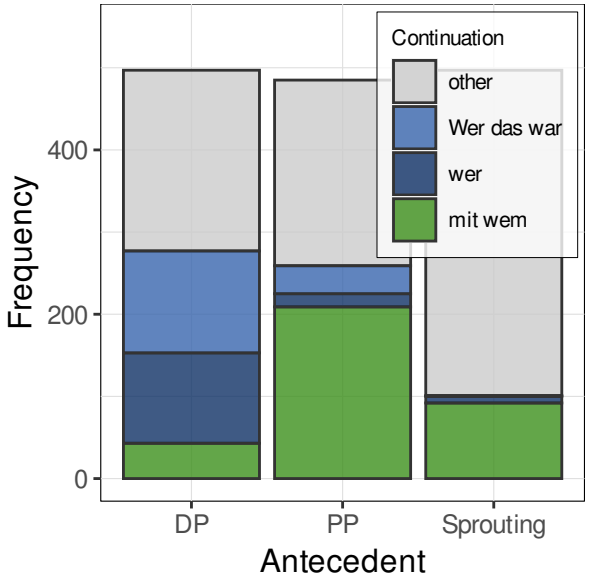

Figure 5. Ratio of continuations as a function of the antecedent

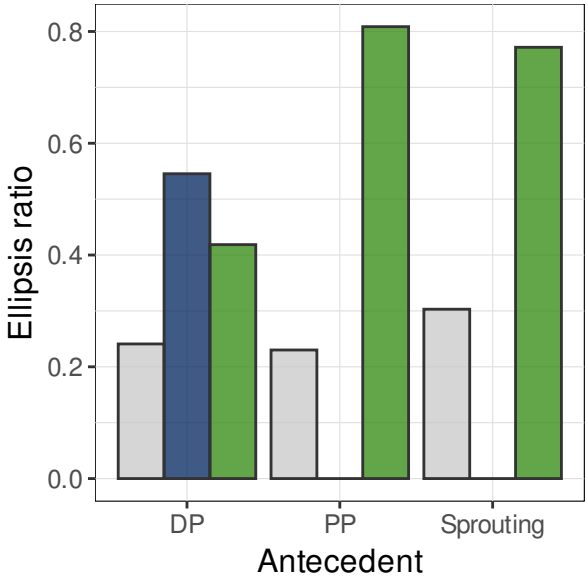

Figure 6. Ratio of ellipsis as a function of antecedent and target

We analyzed the data with logistic regressions in R (R Core Team, 2019). We first investigated whether the FORM of the antecedent (PP, DP, implicit), and the likelihood of the (potentially implicit) second participant given the VERB determine (i) the likelihood of continuations referring to the second participant, and (ii) whether they condition the likelihood of mismatches. Both of these analyses do not distinguish between actual sluices, i.e. ellipses, and the corresponding full forms. In a third analysis, we tested whether more likely continuations of the stimuli are more likely to be reduced by ellipsis, as our account predicts.

6.5.1 How Likely are Continuations Referring to the Second Participant?

Our first series of analyses investigated how likely continuations referring to a second participant are with mixed effects logistic regressions (Bates et al., 2015). For this purpose, we predicted the outcome of a binary DV that encoded whether the potentially sluiced continuation began with a related *wh*-phrase (*who*/*with whom*) or not (e.g. *why*, *when*, *where* or polar embedded questions like *whether* . . .). Since FORM is a ternary variable, in a first step we compared the two (potential) sluicing conditions with explicit antecedents (11a,b) to each other and then compared implicit to explicit antecedents in a second step. Figure 7 illustrates this relationship and suggests that the effect of the verb is particularly strong in the (potential) sprouting condition, i.e. when the antecedent is left implicit.

The full model of the explicit antecedents analysis contained main effects of the FORM of the antecedent (DP/PP), the z-transformed pre-test score for the main VERB, the POSITION of the trial in the experiment and all two-way interactions between these predictors, as well as by-subject and by-item random intercepts. All nominal predictors in the models described in this section were sum-coded. Like in the case of the rating data analyzed in Section 5.4, we used likelihood ratio tests to assess whether including a term in our model significantly improved model fit and removed those that did not. The analysis showed that none of the predictors had a significant effect on the likelihood of a related continuation, including FORM ($\chi^2 = 0.08, p > 0.7$).

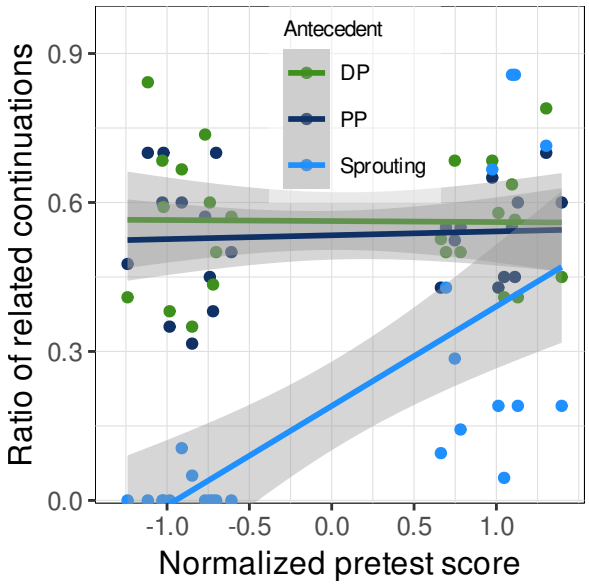

Figure 7. Ratio of continuations related to the second participant by antecedent and pretest score

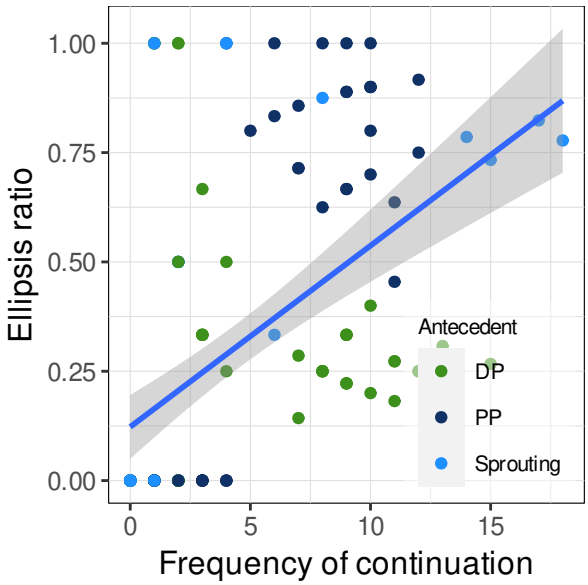

Figure 8. Ellipsis ratio as a function of the frequency of each continuation (by token set and antecedent)

Therefore, in a second step we pooled the conditions with explicit antecedents (11a,b) and compared them to the implicit antecedent condition (11c), where the second participant is not mentioned in the antecedent. The full model contained main effects of FORM (antecedent explicit/implicit), VERB and POSITION, all two-way interactions and by-subject and by-item random intercepts and a by-item random slope for CONSTRUCTION. In the final model there were significant effects of VERB, FORM and of their interaction. The main effect of VERB ($\chi^2 = 19.73, p < 0.01$) shows that related continuations are more likely when a second participant is likely given the verb, and the effect of FORM ($\chi^2 = 38.35, p < 0.001$) suggests that this is also the case when the second participant is explicitly mentioned in the antecedent. The VERB:CONSTRUCTION interaction shows that the effect of the verb is particularly strong when the second participant is implicit ($\chi^2 = 27.43, p < 0.001$).

Our results support our expectations about the likelihood of continuations referring to a second participant. Such continuations are more likely when the second participant is explicitly mentioned and when the second participant is likely given the verb. Additionally, the interaction confirms our expectation that the verb bias is overridden by explicitly mentioning the participant.

6.5.2 Does the Likelihood of a Continuation Predict its Reduction?

Independently of the ratio of mismatches, our account predicts that continuations that are more likely are more often reduced by ellipsis. We investigated this with a logistic regression (R Core Team, 2019) that predicted the likelihood of a continuation's reduction by sluicing from its overall likelihood. This analysis was performed on a subset of the data that contained only continuations referring to the second participant, i.e. those with *wer* or *mit wem* ('(with) who(m)'). The IV was the frequency of a continuation in the production data, grouped by MATCH, FORM and TOKENSET. For instance, we expect that if a matching related continuation is very likely given the verb in a specific token set, it will be more likely to be reduced. The analysis shows that this prediction is borne out, since there is a significant effect of PREDICTABILITY on omission ($F = 50.68, p < 0.001$). Figure 8 illustrates this relationship.

6.5.3 Discussion

The results of the production study support our expectations about the likelihood of sluices and sluiceable continuations referring to a potentially implicit second participant. Furthermore, as our processing account predicts, more likely continuations are more often reduced. Both

explicit antecedents and a higher probability of a second participant given the verb increase the likelihood of a corresponding sluice, and the latter does so specifically when the second participant is not mentioned in the antecedent. Second, our study suggests that predictable continuations are more likely to be reduced by ellipsis.

In our experiment, only few elliptical argument structure mismatches were produced. However, their distribution is in line with the assumptions and predictions of our processing account: If mismatches are less likely, and our data indicate that they are, they will also be perceived as more strongly degraded, as our rating study showed. For instance, mismatches with DP antecedents and PP sluices were perceived as more acceptable than DP sluices with PP antecedents, and they were also the only elliptical mismatches produced at all.

7 Self-Paced Reading Study

7.1 Motivation

The rating study and the production study support the main prediction of our processing account that continuations that are more likely are also more likely to be reduced to sluicing and that their reduction is perceived as more acceptable. However, in the production study subjects produced only few acceptable mismatches, and the differences in likelihood that we found for full forms do not contradict the assumption of syntactic identity constraints on ellipsis that Chung (2006, 2013) proposes. Similarly, the rating study alone does not show whether some conditions are degraded because they are ungrammatical, as syntactic identity constraints predict, or whether they are grammatical but very hard to process. We address this question with a self-paced reading paradigm that investigates whether degraded continuations, which are unexpected given our production data, are also harder to process. If this prediction was borne out, the data would be explained by our processing account without having to assume syntactic identity conditions which are specific to sluicing.

7.2 Materials

We tested the full forms of our 24 items in the six conditions in (8) above. Testing the full forms allows us to use the TP which is omitted under sluicing as spill-over region which is relatively homogeneous across materials, since the material that is contained in the TP is given in the antecedent. We measured reading times on the first four words of this TP, i.e. the *wh*-phrase, the preposition, the subject and the participle, which are bold-faced in the two sample conditions in (12). This way we present full forms, but measure the processing effort on the *wh*-phrase, i.e. the effort associated with a continuation that up to this point could result in ellipsis. In what follows we refer to these structures as sprouting or sluicing *continuations* depending on the form of the antecedent.

(12) a. *Hans hat mit jemandem gechattet, aber ich weiß nicht,* ***mit wem Hans gechattet*** *hat.*
Hans has with somebody texted but I know not with whom Hans texted has (SL, PP, MA)

b. *Hans hat mit jemandem gechattet, aber ich weiß nicht,* ***wer mit Hans gechattet*** *hat.*
Hans has with somebody texted but I know not who with Hans texted has (SL, PP, MM)

7.3 Procedure

The study was conducted on the Internet using IBEX (Drummond, 2019). 48 participants were recruited on the Clickworker crowd-sourcing platform and rewarded with €2.50 for participation.

They read the materials word-by-word in a moving-window masked self-paced reading task. Each subject saw 24 items (4 per condition), which were mixed with the same 60 fillers that we tested in the rating study and presented in individual pseudo-randomized order. 12 fillers were followed by a comprehension question which served as an attention check. 6 of the questions had to be correctly answered as "true" and 6 as "false". Two subjects who provided incorrect answers to more than 25 % of the comprehension questions were excluded from further analysis.

We excluded all by-word reading times that were faster than 90 msec and slower than 2500 msec, which resulted in a loss of 2.1 % of the total data. We then residualized by-word log reading times following the method proposed by Jaeger (2008): We trained a linear mixed effects model (Bates et al., 2015) with fixed effects for the word length in characters, the position of the trial in the experiment, the position of the critical words in the utterance, the type of the trial (item or filler) and a by-subject intercept and used the residuals of this model in the analyses described in the results section in order to factor out these effects. After residualization, we summed the by-word residual log reading times for the critical region in order to obtain one data point per trial. If we did not have reading time data for each of the words in the region in a specific trial (for instance because a subject pressed the space bar too fast for a specific word) we excluded the trial from further analysis. This resulted in a loss of 0.3 % of the data for DP sluices and none of the data for PP sluices. Finally, in order to minimize the effect of outliers, we cut off all aggregated reading times higher or lower than the mean reading time $\pm$ 2.5 standard deviations at that value.

7.4 Results

Figure 9 shows the mean residual log reading times by condition. We analyzed the data using linear mixed effects models (Bates et al., 2015) in R that predicted the aggregated residual log reading time from the FORM of the antecedent, the CONSTRUCTION (sluicing or sprouting), the pre-test score of the VERB and the log-transformed POSITION of the trial in the experiment. Since the antecedent was a ternary DV, and reading times for DP and PP continuations might differ from each other, we conducted a series of pairwise comparisons. First, we investigated whether mismatching DP and PP sluicing continuations were read more slowly and then we tested our predictions that the sprouting continuations are overall read more slowly than sluicing continuations and that the likelihood of a second participant given the verb, operationalized by the pretest score, would specifically reduce reading times for sprouting continuations.

7.4.1 Are Mismatches under Sluicing Read More Slowly than Matches?

Our first two analyses were conducted on subsets of the data containing only the sluicing continuations, one with all DP sluicing continuation conditions and another with all PP sluicing continuation conditions. Both for DP and PP continuations, the full model contained fixed effects for the FORM of the antecedent, the POSITION of the trial in the experiment, the pretest score of the main VERB and all two-way interactions. Models had by-item random slopes for FORM and by-subject and by-item random intercepts. In the case of PP continuations, a main effect of FORM shows that mismatches were read significantly more slowly than matches ($\chi^2 = 5.39$, $p < 0.05$), furthermore there was a FORM:POSITION interaction indicating that reading times for mismatches decreased in the course of the experiment as participants got used to them ($\chi^2 = 5.0$, $p < 0.05$). For DP continuations, mismatches were read marginally more slowly ($\chi^2 = 3.59$, $p = 0.06$) and there was a POSITION main effect indicating an context-independent decrease of reading time across the study ($\chi^2 = 19.12$, $p < 0.001$). Taken together, we find that, as our processing account predicts, mismatching continuations are read more slowly than matching ones.

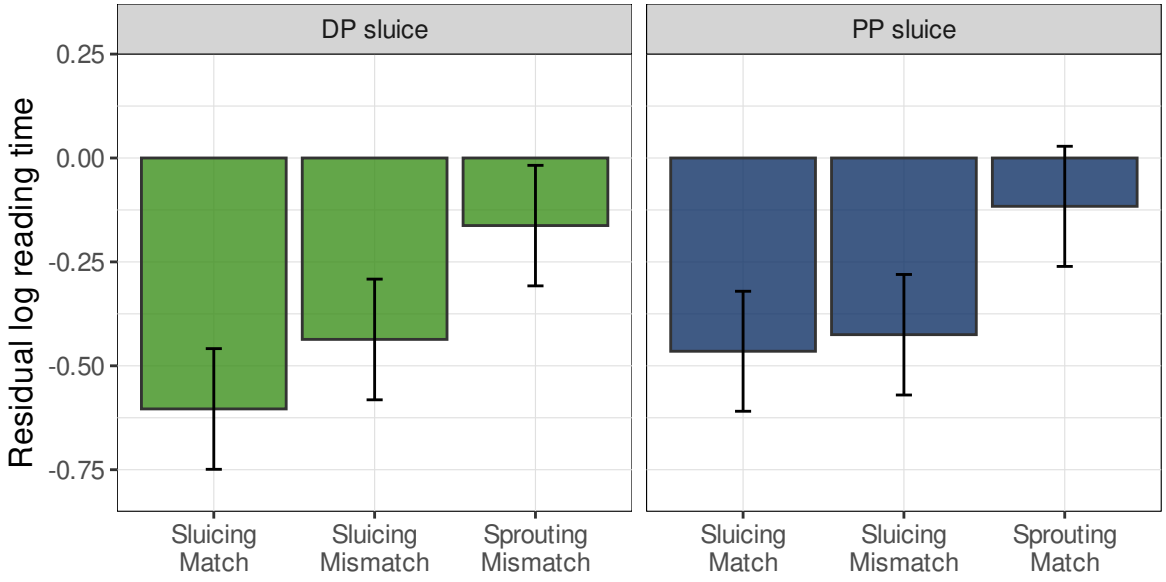

Figure 9. Mean residual log reading times and 95 % confidence intervals by condition. Note that the PP sluicing mismatch condition appears to be as fast as the match condition, but it is subject to a significant interaction with the POSITION of the trial in the experiment

7.4.2 Are Continuations Referring to Implicit Antecedents Read More Slowly?

Our next two series of analyses investigated whether there are differences in reading time between antecedents with explicit and those with implicit antecedents for the *wh*-phrase. We created one subset containing the sprouting mismatch continuation condition and the corresponding DP sluicing mismatch continuation condition, and a second one containing the sprouting match continuation condition and the corresponding PP sluicing match continuation condition. In these analyses we used the same predictors as in the sluicing continuations analyses, except for the fact that the FORM did not distinguish between PP and DP antecedents, but between explicit and implicit ones. The DP model had by-item and by-subects random intercepts but no random slopes, whereas the PP model additionally had an by-items random slope for FORM. The analysis of PP continuations shows that matching continuations are read more slowly when the antecedent is implicit than when the antecedent is explicit ($\chi^2 = 14.6$, $p < 0.001$). Furthermore there is a marginal POSITION effect that reveals a condition-independent decrease in reading time throughout the study ($\chi^2 = 3.01$, $p = 0.08$). The analysis for DP continuations shows that those with implicit antecedents are read more slowly than those with explicit ones ($\chi^2 = 17.49$, $p < 0.001$). In this model there was an effect of POSITION ($\chi^2 = 10.5$, $p < 0.01$) too.

7.5 Discussion

The results of our self-paced reading study provide clear evidence that mismatches are harder to process than matches and that sprouting continuations are harder to process than sluicing continuations. Both PP and DP continuations are read more slowly when they do not match the form of the antecedent or when they refer to an implicit antecedent. The slower reading times for mismatches are in line with the acceptability rating data, which show that mismatches are degraded, and the production data, where mismatches were less frequent. Since unpredictable expressions require more processing effort, this suggests that mismatches are degraded due to their low probability. Similarly, the observation that continuations referring to implicit antecedents (which result in sprouting in case of ellipsis) are read more slowly than those referring to explicit ones matches with the rating and production data. In the rating study,

sprouting turned out to be less acceptable than sluicing, and in the production study an implicit antecedent resulted in a lower rate of continuations referring to this antecedent than when it was explicit.

In the self-paced reading study we found no effect of the likelihood of the second participant given the verb that we had assessed with the pre-test. Since the production study clearly showed that verbs describing an action which is likely to involve a second participant increase the probability of a continuation referring to it, we expected that the pre-test score would also reduce processing effort and hence reading times in the sprouting continuation conditions. A possible explanation for the absence of such an effect is the difference between the probability that there *is* a second person participating in the action and the probability of somebody *talking* about this person. In case of a verb like 'to be on the phone' it might be more interesting to a potential listener who the interlocutor is, whereas in case of e.g. *feiern* 'to party' or *grillen* 'to barbecue' it might be more common to say *where* the subject of the utterance was partying or barbecuing than to tell who else participated in the action. The production data support this hypothesis, since for *telefonieren* 'to be on the phone' 68.18 % of the responses in the condition without explicit antecedent referred to the interlocutor, whereas only 14.29 % for *grillen* 'to barbecue' and 19.05 % for *feiern* 'to party' did.[13]

Taken together, we find that potential sluices are harder to process when the antecedent is implicit and when there are argument structure mismatches. Given the observation that these continuations are less likely in the production task and their elliptical variants are less likely in the rating task, this supports our processing account of sluicing and sprouting mismatches.

8 General Discussion

We used experimental methods to investigate the predictions of two syntactic identity constraints on sluicing proposed by Chung (2006, 2013) and of a processing account on the acceptability and processing of antecedent-target mismatches under sluicing and sprouting. While syntactic identity constraints predict a categorial difference between grammatical and ungrammatical mismatches, our processing account predicts a gradual effect of the likelihood of a potentially reduced phrase on its processing and the acceptability of its omission. Our experiments show that the acceptability of mismatches is indeed related to differences in predictability and processing difficulties. The data challenge both the constraints in Chung (2006) and those in Chung (2013).

To test whether the acceptability pattern reported in the literature is empirically supported, we first conducted an acceptability rating study. The rating study confirms the intuition that underlies the *Numeration Condition* (NC) in Chung (2006) that sprouting mismatches are specifically degraded, but, unlike the NC predicts, argument structure mismatches under sluicing are also heavily degraded. This in turn is expected given the *Argument Structure Condition* (ASC) by Chung (2013), but the ASC does not predict that sprouting mismatches are particularly unacceptable. Furthermore, mismatches involving adjunct sluices are as degraded as argument sluices, and this is totally unexpected given the ASC. Taken together, neither the 2006 nor the 2013 version of Chung's theory can explain the full acceptability pattern. Since our production study shows that sluices referring to implicit antecedents and mismatches are overall less likely, the results of the rating study are in line with the processing account.

In the production study, subjects produced higher ratios of continuations referring to a second participant involved in the action described by the verb if this participant was explicitly mentioned, i.e. in the configuration that leads to sluicing when the continuation is elliptical. In the case of implicit antecedents, there was a strong effect of the likelihood of a second

[13] Similarly, Poppels & Kehler (2019) show that violations of Chung's constraints improve when the potentially sluiced phrase corresponds to a salient QuD. From the perspective of our account, salient QuDs increase the likelihood of the omitted material and consequently favor ellipsis.

participant given the verb, which we did not observe for explicit antecedents. This supports our intuitions about the likelihood of sluices and is in line with the rating data: We also observed that continuations that were overall more likely are more likely to be reduced. Subjects produced only few argument structure mismatches and even less elliptical ones, which is also in line with the prediction of our processing account that predictable expressions are more likely to be reduced.

Finally we conducted a self-paced reading study in order to investigate whether less predictable continuations are harder to process. Our processing account predicts that in this case the full form should be preferred in order to distribute processing effort more uniformly across a larger number of words. The analyses of the data show that (i) mismatches are harder to process than matches and that (ii) continuations referring to implicit antecedents are harder to process than those referring to explicit ones. In case of ellipsis, this translates into a higher processing effort for sprouting than for sluicing. Surprisingly, we did not find an effect of the likelihood of a second participant given the verb on sluices with implicit antecedents, but this might be due to the relatively large degree of variation between items evidenced by the production study: For some verbs, like *grillen* 'to barbecue', it is very likely that other people than the subject are involved, however, other issues, like the food or location, are more likely to be talked about.

Except for the verb effect in the reading time data, the results of all three experiments are in line with our proposed processing account. The rating study gave clear evidence that none of the two versions of the syntactic identity account in Chung (2006, 2013) can explain the complete pattern. Whereas the coverage of these rules is restricted to specific types of ellipsis and they are not straightforwardly integrated into e.g. a Generative framework, processing principles are independently required and motivated and do not increase the complexity of the syntactic system. Our case study on argument structure mismatches under sluicing and sprouting thus suggests that probabilistic processing accounts might be a promising line of research to account for other syntactic identity contrasts under ellipsis, e.g. for VPE, gapping or right node raising, as well.

Acknowledgements

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 232722074 – SFB 1102 / Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102

References

Arregui, A., Clifton, C., Frazier, L., & Moulton, K. (2006). Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language*, *55*(2), 232–246.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Christensen, R. H. B. (2019). *ordinal – Regression models for ordinal data*. R package version 2019.04-25. https://CRAN.R-project.org/package=ordinal.

Chung, S. (2006). Sluicing and the lexicon: The point of no return. In *Annual Meeting of the Berkeley Linguistics Society*, volume 31, (pp. 73–91).

Chung, S. (2013). Syntactic Identity in Sluicing: How Much and Why. *Linguistic Inquiry*, *44*(1), 1–44.

Chung, S., Ladusaw, W. A., & McCloskey, J. (1995). Sluicing and logical form. *Natural Language Semantics*, *3*(3), 239–282.

de Lange, J. (2008). *Article omission in headlines and child language: a processing approach*. PhD thesis, LOT, Utrecht.

Delogu, F., Crocker, M. W., & Drenhaus, H. (2017). Teasing apart coercion and surprisal: Evidence from eye-movements and ERPs. *Cognition*, *161*, 46–59.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*, 193–210.

Drummond, A. (2019). Ibex 0.3.6 Manual.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL (Vol. 2)*, (pp. 159–166).

Jaeger, T. F. (2008). Modeling self-paced reading data: Effects of word length, word position, spill-over, etc. HLP/Jaeger lab blog. https://hlplab.wordpress.com/2008/01/23/modeling-self-paced-reading-data-effects-of-word-length-word-position-spill-over-etc/. Accessed on 7/5/2020.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62.

Lemke, R. (2021). Experimental investigations of the syntax and usage of fragments. Open Germanic Linguistis 1. Berlin: Language Science Press.

Lemke, R., Horch, E., & Reich, I. (2017). Optimal encoding! – Information Theory constrains article omission in newspaper headlines. In *Proceedings of the 15th Conference of the {E}uropean Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, (pp. 131–135).

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Levy, R. P., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt & T. Hoffman (Eds.), *Advances in neural information processing*, volume 19 (pp. 849–856). Cambridge, MA: MIT Press.

Merchant, J. (2001). *The syntax of silence. Sluicing, islands, and the theory of ellipsis*. New York: Oxford University Press.

Merchant, J. (2004). Fragments and ellipsis. *Linguistics and Philosophy*, *27*(6), 661–738.

Merchant, J. (2013). Voice and ellipsis. *Linguistic Inquiry*, *44*(1), 77–108.

Poppels, T., & Kehler, A. (2019). Reconsidering asymmetries in voice-mismatched VP-ellipsis. *Glossa: a journal of general linguistics*, *4*(1).

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. In J. H. Yoon & A. Kathol (Eds.), *Ohio State University Working Papers in Linguistics*, volume 49. Ohio State University.

Ross, J. R. (1969). Guess who? In Binnick, R., Davidson, A., & Green, G. (Eds.), *Papers from 5th regional meeting of the Chicago Linguistic Society*, (pp. 252–286)., Chicago, IL. Chicago Linguistic Society.

Runner, J., & Dozat, T. (2011). A brief report on voice mismatch effects in verb phrase ellipsis and sluicing. In A. Fine (Ed.), *University of Rochester Working Papers in the Language Sciences*, volume 6. University of Rochester.

Sag, I. A., Hofmeister, P., & Snider, N. (2007). Processing complexity in subjacency violations: the complex noun phrase constraint. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 43, (pp. 215–229).

Schäfer, L. (2021). Topic drop in German: Empirical support for an information-theoretic account to a long-known omission phenomenon. *Zeitschrift für Sprachwissenschaft*, *40*(2), 161–197.