The Preposition Stranding Generalization and ellipsis alternation

Joanna Nykiel

Abstract

Ellipsis alternation refers here to the alternation between two kinds of ellipsis remnants whose correlates are prepositional phrases. One kind of remnant includes the preposition hosted by its correlate and the other does not. I focus on the Preposition Stranding Generalization, which attempts to explain the availability of this alternation in syntactic terms, by connecting it to the possibility of preposition stranding in nonelliptical clauses. Drawing on two kinds of corpus evidence from English, I argue that the Preposition Stranding Generalization does not capture syntax-relevant information, but performance-based preferences. The first piece of evidence reveals that the content of a remnant and its correlate affects ellipsis alternation not only in languages without preposition stranding, but also in English. The second piece of evidence shows that the availability of preposition stranding in English nonelliptical clauses supports the use of prepositionless remnants via structural persistence, that is, reuse of syntactic structure found in antecedent clauses. These data lead me to conclude that the Preposition Stranding Generalization is wrongly assumed

to be a reliable test of whether underlying structure is present in elliptical constructions.

1 INTRODUCTION

On the basis of English corpus data, this paper explores what kind of generalization is captured by the Preposition Stranding Generalization (PSG) of Merchant (2001: 107). The PSG is formulated as follows:

(1) A Language L will allow preposition-stranding under sluicing just in case L allows preposition stranding under regular WH-movement.

The PSG refers to the elliptical construction sluicing, where a wh-phrase is left stranded and has an overt correlate in the antecedent clause (Ross 1969), as shown in (2)–(3). The stranded wh-phrases (remnants) and their correlates are marked in bold. Note that in the type of sluicing the PSG captures, the correlate is always a prepositional phrase (PP). This allows an alternation between two realizations of the remnant: either as a PP hosting the preposition present in the correlate, or as an NP lacking this preposition. Let us term this alternation ellipsis alternation.

(2) Katie is staying in Paris **with someone**, but I don't know **with who/who**.

(3) A: I'm counting **on my other friends**.
B: **On which other friends**?/**Which other friends**?

The PSG intends to capture an apparently categorical split between languages allowing preposition stranding in interrogative clauses and those

not allowing it. The availability of ellipsis alternation follows from movement and deletion operations possible only in the former kind of language. Both (4) and (5) can be derived in English through movement of the wh-phrase, with or without the preposition, to clause-initial position and subsequent deletion of the rest of the material (the deleted material is indicated by strikethrough). Example (5) illustrates preposition stranding, such that the preposition *with* remains at the end of the interrogative clause.

(4) Katie is staying in Paris with someone, but I don't know with who ~~Katie is staying in Paris~~.

(5) Katie is staying in Paris with someone, but I don't know who ~~Katie is staying in Paris with~~.

The PSG predicts that languages that disallow preposition stranding also disallow ellipsis alternation, allowing only PP remnants.

The status of the PSG as a syntactic generalization has rarely been questioned in the linguistic literature. It was in fact extended to another elliptical construction, fragment answers, shown in (6) (Merchant 2005). The PP correlate and alternating remnants are marked in bold. As in sluicing, the remnant may be realized as a PP or an NP, but is here argued to derive from an underlying declarative clause through fronting and deletion, as in (7) and (8).

(6) A: **What meds** are you **on**?
B: **On none/None**.

(7) None ~~I am on~~.

(8) On none ~~I am~~.

However, understanding what kind of generalization the PSG really is has implications for cross-linguistic work on elliptical constructions. The PSG is commonly assumed to be a reliable diagnostic of movement and with it, underlying syntactic structure in elliptical constructions (Stjepanović 2008, Szczegielniak 2008, Vicente 2008, Rodrigues, Nevins & Vicente 2009, Craenenbroeck 2010a, b, Kluck 2011, Yoshida, Nakao & Ortega-Santos 2015). This assumption is only correct if we can be certain that the PSG captures facts about the syntax of the languages to which it is applied. I argue that it does not, but rather, that it at best captures a pattern of preference found in languages with and without preposition stranding in nonelliptical clauses. More specifically, I propose that NP remnants are preferred over and more frequent than PP remnants in ellipsis alternation in languages that allow preposition stranding. I offer two kinds of evidence in support of this proposal. The significance of the evidence presented here is that the PSG is too strong and hence unable to support deletion-based theories of ellipsis (Ross 1969, Sag 1976, Merchant 2001, 2005) over direct interpretation ones, which generate ellipsis remnants without any underlying structure and rely on the surrounding context in assigning an interpretation to them (Ginzburg & Sag 2000, Culicover & Jackendoff 2005, Sag & Nykiel 2011).

The rest of the paper proceeds as follows. The next section turns to one factor known to affect the acceptability of ellipsis alternation cross-linguistically—the semantic and syntactic content of the wh-phrase used in the remnant. Section 3 addresses the nature of the correspondence between ellipsis alternation and preposition stranding in nonelliptical clauses, proposing that it is not syntactic, but based on structural persistence. Sec-

tion 4 presents two corpus studies of ellipsis alternation in US English. In section 5, I discuss the implications of the findings for theories of ellipsis. Section 6 concludes.

2 SEMANTIC AND SYNTACTIC CONTENT OF THE WH-PHRASE

The PSG is not exceptionless. One factor that affects the availability of ellipsis alternation in non-preposition-stranding languages is the semantic and syntactic content of the wh-phrase serving as a remnant. Examples (9) and (10) from Spanish illustrate a contrast between a bare wh-phrase and a which-NP phrase, such that (10) is preferred over (9) if the remnant is an NP (Rodrigues, Nevins & Vicente 2009: 2).

(9) Juan ha hablado con alguien, pero no sé quién.
Juan has talked with someone but not I know who
'Juan has talked with someone but I don't know who.'

(10) Juan ha hablado con una chica, pero no sé cuál.
Juan has talked with a girl but not I know which
'Juan has talked with a girl but I don't know which.'

Rodrigues, Nevins & Vicente (2009) stress that the status of examples like (9) ranges from marginal to acceptable, while examples like (10) are fully acceptable. These judgments are inconsistent with the PSG. However, Rodrigues, Nevins & Vicente (2009) argue that these examples derive from different underlying sources: examples like (9) have interrogative clause sources (see examples (4) and (5)), but examples like (10) have cleft sources, as shown in (11). On this analysis, Spanish sluicing uses two underlying structures.

(11) Juan ha hablado con una chica, pero no sé cuál es la
Juan has talked with a girl but not I know which is the
chica con la que ha hablado Juan
girl with the that has talked Juan
'Juan has talked with a girl but I don't know which is the girl that Juan has talked with.'

There are two problems with this analysis. First, the proposed cleft source does not easily generalize to sluicing outside of Spanish (as already noted by Vicente 2008). In languages with overt case marking, the which-NP phrase is marked for the same case as its correlate (Ross 1969), while the proposed cleft requires the nominative case. This is illustrated for Polish in (12) and (13).[1]

(12) Juan rozmawiał z jakąś dziewczyną ale nie wiem
Juan talked with some girl.INSTR but not I know
którą.
which.INSTR
'Juan talked with a girl, but I don't know which.'

(13) *Juan rozmawiał z jakąś dziewczyną ale nie wiem
Juan talked with some girl.INSTR but not I know
która jest dziewczyna z którą Juan
which.NOM is girl with whom.INSTR Juan
rozmawiał.
talked
'Juan talked with a girl, but I don't know which is the girl Juan talked with.'

Non-preposition-stranding languages are now known to be non-uniform with respect to having any cleft sources. While some (Brazilian Portuguese, Spanish) have been argued to have them, others have been shown to have none (Indonesian, Amis, Serbo-Croatian, Polish, Russian, Emirati Arabic), although ellipsis alternation is available in all of them (Stjepanović

[1]The notation X.INSTR and X.NOM indicates the nominative and instrumental case.

2008, Vicente 2008, van Craenenbroeck 2010a, b, Sag & Nykiel 2011, Sato 2011, Wei 2011, Nykiel 2013a; pace Szczegielniak 2008, Leung 2014, Philippova 2014).

The second problem with Rodrigues, Nevins & Vicente's (2009) analysis relates to the acceptability contrast between examples (9) and (10). Beyond noting that (9) is less acceptable than (10), Rodrigues, Nevins & Vicente (2009) do not explicitly address its status with respect to unacceptable instances of preposition stranding in nonelliptical clauses. Let us note first that the contrast does not reduce to the content of the wh-phrases alone, but to both the wh-phrases and their counterparts in the correlates. In example (9), an indefinite pronoun is paired with a bare wh-phrase, and in example (10), an NP is paired with a corresponding which-NP phrase, matching in terms of their syntactic and semantic content. There is evidence that mismatch in such content degrades the acceptability of sluicing, whether or not ellipsis alternation is involved, as in (14)–(15) (Dayal & Schwarzschild 2010, Nykiel 2013b).[2]

(14) *Kim saw someone on the beach, but I don't know which actor.

(15) *Kim saw a restaurant on the beach, but I don't know what (Intended: I don't know what she saw).

Further, independent evidence shows that matching contentful correlates and remnants (NPs and which-NP phrases) are found more acceptable than matching noncontentful ones (indefinite pronouns and bare wh-phrases) in English sluicing with appositive antecedents, as in (16)-(17) from Collins, Popova, Sag & Wasow (2015: Appendix).

[2]These results provide an independent explanation for why a bare wh-phrase is more degraded than a which-NP phrase in Spanish when the correlate contains an NP, see discussion in Rodrigues, Nevins & Vicente (2009: ex. 50b).

(16) My brother Steve, who says he read **something** last week, can't remember **what**.

(17) My brother Steve, who says he read **an interesting book** last week, can't remember **which book**.

If the acceptability contrast between contentful and noncontentful correlates and remnants is a general characteristic of sluicing, it is not surprising to find it in ellipsis alternation, as well. In fact, this contrast in ellipsis alternation is robust in several non-preposition-stranding languages beyond Spanish: French, Serbo-Croatian, Polish, Czech, and Russian (Stjepanović 2008, Szczegielniak 2008, Rodrigues, Nevins & Vicente 2009, Caha 2011, Sag & Nykiel 2011, Nykiel 2013a, Tatiana Philippova p. c.).[3] Note that this list includes languages that have been shown to lack any underlying cleft sources. This suggests that a larger generalization is missed if we analyze constraints on ellipsis alternation in Spanish sluicing as different than constraints on ellipsis alternation in the languages listed above, and in particular, if we ignore the possibility that examples like (9) are not unacceptable, but merely more degraded than examples like (10). This possibility has already been tested and supported for Polish in Nykiel (2013a). Given the cross-linguistic robustness of the effect of the content of a wh-phrase and its counterpart in the correlate on the acceptability of ellipsis alternation, it has been suggested that mental accessibility of the correlate determines the specificity of instructions that the remnant has to provide for successful resolution (Sag & Nykiel 2011, Nykiel 2013a,

[3] An acceptability contrast between contentful and noncontentful wh-phrases in not unique to sluicing. For example, Hofmeister and Sag (2010) demonstrate experimentally that increasing the semantic and syntactic content of wh-phrases improves the acceptability of island violations in English filler-gap constructions, and attribute these results to constraints on language processing. Specifically, they argue that the perception of how acceptable island violations are depends both on grammatical principles and performance-based constraints, such as memory limitations.

2015). These proposals build on research on anaphora resolution (Ariel 1990, 2001, Martin & McElree 2011) but their details are not our concern here.

I hypothesize that contentful wh-phrases and their NP counterparts in the correlates raise the acceptability of NP remnants in ellipsis alternation even in languages allowing preposition stranding. I test this hypothesis in English corpus data in section 4 and find clear support for it, which strongly suggests that there is no compelling reason to propose different underlying sources for sluicing in Spanish for examples with contentful and noncontentful wh-phrases. Rather, I explore a more empirically accurate alternative: the PSG is formulated based on examples with noncontentful wh-phrases, and their acceptability is degraded with respect to examples with contentful wh-phrases across all the languages considered.[4] This alternative raises the obvious question of why there should be an acceptability contrast between the preposition-stranding languages and the non-preposition-stranding ones. I turn to this question next.

3 STRUCTURAL PERSISTENCE

We have seen thus far that a cross-linguistically robust pattern of preference is found in ellipsis alternation that challenges the status of the PSG as a syntactic generalization. However, the availability of preposition stranding in nonelliptical clauses is likely to affect ellipsis alternation in ways other than syntactic. To see this, consider the examples of sluicing and fragment answers in (18)–(21) (the correlates are marked in bold).[5]

[4]Merchant (2001) admits that there is some unexplained variation in acceptability judgments in his data, for example in French.

[5]These examples come from the dataset collected for this study. The annotations SB and S refer to the Santa Barbara and Switchboard corpora.

(18) A: When you think of a wedding cake, **what** do you think **of**?
B: Marriage. (SB)

(19) A: **What area** do you live **in**?
B: North Carolina. (SB)

(20) A: He's **in the army**.
B: Which one?
A: Ours. (S)

(21) A: Mostly Sharon got on her bandwagon **about Missus Jackson**.
B: About who? (SB)

In (18) and (19), the remnants have correlates in interrogative clauses with preposition stranding, with the result that the correlates are not continuous PPs. The likelihood that remnants are realized as NPs is higher for such examples than for examples like (20) and (21), in which the correlates are continuous PPs. A remnant also has a high chance of being realized as an NP if its correlate has also been realized as an NP. This is the case in (20): the phrase *Which one*, itself a sluicing remnant realized as an NP, serves as the correlate for *Ours*. These patterns reflect structural persistence—reuse of syntactic structure encountered in the surrounding discourse. That is, speakers reuse the structure of the correlate in the remnant by either realizing the remnant as a PP, which repeats the structure of a PP correlate whose constituents are adjacent, or by realizing it as an NP, which repeats the structure of a PP correlate where the preposition is separated from the complement or altogether missing.

Structural persistence occurs at different levels of linguistic description: semantics, syntax, morphology, lexis and phonetic form (Meyer &

Schvaneveldt 1971, Tanenhaus, Flanigan & Seidenberg 1980, Kempley & Morton 1982, Levelt & Kelter 1982, Weiner & Labov 1983, Bock 1986, Hartsuiker & Westenberg 2000, Branigan, Pickering & Cleland 2000, Szmrecsanyi 2005). Structural persistence effects are observed at the level of syntax when speakers repeat syntactic structure as their discourse evolves. These effects persist both within individual speakers (self-priming) and in dialogue. Such a reuse of syntactic structure is argued to facilitate language production by allowing speakers to use structures that they have implicitly learned from exposure to other speakers' output, as well as their own (MacDonald 2013). It has been suggested that patterns of preference found in syntactic variation can at least in part be explained by speakers' exposure to similar structures appearing in prior discourse (Branigan, Pickering & Cleland 2000, Szmrecsanyi 2005, 2006, Bresnan 2007, Bresnan, Cueni, Nikitina & Baayen 2007, Ford & Bresnan 2013, Nykiel 2015).

Of relevance to the argument developed here is Levelt & Kelter's (1982: 80) study of Dutch fragment answers. They demonstrate experimentally that Dutch speakers reuse the syntax of questions like (22) and (23), which optionally contain prepositions, in their elliptical responses. The response to (22) usually corresponds to the structure of the question by also containing the preposition, as in (24), and the response to (23) corresponds to it by containing no preposition, as in (25). These correspondences are statistically significant. But Levelt & Kelter (1982) demonstrate that reuse effects disappear when additional linguistic material intervenes between the question and the answer, as in (26).

(22) Aan wie laat Paul zijn viool zien?
to whom lets Paul his violin see
'Who does Paul allow to see his violin?'

(23) Wie laat Paul zijn viool zien?
whom lets Paul his violin see
'Who does Paul allow to see his violin?'

(24) Aan Toos.
to Toos
'Toss.'

(25) Toos.
Toos
'Toos.'

(26) Om hoe laat/Hoe laat gaat uw winkel dicht, want ik
at what time/what time goes your store closed because I
moet er speciaal voor naar de stad komen, ziet u?
must there especially for to the town come, know you
'At what time/What time does your store close, since I have to come to town especially for that, you know?'

In the data to be presented here, the correlate always appears in the immediately preceding sentence, and hence we may expect to see clear structural persistence effects. Of the two constructions we have discussed, fragment answers are directly affected by preposition stranding if the correlate appears in an interrogative clause as an noncontinuous PP (see (18) and (19)). There is only one example in the current data where sluicing is so affected, shown in (27). As predicted, the remnant is realized as an NP following its noncontinuous correlate.

(27) A: **What** area are you **in**?
B: **What area of the country**? I'm in southwest Houston. (SB)

In the next section, I explore how strong exactly structural persistence effects are across both sluicing and fragment answers. It would appear that these effects can account for the distribution of PP remnants

and NP remnants in languages with and without preposition stranding, because remnants will be more often realized as NPs in the former. Existing research on ellipsis alternation suggests that NP remnants are more frequent than PP remnants in preposition-stranding languages, but not in non-preposition-stranding languages. The environments in which NP remnants are reported to be acceptable (contentful wh-phrases and correlates) in non-preposition-stranding languages are a subset of the environments in which NP remnants are acceptable (contentful and noncontentful wh-phrases and correlates) in preposition-stranding languages. Unsurprisingly, the current data show that the ratio of NP remnants to PP remnants is 276 (67.2%) to 135 (32.8%) for English. A small sample (25 examples) from Norwegian puts the ratio of NP remnants to PP remnants at 22 (88%) to 3 (12%). In contrast, the ratio of NP remnants to PP remnants is 57 (18.3%) to 254 (81.7%) for Polish.[6] We can conclude from these data that the PSG does not capture a grammatical constraint on ellipsis alternation, but an acceptability contrast linked to the frequency of NP and PP remnants in languages with and without preposition stranding.

It is well known that frequency correlates with acceptability ratings. For example, Keller (2000) compares corpus frequencies and acceptability ratings and finds that the most frequent structures receive the highest ratings, with a decrease in frequency corresponding to a decrease in acceptability ratings. This correspondence is often not perfect: structures that are actually not found in corpora can be rated as acceptable to various degrees (Konieczny 2000, Featherston 2005, Kempen & Harbusch 2005). However, for structures that are found in corpora, their corpus frequency closely tracks their acceptability. This evidence, along with the corpus

[6]The Norwegian and Polish samples were collected from the Norwegian Speech Corpus–the Oslo part (NoTa–Oslo) and the National Corpus of Polish.

frequencies for NP and PP remnants for English, Norwegian, and Polish, challenges the assumption that the PSG is a syntactic generalization. We already know that the acceptability of any NP remnants is higher than the acceptability of grammatical violations such as preposition stranding in interrogative clauses in Polish and in Spanish, contrary to prediction. So long as we lack sufficient empirical evidence for the rest of the languages on which the PSG is based, we cannot rule out the possibility that the PSG can be explained in terms of performance-based factors. The next section provides evidence for one such factor.

4 CORPUS STUDIES

The data collected for these studies come from three corpora of spoken American English: the Switchboard corpus (henceforth S), Santa Barbara (henceforth SB) and the Corpus of Contemporary American English (henceforth COCA). I first extracted all wh-phrases from the first two corpora and then identified those that were instances of sluicing with PP correlates. Next, I identified those wh-phrases which were embedded in interrogative clauses as prepositional objects and selected those that had elliptical responses. Following these procedures, I extracted remnants that instantiated either sluicing or fragment answers. These data constitute 60% of the entire dataset. As for COCA, I extracted an equal sample of bare wh-phrases and which/what/whose-NP phrases from the spoken part of the corpus to avoid a potential bias toward a particular kind of wh-phrase. In extracting relevant remnants, I followed the same procedures as before. I extracted the total of 411 ellipsis remnants, of which 276 (67.2%) were NP remnants, and coded them for the factors discussed in sections 2 and 3, as described in detail below.

For statistical analysis, this dataset was reduced to a smaller sample of 310 items. This step was necessary because two types of items showed a pattern of behavior distinct from the rest. These two types are illustrated in (28) and (29) (the correlates and remnants are shown in bold). Example (28) resembles fragment answers except that both the question and answer occur within a single speaker's turn; such examples are termed split questions in the literature (Camacho 2002, Arregi 2010). Example (29) resembles sluicing except that the remnant here has a clarificational function: note that the proper noun *Aguilar* has not been heard correctly, and hence its referent has not been identified at the point the remnant occurs. Such examples are referred to as reprise questions (Ginzburg & Sag 2000). These items were removed from the data because split questions have an independent preference for NP remnants over PP remnants (46 to 10), and reprise questions have an independent preference for PP remnants over NP remnants (30 to 15), which is unusual for English.[7] After removing these two types of items from the data, only canonical sluicing and fragment answers remained.

(28) This cake is filled **with what**, **whipped cream**? (S)

(29) But he couldn't get along **with Aguilar**.
B: **With who**?
A: I mean Aguirre.
B: Oh, Aguirre. (S)

[7]For more on these patterns, see Nykiel (2014, 2015).

4.1 *Corpus study 1*

The data were annotated in terms of the content of the correlate. As discussed in section 2, correlates and remnants normally match in content under sluicing, such that, for example, a correlate hosting an NP is coupled with a remnant hosting a which/what/whose-NP phrase. However, the acceptability of fragment answers is not affected by a distinction between matching and nonmatching correlates and remnants. In example (30), the correlates host a contentful phrase (the NP *what college*) and a noncontentful one (the interrogative pronoun *what*), while the remnants are both contentful phrases—NPs. By coding the data for the content of the phrase hosted by the correlate, we capture the pattern relevant for fragment answers and at the same time ensure that a match between the correlate and remnant can also be captured for sluicing. The coding scheme simply captures the binary distinction between NPs and (indefinite or interrogative) pronouns, that is, between contentful and noncontentful correlates. The hypothesis tested here is that remnants are more frequently realized as NPs if their correlates are contentful than if they are noncontentful.

(30) A: So you went **to what college**?
B: **Northern State University in Aberdeen, South Dakota**.
A: And **what** did you graduate **in**?
B: **International business**. (S)

4.1.1 *Results*

The results are given in Table 1. Consistent with our hypothesis, the percentage of NP remnants is higher for contenful correlates than noncontentful ones. Statistical significance of this pattern was tested by fit-

ting a mixed-effects logistic regression model to the data (Baayen 2008). The model predicted the realization of remnants based on the content of their correlates.[8] Compared to contentful correlates, noncontentful correlates lowered the likelihood that remnants would appear as NPs ($\beta = -0.97, SE = .42, zvalue = -2.3, p < .05$).

Realization of correlate	PP remnant	NP remnant	Total remnants
contentful	42(24.1%)	132(75.9%)	174(100%)
noncontentful	53(39%)	83(61%)	136(100%)

Table 1: Realization of remnants by correlate content

4.2 *Corpus study 2*

To control for the effect of structural persistence, I annotated each antecedent for whether it contained a continuous PP correlate or a discontinuous one (with the preposition either separated from its object or missing altogether). If preposition stranding has applied in an antecedent clause, the correlate consists of a preposition separated from, and following, its object, as in (18) repeated here as (31). It might seem that for antecedent clauses with preposition stranding, the correlate does not contain the preposition but only its object. As can be seen from Table 2 below, however, PP remnants appear with discontinuous correlates, suggesting that the correlate is the entire PP, not only its object. For convenience, I used the label discontinuous to also code correlates appearing as NPs rather than PPs, as in (20) repeated here as (32). The separation of the preposition from its object or its absence could prime speakers to use NP

[8] The model was developed in the free software R, using the package lme4. Combinations of prepositions and verbs/nouns/adjectives were entered as a random effect, since some appeared in the data more than once.

remnants. In contrast, a continuous correlate hosts a preposition preceding its object and is embedded in one of three environments: a declarative clause (21) (repeated here as (33)), an in-situ interrogative (34) or an elliptical PP (35). I hypothesized that, all else being equal, discontinuous correlates are more likely than continuous correlates to yield NP remnants.

(31) A: When you think of a wedding cake, **what** do you think **of**?
B: Marriage. (SB)

(32) A: He's **in the army**.
B: Which one?
A: Ours. (S)

(33) A: Mostly Sharon got on her bandwagon **about Missus Jackson**.
B: About who? (SB)

(34) A: We're products **of what**?
B: A cultural process. (SB)

(35) A: Learn the rules of the game, play the game.
B: **For what**?
A: For whatever you want. (S)

4.2.1 *Results*

Table 2 gives the statistics for the realization of remnants with respect to the realization of all correlates. Table 3 gives a subset of the data in Table 2: discontinuous correlates hosted by clauses with preposition stranding vs. all continuous correlates.

Realization of correlate	PP remnant	NP remnant	Total remnants
discontinuous	7(6.9%)	94(93.1%)	101(100%)
continuous	88(42.1%)	121(57.9%)	209(100%)

Table 2: Realization of remnants by realization of correlates (all correlates)

Realization of correlate	PP remnant	NP remnant	Total remnants
discontinuous	5(7.7%)	60(92.3%)	65(100%)
continuous	33(30.5%)	75(69.5%)	108(100%)

Table 3: Realization of remnants by realization of correlates (continuous correlates and discontinuous correlates hosted by clauses with preposition stranding)

Although both types of correlate prefer NP remnants across both Table 1 and 2, an increase can be seen in the number of NP remnants where the correlates are discontinuous. To verify whether these patterns were statistically significant, I fitted a mixed-effects logistic regression model to these data predicting the realization of remnants as NPs or PPs, based on the form of the correlate.[9] Remnants were found to be significantly less likely to be realized as NPs for continuous correlates than for discontinuous ones ($\beta = -2.37, SE = .45, zvalue = -5.2, p < .001$). This was also true when continuous correlates were compared only with discontinuous correlates hosted by interrogative clauses with preposition stranding ($\beta = -1.85, SE = .61, zvalue = -3, p < .01$). These results confirm that the availability of preposition stranding in English nonelliptical clauses is relevant to the realization of remnants in terms of structural persistence.

[9] As before, the model was developed in the free software R, using the package lme4. It also included combinations of prepositions and verbs/nouns/adjectives as a random effect.

Note also that discontinuous correlates have a strong presence in the data, since they constitute about one third of all correlates.

5 DISCUSSION

The results discussed in the previous section provide clear answers to the two questions that this paper asks. First, is there a preference for NP remnants to occur with contentful correlates that generalizes to preposition-stranding and non-preposition-stranding languages alike? Second, is there a nonsyntactic reason why NP remnants are overall more frequent and more acceptable in preposition-stranding languages than non-preposition-stranding ones? Both questions relate to the status of the PSG and its use as a diagnostic of underlying syntactic structure in ellipsis remnants.

As for the first question, it is clear that contentful correlates are more frequently coupled with NP remnants than noncontentful correlates are in English. This is the same pattern as that observed in Spanish and the non-preposition-stranding languages listed in section 2. The cross-linguistic generality of this pattern strengthens the possibility that NP remnants appearing with noncontentful correlates are not unacceptable or ungrammatical in any language but merely degraded with respect to NP remnants appearing with contentful correlates. It is unjustified to impose on NP remnants a binary division into acceptable and unacceptable ones without an empirically grounded comparison with acceptable and unacceptable structures for any given language. Rodrigues, Nevins & Vicente's (2009) decision to collapse NP remnants appearing with noncontentful correlates with unacceptable structures in Spanish is problematic for this reason. Recall further that noncontenful correlates (with matching remnants) are more degraded than contentful correlates (with matching rem-

nants) under sluicing outside of ellipsis alternation, which, together with the current data, suggests that this degradation is a general characteristic of both sluicing and fragment answers rather than evidence for an alternative underlying structure for Spanish sluicing. The bulk of support for the PSG comes from noncontentful phrases serving as correlates and remnants, without proper attention to their status with respect to contentful phrases or structures that are unacceptable in the languages considered. This means that the PSG is only a partial statement about the data.

The second question address a difference in the acceptability of NP remnants with noncontentful correlates between preposition-stranding language and non-preposition-stranding ones. There is at least one reason to believe that this difference is not syntactic, but follows from reuse of the structure of the correlate. We have seen that the availability of preposition stranding in English has a measurable effect on ellipsis alternation by promoting NP remnants. This results in an overall higher frequency of NP remnants than PP remnants, and other preposition-stranding languages may be expected to behave like English. On the assumption that high-frequency structures receive high acceptability ratings, all NP remnants should receive high acceptability ratings in English, possibly higher than PP remnants. For non-preposition-stranding languages, the set of non-continuous correlates is reduced to remnants that have themselves been realized as NPs and serve as correlates for subsequent remnants, with the result that PP remnants outnumber NP remnants. Since the overall frequency of NP remnants is lower than the frequency of PP remnants in such languages, their overall acceptability should also be lower than the acceptability of PP remnants (see Nykiel 2013a for empirical evidence). Hence, an explanation for acceptability differences between NP remnants and PP remnants in various languages may be found in the distributional

frequencies of these remnants.

Both the observed structural persistence effects and the effects following from the content of the remnant and its correlate are performance-based. Structural persistence effects follow directly from the architecture of language production and are commonly found in studies of syntactic variation in the sense that the choice of syntactic alternatives is constrained by structures appearing in previous discourse, as discussed in section 3. Structural persistence effects arise in ellipsis alternation because PP correlates are hosted by clauses with different possibilities for preposition placement, and these possibilities are then repeated in the structure of remnants. Why the content of a correlate and remnant should affect ellipsis alternation, as well as sluicing and fragment answers in general, is not well known yet. One possibility, building on Ariel (1990, 2001), is that for all anaphoric constructions, the anaphor serves to provide instructions necessary to retrieve the antecedent, that is, its explicitness tells the hearer how accessible the antecedent is. Ariel (1990) observes that the less accessible the antecedent is, the more explicit the anaphor. If noncontentful phrases are less accessible correlates than contentful phrases under sluicing, as is suggested by an acceptability difference between them, then PP remnants are more explicit than NP remnants, serving as better retrieval instructions. This possibility, which is explored in detail in Nykiel (2013b, 2015), also follows from the architecture of language production.

The current results strongly suggest that the PSG cannot diagnose underlying syntactic structure in ellipsis remnants, because it does not constitute syntax-relevant evidence. It has been argued in the literature that patterns that can be explained by performance-based factors are not relevant to syntax and should not be analyzed as such (Featherston 2005). Even in English, the data surrounding ellipsis alternation are not free of

variation that is unexpected from the point of view of the PSG. For the examples in (36) and (37), preposition stranding is unavailable in the corresponding interrogative clauses shown in (38) and (39), and hence the acceptable NP remnants cannot be derived from these sources (Chung, Ladusaw & McCloskey 1995: 273). Preposition stranding is also unavailable for deriving the attested example in (40), because the phrase *let alone* allows no sentential source (for more detail, see Culicover & Jackendoff 2005).

(36) We will use force **under certain circumstances**, but we can't say **what**.

(37) This theory is right **in some sense**, but I'm not sure **what**.

(38) *What circumstance will we use force under?

(39) *What sense is this theory right in?

(40) When he started, he was unable to speak **to me**, let alone **other students and tutors**.[10]

However, if the availability of preposition stranding only affects ellipsis alternation as a performance-based factor, these NP remnants are entirely unproblematic. They are, in fact, expected, given their contentful correlates. These facts, taken together with the current results, are easily accommodated on direct interpretation approaches to ellipsis (Ginzburg & Sag 2000, Culicover & Jackendoff 2005, Sag & Nykiel 2011) by allowing the grammar to place no constraints on ellipsis alternation. Instead,

[10](http://www.thefreelibrary.com/Youn+Rhys+now+no+longer+forced+to+suffer+in+silence%3B+Teenager's+life...-a0192081449)

constraints on language production are responsible for the cross-linguistic distribution of remnants.

6 CONCLUSION

This paper has argued that the PSG is wrongly taken to be a reliable test of whether underlying structure is present in elliptical constructions. In support of this argument, I have offered evidence from two corpus studies of ellipsis alternation in English. The first of them demonstrates that the content of a remnant and its correlate affects ellipsis alternation not only in languages without preposition stranding, but also in English. The second study shows that the availability of preposition stranding in English nonelliptical clauses supports the use of NP remnants via structural persistence. Both of these patterns of preference follow from the architecture of language production, and hence are performance-based. I conclude that the PSG attempts to explain patterns that are fundamentally not syntactic in nature.

REFERENCES

Ariel, Mira. 1990. *Accessing Noun Phrase Antecedents*. London-New York: Routledge.

Ariel, Mira. 2001. Accessibility theory: An overview. In: T. Sanders, J. Schilperoord, and W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects*, 29–87. Amsterdam: John Benjamins.

Arregi, Karlos. 2010. Ellipsis in split questions. *Natural Language and Linguistic Theory* 28: 539–592.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18: 355–387.

Branigan, Holly P., Martin J. Pickering & Alexandra A. Cleland. 2000. Syntactic coordination in dialogue. *Cognition* 75(2): B13–B25.

Bresnan, Joan. 2007. Is knowledge of syntax probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in Search of Its Evidential Base*, Series: Studies in Generative Grammar, 75–96. Berlin and New York: Mouton de Gruyter.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive Foundations of Interpretation*, 69-94. Amsterdam: Royal Netherlands Academy of Science.

Caha, Pavel. 2011. Case in adpositional phrases. Ms., CASTL, Tromsø.

Camacho, José. 2002. Wh-doubling: implications for the syntax of wh-movement. *Linguistic Inquiry* 33: 157–164.

Chung, Sandra, William Ladusaw & James McCloskey. 1995. Sluicing and logical form. *Natural Language and Linguistic Theory* 3: 239–282.

Collins, James, Daria Popova, Ivan A. Sag & Thomas Wasow. 2015. Sluicing and the inquisitive potential of appositives. In Marlies Kluck, Dennis Ott & Marc de Vries (eds.), *Parenthesis and ellipsis: Cross-linguistic and theoretical perspectives*, 47–74. Berlin/Boston/Munich: Walter de Gruyter.

Craenenbroeck, Joeren van. 2010a. *The syntax of ellipsis: Evidence from Dutch dialects*. New York, NY: Oxford University Press.

Craenenbroeck, Joeren van. 2010b. Invisible Last Resort. A note on clefts as the underlying source for sluicing. *Lingua* 120: 1714–1726.

Culicover, Peter & Ray Jackendoff. 2005. *Simpler syntax*. New York: Oxford University Press.

Dayal, Veneeta & Roger Schwarzschild. 2010. Definite inner antecedents and wh-correlates in sluices. In Peter Staroverov, Daniel Altshuler, Aaron Braver, Carlos A. Fasola & Sarah Murray (eds.), *Rutgers Working Papers in Linguistics* Vol. 3, 92–114. New Brunswick, NJ: LGSA.

Featherston, Sam. 2005. The decathlon model of empirical syntax. In M. Reis and S. Kepser (Eds.), *Linguistic evidence: Empirical, Theoretical and Computational Perspectives*, 187–208. Berlin: de Gruyter.

Ford, Marilyn and Joan Bresnan. 2013. Using convergent evidence from psycholinguistics and usage. In Manfred Krug & Julia Schlüter (Eds.), *Research Methods in Language Variation and Change*, 295–312. Cambridge: Cambridge University Press.

Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative investigations. The form, meaning and use of English interrogative*. Stanford, CA: CSLI Publications. [Distributed by U. Chicago Press].

Hartsuiker, Robert J. & Casper Westenberg. 2000. Word order priming in written and spoken sentence production. *Cognition* 75: B27–B39.

Hofmeister, Philip & Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86: 366–415.

Keller, Frank. 2000. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Ph.d. thesis, University

of Edinburgh, Edinburgh, UK.

Kempen, Gerard & Karin Harbusch. 2005. The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses. In Marga Reis & Stephan Kepser (eds.), *Linguistic evidence: Empirical, Theoretical and Computational Perspectives*, 329–350. Berlin: de Gruyter.

Kempley, Steven T. & John Morton. 1982. The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology* 73: 441–445.

Kluck, Marlies. 2011. Sentence amalgamation. Doctoral dissertation, University of Groningen.

Konieczny, Lars. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research* 29: 627–645.

Leung, Tommy. 2014. The Preposition Stranding Generalization and conditions on sluicing: Evidence from Emirati Arabic. *Linguistic Inquiry* 45: 332–340.

Levelt, Willem & Stephanie Kelter. 1982. Surface form and memory in question answering. *Cognitive Psychology* 14: 78–106.

Martin, Andrea E. & Brian McElree. 2011. Direct-access retrieval during sentence comprehension: Evidence from sluicing. *Journal of Memory and Language* 64: 327–343.

MacDonald, Maryellen. C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4, Article 226.

Merchant, Jason. 2001. *The syntax of silence: sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.

Merchant, Jason. 2005. Fragments and ellipsis. *Linguistics and Philosophy* 27: 661–738.

Meyer, David & Roger Schvaneveldt. 1971. Facilitation in recognizing pairs of words: Evidence of dependence between retrieval operations. *Journal of Experimental Psychology* 90: 227–234.

Nykiel, Joanna. 2013a. Clefts and preposition omission in sluicing. *Lingua* 123:74–117.

Nykiel, Joanna. 2013b. Wh-phrases in sluicing: An interaction of the remnant and the antecedent. In: Hofmeister, P. and E. Norcliffe (Eds.) *The core and the periphery. Data-driven perspectives on syntax inspired by I. A. Sag*, 251–271.

Nykiel, Joanna. 2014. The ellipsis alternation: Remnants with and without prepositions. Manuscript.

Nykiel, Joanna. 2015. Constraints on ellipsis alternation: A view from the history of English. *Language Variation and Change* 27: 227–254.

Philippova, Tatiana. 2014. P-omission under sluicing, [P clitic] and the nature of P-stranding. In Martin Kohlberger, Kate Bellamy & Eleanor Dutton (eds.),*Proceedings of ConSOLE XXII*, 133–155.

Rodrigues, Cilene, Andrew Nevins, and Luis Vicente. 2009. Cleaving the interactions between sluicing and preposition stranding. In Daniele Torck & Leo Wetzels (eds.), *Romance Languages and Linguistic Theory 2006*, 175–198. Amsterdam: John Benjamins Publishing Company. variation in English. *Language* 8: 613–644.

Ross, John R. 1969. Guess who? In: R. Binnick, A. Davison, G. Green, and J. Morgan (Eds.), *Papers from the 5th Regional Meeting of the*

Chicago Linguistic Society, 252–286, Chicago: Chicago Linguistic Society.

Sag, Ivan A. 1976. Deletion and logical form. Doctoral dissertation. MIT [Published in 1980, New York: Garland Press].

Sag, Ivan A. & Joanna Nykiel. 2011. Remarks on sluicing. In: S. Mueller (Ed.) *Proceedings of the HPSG 11 Conference. CSLI Publications*. (http://csli-publications.
stanford.edu/HPSG).

Sato, Yosuke. 2011. P-stranding under sluicing and repair by ellipsis: why is Indonesian (not) special? *Journal of East Asian Linguistics* 20: 339–382.

Stjepanović, Sandra. 2008. P-stranding under sluicing in a non-P-stranding language? *Linguistic Inquiry* 37(1): 179–190.

Szczegielniak, Adam. 2008. Islands in sluicing in Polish. In: N. Abner and J. Bishop (Eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics, Cascadilla Proceedings Project*, 404–412. Somerville, MA, USA.

Szmrecsanyi, Benedikt. 2005. Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 11: 113–150.

Szmrecsanyi, Bendeikt. 2006. *Morphosyntactic persistence in spoken English. A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. (Series: Trends in Linguistics. Studies and Monographs [TiLSM] 177). Berlin/New York: Mouton de Gruyter.

Tanenhaus, Michael K., Helen P. Flanigan & Mark Seidenberg. 1980. Orthographic and phonological activation in auditory and visual word

recognition. *Memory and Cognition* 8: 513–520.

Vicente, Luis. 2008. Syntactic isomorphism and non-isomorphism under ellipsis. Ms. UCSC.

Wei, Ting-Chi. 2011. Parallelism in Amis sluicing. *Concentric: Studies in Linguistics* 37(1): 1–44.

Weiner, E. Judith & William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19: 29–58.

Yoshida, Masaya, Chizuru Nakao & Iván Ortega-Santos. 2015. The syntax of *why-stripping*. *Natural Language and Linguistic Theory* 33: 323–370.