

Preposition omission in English ellipsis

Joanna Nykiel (University of Silesia)

Introduction

The availability of preposition omission in elliptical constructions, has since Merchant (2001) been assumed to follow from the possibility of preposition stranding under *wh*-movement. This correspondence is illustrated in (1), where the A-sentence instantiates an interrogative clause with preposition stranding, and hence, the answer in B (an example of Bare Argument Ellipsis, henceforth BAE) has the option of omitting the preposition *with*. If correct, this correspondence, known as the Preposition Stranding Generalization (PSG), suggests that underlying structure is present in elliptical constructions such that BAE in (1) arises via movement and deletion (Merchant 2001), as in (2):

- (1) A: What's this cake filled with? B: (With) whipped cream.
- (2) Whipped cream ~~this cake is filled with~~.

Actual linguistic behavior, however, fails to support this correspondence. Much research on languages without preposition stranding shows that preposition omission is available in sluicing, another elliptical construction shown in (3) (Almeida and Yoshida 2005, Szczegielniak 2008, Stjepanović 2008, Rodrigues et al. 2009, Sag and Nykiel 2011).

- (3) I heard it from someone, but I don't even remember who.

That the behavior of preposition-stranding languages, such as English, too, may be problematic for the PSG has already been noted in the literature (Chung et al. 1995, Fortin 2007). Some English prepositions do not strand easily (4)-(5), and some constructions do not allow preposition stranding (6)-(7). And yet, all these sentences have acceptable sluicing variants without prepositions, shown in (8)-(11).

- (4) ?Which version has this quest been in the game since?
- (5) ?Whose efforts did everything look rather tarnished despite?
- (6) *What car what happened with? (cf. What happened with what car?)
- (7) *What cake what did Jani think of? (cf. What did Jani think of what cake?)
- (8) This quest been in the game since some version, but I don't remember which.
- (9) A: Everything looked rather tarnished despite someone's efforts. B: Whose?
- (10) A: What happened with the car? B: What car?
- (11) A: What did Jani think of the cake? B: What cake?

Based on these data I hypothesized that the possibility of preposition omission in ellipsis cannot be accurately predicted from the availability of preposition stranding in nonelliptical clauses in any language. My goal here is to empirically test this hypothesis for English and identify factors that can predict omission of prepositions in elliptical constructions. More generally, I explore

whether the pattern of preposition omission indicates that underlying structure is present in ellipsis, and hence, that ellipsis arises through movement and deletion.

Data and method

Data were collected in two stages: a corpus study and a follow-up experiment. I sampled 409 ellipsis remnants from three corpora of spoken American English (Switchboard, Santa Barbara and the Corpus of Contemporary American English), and developed a generalized mixed-effects model of speakers' choices of remnants with and without prepositions. Such modeling is used to assess the influence of each of several factors interacting simultaneously to produce an outcome. Two elliptical constructions, sluicing and BAE, were included, each with two subtypes. For sluicing, one subtype was what I dub regular sluicing (see (3)) and the other reprise utterances (12), a type of clarificational utterances (see Ginzburg and Sag 2000).

(12) A: Have you ever heard of Yani? B: Of who?

For BAE, one subtype was regular BAE (see (1)), and the other split questions (see Arregi 2010), shown in (13). Split questions differ from BAE in that the question and answer appear within a single speaker's turn.

(13) A: I'll have you arrested. B: By who? The sand police?

I coded the data for the acceptability of preposition stranding in the corresponding nonelliptical clauses (this information was obtained via Amazon's Mechanical Turk), construction type, syntactic category of the correlate (PP or NP), phrasal complexity of the correlate, and semantic dependence between a preposition and verb (or noun or adjective). The last two factors are illustrated below. In (14)-(15), the correlates contain an NP (*audition*) and a which-NP phrase (*which state*), which are more complex phrases (because they contain more semantic and syntactic information, see Hofmeister et al. In press¹) than the correlates in (1) and (3). The examples in (16)-(17) differ in terms of whether the prepositions are dependent on the verbs for interpretation (see Hawkins 2000, 2004). (16) hosts the preposition *for*, which can be interpreted independently of the verb *finish*, but this is not the case with the preposition *of* in (17). English prepositions and verbs show a gradient of such semantic dependencies, which has its origin in the rise of prepositional verbs in the Middle English period (Brinton and Traugott 2005). In coding the data for semantic dependencies, I used speakers' judgments (using Hawkins' (2000) Entailment tests) collected via Amazon's Mechanical Turk and Pointwise Mutual Information (a measure of how likely X is to appear, given the appearance of Y).

(14) A: I'm here for the audition. B: Which audition?

(15) Which state are we talking about? B: Oklahoma.

(16) A: I promised I'd finish for Christmas. B: You just didn't say which Christmas.

(17) What do you think of when you think of marriage? B: Wedding cake.

Selected data were then used as experimental stimuli in a study of speakers' judgments (100-split task) about the naturalness of these data, given the antecedents (also done via Amazon's Mechanical Turk). This methodology, following Bresnan (2007) and Ford and Bresnan (2010), allowed me to verify whether speakers' ratings were significantly correlated with corpus data, and hence, could be explained in the same way.

¹For more information on phrasal complexity and its effects on memory retrieval, see Criss and McClelland 2006, Nairne 1990, 2001, 2006, Oberauer and Lewandowsky 2008, Hofmeister 2009.

Results and discussion

The model of the corpus data, which correctly predicts 92% of the remnants, shows that the availability of preposition stranding makes a nonsignificant contribution to predicting the appearance of prepositionless remnants ($p = 0.204$). This finding confirms what raw frequency data shown in Table 1 also suggest. Table 1 crosses the appearance of prepositionless remnants with the possibility of preposition stranding, and yet the ratio of preposition omission remains largely unaffected.

Possibility of preposition stranding	Remnant with P	Remnant without P	Total
possible	108(31.7%)	233(68.3%)	341(100%)
impossible	27(39.7%)	41(60.3%)	68(100%)

Table 1: Possibility of preposition stranding by realization of ellipsis remnant

According to the model, factors making a significant contribution to predicting the appearance of prepositionless remnants are not syntactic and include:

1. Priming (a correlate containing a preposition is more likely to co-occur with a remnant containing one ($p < 0.001$))
2. Construction type (reprise utterances favor retention of prepositions in remnants ($p < 0.00001$), while split questions favor preposition omission ($p < 0.00001$))
3. Correlate complexity (a complex correlate increases the likelihood of preposition omission ($p < 0.002$))
4. Semantic dependency (pairs of verbs and prepositions that are semantically dependent on each other, that is, noncompositional, increase the likelihood of preposition omission ($p < 0.0001$))

I next compared corpus probabilities for selected remnants from the data with speakers' judgments about the naturalness of these remnants, given their antecedents. Both values were significantly correlated ($p < 0.0005$), which indicates that the same factors can explain both corpus data and speaker judgments. The last two factors (3 and 4) are the most interesting from the point of view of theoretical treatments of ellipsis.

The involvement of correlate complexity strongly suggests that complex correlates are more accessible to the parser, and hence, easier to retrieve. Extensive research on memory retrieval demonstrates that encoding unique semantic and syntactic features in a phrase aids its subsequent retrieval (Criss and McClelland 2006, Nairne 1990, 2001, 2006, Oberauer and Lewandowsky 2008, Hofmeister 2009). And in the case of sluicing, correlates are reaccessed by means of cues that remnants provide to retrieve them, as part of a 'direct-access mechanism' (Martin and McElree In press). I propose that prepositions serve as additional retrieval cues for a remnant's correlate, if such cues are needed, given the current accessibility of the correlate. The mediating effect of complexity on preposition omission in English aligns with results reported for several non-preposition-stranding languages (Vicente 2008, Szczegielniak 2008, Rodgrigues et al. 2009, Caha 2011, Nykiel and Sag 2011).² If this indeed is a cross-linguistic generalization,

²In this research, the focus remains on the complexity of the remnant, not the correlate. Note, however, it must be the complexity of the correlate that plays a role in BAE shown in (i)-(ii), because the remnant remains the same:

- (i) A: Which teacher did you complain to? B: Kate
- (ii) A: Who did you complain to? B: Kate.

an explanation for it may be found not in the syntax of these languages, but in the process of encoding and retrieval of linguistic signs.

One remaining question is of course why preposition omission seems to be more widespread in English than in other languages despite the shared sensitivity to the complexity of the correlate. The results of regression modeling suggest that English is assigned a special status due to having prepositional verbs. Such verbs increase the number of verb-preposition pairs that are noncompositional, providing a boost for preposition omission (and from a purely usage-based perspective, the availability of preposition stranding in English may have helped create more semantically dependent pairs through entrenchment).

While a growing body of research has focused on whether the PSG holds for languages without preposition stranding, none has rigorously explored whether it correctly captures the behavior of English ellipsis. If similarity between ellipsis remnants and nonelliptical clauses in terms of preposition placement is evidence for underlying structure in ellipsis remnants, then the current data provide no such evidence. The data challenge deletion-based accounts of ellipsis (Ross 1969, Hankamer and Sag 1976, Merchant 2001, 2004), while strengthening accounts eschewing deletion (Ginzburg and Sag 2000, Culicover and Jackendoff 2005), which do not posit any syntactic connection between ellipsis remnants and nonelliptical clauses. At the same that this research probes the viability of the PSG, it makes a contribution to studies of syntactic variation by identifying constraints on speakers' choices between two kinds of ellipsis remnants.

Selected references

- Bresnan, J. and M. Ford.** 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1): 168-213. **Brinton, L. and E. C. Traugott.** 2005. *Lexicalization and language change*. Cambridge: Cambridge University Press. **Chung, S., W. Ladusaw, and J. McCloskey.** 1995. Sluicing and logical form. *Natural Language and Linguistic Theory* 3: 239-282. **Culicover, P. and R. Jackendoff.** 2005. *Simpler syntax* New York: Oxford University Press. **Ginzburg, J. and I. A. Sag.** 2000. *Interrogative investigations. The form, meaning and use of English interrogatives*. Stanford, CA: CSLI Publications. [Distributed by U. Chicago Press]. **Hofmeister, P., T. F. Jaeger, I. Arnon, I. A. Sag, and N. Snider** In press. The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes*. **Martin, A. E. and B. McElree.** In press. Direct-access retrieval during sentence comprehension: Evidence from sluicing. *Journal of Memory and Language*. **Merchant, J.** 2001. *The syntax of silence: sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press. **Merchant, J.** 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27: 661-738. **Rodrigues, C., A. Nevins, and L. Vicente.** 2009. Cleaving the interactions between sluicing and preposition stranding. In Wetzels, L., Weijer, J. van der (eds.), *Romance Languages and Linguistic Theory 2006*, 175-198. John Benjamins, Amsterdam.