Title: Language production and online language comprehension behavior

Author: Joanna Nykiel

Citation style: Nykiel Joanna. (2015). Language production and online language comprehension behavior W: D. Gabryś-Barker (red.), "Topics in applied psycholinguistics" (s. 51-63). Katowice : Wydawnictwo Uniwersytetu Śląskiego.

Joanna Nykiel

Language production and online language comprehension behavior*

Abstract: This chapter provides an overview of three different sources of data (corpus data, acceptability judgment data, and eye movement data) regarding a single constraint, semantic dependency between verbs and prepositions, on ellipsis alternation in Norwegian. One of these sources is provided by an eye tracking study of speakers' online comprehension of instances of ellipsis alternation. Using the eye tracking methodology allows us to probe language production and online language comprehension behavior in terms of whether the same constraint is seen operative in both kinds of linguistic behavior.

Keywords: ellipsis alternation, eye movement, language comprehension, language production

1. Introduction

This chapter reviews one instance of syntactic variation from the point of view of language production and language comprehension. It focuses on the eye tracking methodology as a means to explore the extent to which comprehension behavior reflects patterns found in language production. The instance of syntactic variation in question here is what I have elsewhere dubbed ellipsis alternation (Nykiel, 2014a, b; 2015), which refers to the possibility of retaining or omitting prepositions from the elliptical constructions shown in (1)–(3). Example (1) illustrates sluicing, examples (2) and (3) illustrate Bare Argument Ellipsis (BAE),[1]

* I am grateful to Per Henning Uppstad, Philip Hofmeister, Jacob Thaisen, and Rebecca Charboneau. This research was funded by the Research Council of Norway through an Yggdrasil grant.

[1] Example (3) is in fact described in the literature as instantiating split questions, where the same speaker is asking a question and answering it (Arregi, 2010). Thus, split questions differ from those types of BAE that involve two speakers (such as example (2)). However, there is nothing riding on this difference here.

both of which feature stranded phrases (marked in bold) which I will refer to as remnants. The meanings for these remnants are provided by the antecedent clauses. Although sluicing and BAE differ in terms of what phrases serve as remnants (wh-phrases in sluicing and non-wh-phrases in BAE), they both allow the remnants to optionally include prepositions, if the antecedent clauses contain PP correlates for the remnants. The PP correlates are as follows: *with a friend* in (1), *what … about* in (2), and *on what* in (3). In the rest of this paper, I will call the alternation between remnants with prepositions and remnants without prepositions ellipsis alternation.

(1) Kim is in Paris with a friend, but I don't know **with which one/which one**.
(2) A: What's Kim talking about?
B1: **About Eggs**.
B2: **Eggs**.
(3) A: And you're counting on what? **On people going into the booth/ People going into the booth?**

Ellipsis alternation has been found to be constrained by several factors (see Nykiel, 2014a, b; 2015), only one of which is our focus here. This constraint is linked to semantic relationships between prepositions and other lexical categories, of which verbs are of interest to us. The idea is that some prepositions are required by verbs for grammaticality, but some are merely optional, although they are listed in the verbs' lexical entries. The first kind of preposition is exemplified by *on* in *rely on something* and the second kind by *for* in *wait for something*. It would seem that the distinction between adjunct PPs and argument PPs should be able to capture the difference in behavior between these prepositions and the verbs. However, it does not, because the preposition *for* is adjunct-like in character (it is optional) despite being selected for by the verb *wait*, the way arguments are. To avoid this problem, Hawkins (2000; 2004) proposes that combinations of verbs and prepositions be analyzed in terms of whether one can be processed independently of the other. For instance, example (4) from Hawkins (2000, p. 243) shows the verb *die* followed by two PPs, where the first of these PPs depends on the verb for interpretation, but the second does not.

(4) Fatty died $_{PP1}$[of heat] $_{PP2}$[at Lake Rudolph].

To see the dependency, we replace the verb with an appropriate verbal proform (*do something*, in this case), as is (5), and check whether the resulting meanings are entailed by the sentence in (4).

(5) Fatty did something $_{PP1}$[of heat] $_{PP2}$[at Lake Rudolph].

Notice that while *Fatty did something of heat* is semantically odd and not entailed by (4), *Fatty did something at Lake Rudolph* is entailed by (4). These semantic relationships indicate that the first PP is dependent on the verb *die* for its interpretation. We can test the verb's semantic dependence on the two PPs by removing them from (4), which yields (6). Since (6) is entailed by the meaning of the verb *die* in (4), we conclude that this verb is not semantically dependent on either of the two PPs.

(6) Fatty died.

To take another example, consider (7).

(7) Kim came across some toys.

Following the same steps as before, we test the preposition *across* for semantic dependence on the verb *come* and find that it indeed is dependent on it (*Kim did something across some toys* is semantically odd and not entailed by (7)). We next test the verb's dependence on the preposition by removing it from (7) and find that it is dependent on it, because the meaning of *come across* does not entail *Kim came*. These examples show that there are three levels of semantic dependence; these are summarized in Table 1.

Table 1. Levels of semantic dependence

Dependence level	Example
0: No dependence	The verb *die* and the PP *at Lake Rudolph* in (4)
1: One-way dependence	The verb *die* and the PP *of heat* in (4)
2: Two-way dependence	The verb *come* and the PP *across some toys* in (7)

Such semantic dependencies have relevance to the order of postverbal PPs. Their presence favors the order where the semantically dependent PP is immediately adjacent to the verb (which may or may not be semantically dependent on the preposition), that is, the order seen in example (4) (Hawkins, 2000; 2004). This is because immediate adjacency of semantically dependent categories helps the human processor access them simultaneously (and hence immediately construct an appropriate phrasal node) rather than over larger distances.

Beyond postverbal PPs, semantic dependencies affect ellipsis alternation, such that semantically dependent prepositions are the ones typically omitted from remnants. For instance, (8) illustrates a V–P combination, *fall* and *for*, with a two-way dependence (I leave it to the reader to apply the tests) in sluicing. The preposition *for* is unlikely to be left in the remnant, if at all acceptable.

(8) Kim fell for an old trick, but she didn't say **which/?for which**.

Similarly, V–P combinations with a one-way dependence typically feature remnants without prepositions. Example (9) illustrates the above.

(9) A: Kim is working on a project. B: **What project?**

These patterns are robust in data extracted from corpora of spoken American English (Nykiel, 2014a). They suggest that speakers avoid splitting a semantically dependent V–P combination under ellipsis by omitting the preposition from the remnant, given that the verb is also absent from it.

One explanation for why ellipsis alternation is sensitive to semantically dependent V–P combinations is linked to the idea that such combinations form processing domains. Hawkins (2004) proposes that the human processor prefers to access all the elements constituting a processing domain as efficiently as possible, which predicts that what satisfies this preference is immediate adjacency of these elements in nonelliptical clauses. Nykiel (2014a) extends this proposal to elliptical clauses by proposing the following two preferences:

1. If a V–P combination is characterized by semantic dependence (one- or two-way), then it is preferable, all else being equal, to omit the preposition from the remnant, because the processing domain includes the verb, but not the prepositional object present in the remnant (the processing domain here is the V–P combination).
2. If a V–P combination is characterized by no semantic dependence, then it is preferable, all else being equal, to retain the preposition in the remnant, because the processing domain includes the prepositional object present in the remnant (the processing domain here is the phrase headed by the preposition, that is, the preposition and its object).

It is important to point out that the avoidance of splitting semantically dependent combinations is framed in terms of the processing benefits associated with the speaker, and not the hearer/comprehender. In fact, it is sometimes the case that preferences found in language production behavior do not align with patterns found in language comprehension behavior (Konieczny, 2000). At the same time, there is evidence that speakers make use of patterns occurring in their own linguistic output and that of other speakers' when engaged in language comprehension tasks (MacDonald, 1999; 2013; Staub et al., 2006). In this chapter, I focus on ellipsis alternation in Norwegian with the purpose of comparing language production and language comprehension behavior. This chapter is based on research presented in detail in Nykiel (2014c). The data discussed here come

from three sources: a corpus search, an acceptability judgment study, and an eye tracking experiment. The eye tracking methodology provides a particularly useful way of exploring to what extent language production patterns align with language comprehension behavior. I first overview what is known about reading and the advantages of using eye tracking in studies of reading, and next turn to the data.

2. Eye tracking as a means to investigate online language comprehension

Psycholinguistic literature provides plentiful evidence of how language users comprehend written discourse. Monitoring how long they read a text online has long been considered a window into their ability to comprehend that text (Rayner, 1978; Frazier & Rayner, 1982; 1987; Rayner, 1998; Rayner et al., 2001). Eye movement studies, in particular, provide valuable evidence regarding the ease or difficulty of comprehending texts by offering a perspective on moment-to-moment language processes. Eye movements are recorded with an eye tracker interfaced with a PC computer. All experimental sentences are displayed in a single line on a monitor. Viewing is binocular, but only the right eye is typically recorded. Before the beginning of the experiment, participants are instructed to read for comprehension, and to do so at a normal rate. A calibration routine is performed afterward, and re-calibration may be performed in between trials, as needed. A key press is one of the ways to trigger the presentation of the next sentence on the screen. Software is normally provided for implementing experiments and analyzing data.

During reading, the eyes move in jumps (called "saccades") and fixate certain elements of the text in between these saccades. Typically, readers move about eight character spaces per saccade, with each fixation lasting about 200–250 msec (Staub & Rayner, 2007). Fixations, but not saccades, allow readers to register meaningful information picked up from the text. However, while readers are fixating a word, characters to the left and right of the fixation point are also visible to the eye. This so-called perceptual span includes 3–4 characters to the left of fixation and 14–15 characters to the right of it, although readers cannot identify words located further than 5–7 characters to the right of fixation. However, even words that cannot be identified are partially processed at the level of spelling as part of a preview effect: the first few letters of the word are registered (Rayner, 1998). For skilled readers, about 90% of all saccades are progressive (i.e., they move the eyes forward). Regressive saccades occur if fragments of text are difficult to process (e.g., because they are locally ambiguous or unpredictable,

given the context). Thus, processing difficulty can be measured in several ways. Commonly reported are first fixation duration (the time spent fixating a word for the first time and before moving to another part of it or moving off it), first pass time (the time spent fixating a word before leaving it for the first time), go past time (the time spent fixating a word, including the time spent fixating any words to the left of that word, if regressive eye movements have occurred), total fixation duration (the time spent fixating a word, including all fixations on it), the proportion of trials on which a regressive saccade, as opposed to a progressive one, occurs on the first pass through the word, and the proportion of trials on which the word is skipped on the first pass through it. Information about which words are fixated and for how long a period of time provides reliable data about language comprehension.

The experimental data to be reviewed next include data collected in an eye tracking study, in which all the measures listed above were reported.

3. Language production behavior and language comprehension behavior

The data I discuss in this section reveal that patterns found in language production align with patterns found in language comprehension. I begin with corpus data, and then compare them with acceptability judgment data (section 3.2), and with eye movement data (section 3.3).

3.1 Corpus data

These data were extracted from the Norwegian Speech Corpus—the Oslo part (NoTa-Oslo) (Nykiel, 2014c).[2] The total of twenty-five instances of ellipsis alternation were extracted, whose majority (twenty-two instances) featured remnants without prepositions.[3] Some examples are given in (10)–(11).

(10) A: Hva skal vi prate om da?
what shall we talk about then
B: Nei, det vet jeg ikke.
no this know I not

[2] This 900,000-word corpus contains interviews with Norwegian speakers from Oslo and the Oslo area. It is available online at: http://www.tekstlab.uio.no/nota/oslo/english.html.

[3] For more information on how the data were extracted, see Nykiel (2014c).

A: **Cupfinalen?**
the Cup final
"A: What should we talk about then? B: I don't know. A: The Cup final?"

(11) A: Hvilken skole gikk du på?
which school went you to
B: **Ammerud Skole.**
Ammerund School
"A: Which school did you go to? B: Ammerund School."

In only three examples, the remnants appeared with prepositions. These examples are given in (12)–(14) below. They are the really interesting cases, in that they violate the majority pattern, that is, omission of prepositions from remnants, seen in the data.

(12) A: Den går bak Vålerenga kirke.
it goes behind Vålerenga church
B: På hvilken side av den store veien?
on which side of the big road the
A: **På høyresida**.
on right side
"A: It goes behind the Vålerenga church. B: On which side of the big road? A: On the right side."

(13) A: Men er det en film?
but is that a film
B: Ja, ja.
yes, yes
A: Med hvem da? **Med han selv**?
with who then with himself
"A: But is this a film? B: Yes, yes. A: With who then? With himself?"

(14) A: Du dro dit med venner? **Med hvem**?
you went there with friends with who
"A: You went there with friends? With who?"

If we test the examples in (12)–(14) for semantic dependence, what we find is that none of them host semantically dependent combinations, which is expected. Note that in applying the tests, I use the antecedent clauses with in-situ interrogative phrases the way they actually appear, though collapsing them into single sentences for examples (12) and (13). This is to avoid suggesting that remnants may have more structure than is visible, which continues to be a matter of debate (Merchant, 2001; Culicover & Jackendoff, 2005; Nykiel, 2013). In (12), removing the PP, gives *It goes behind the Vålerenga church*,

which is entailed by *It goes behind the Vålerenga church on which side of the big road*. Replacing the verb with a proform gives *It does something on which side of the big road,* which is also entailed. Similarly in (14), *You went there with friends* entails *You went there* and *You did something with friends*. Example (13) hosts a combination of a noun and preposition, thus differing somewhat from the rest of the examples. However, we can test this combination for semantic dependence the same way as before: *Is that a film with who* entails both *Is that a film* and *Is that something with who,* where *something* acts as a nominal proform.

The lesson we learn from this section is that the overall preference in Norwegian is for prepositions to be omitted from remnants. Prepositions that are not omitted from remnants tend to be semantically independent ones. The next section asks whether acceptability judgment data are consistent with corpus data.

3.2 Acceptability judgment data

The participants in this acceptability judgment task were eighty University of Stavanger students. They were asked to provide acceptability ratings (on a 7-point scale) about sentences hosting sluicing remnants with and without prepositions. These sentences were separated out into two groups, one containing semantically dependent V–P combinations (one-way dependence, with the prepositions dependent on the verbs) and the other containing semantically independent V–P combinations. Example (15) illustrates the first group and example (16) the second.[4]

(15) Stig har fortalt om noe, men jeg vet ikke om hva/ hva og jeg er
Stig has told about something but I know not about what/what and I am
likegyldig.
indifferent
"Stig has told (us) about something but I don't know about what/what and I don't care."

(16) Terje har syklet over noe, men jeg vet ikke over hva/ hva og jeg er
Terje has biked over something but I know not over what/what and I am likegyldig.
indifferent
"Terje has biked over something but I don't know over what/what and I don't care."

[4] For more information on the design of this experiment, see Nykiel (2014c).

The results of this study reveal one pattern. Remnants without prepositions are found to be more acceptable than remnants with prepositions for those V–P combinations that show semantic dependence. Otherwise, there is no reliable difference between ratings for remnants with prepositions and remnants without prepositions. Perhaps unexpectedly, given the corpus results discussed in the previous section, no preference for remnants without prepositions was observed across all experimental sentences. The next section is a discussion of what an eye movement study of ellipsis alternation tells us about its comprehension.

3.3 Eye movement data

For the eye movement study, the same experimental sentences were used as those used in the acceptability judgment experiment. A different group of University of Stavanger students participated in this experiment. As is usual in eye movement experiments (see section 2), the participants were asked to read for comprehension and the measures listed in section 2 were all recorded: first fixation duration, first pass time, go past time, total fixation duration, the proportion of trials with regressive eye movements, and the proportion of trials with words skipped.[5]

To collect these measures, each experimental sentence was divided into four regions of interest, as shown in (17). The first region is the antecedent region, the second region is the correlate region, the third region is the remnant region, and the final region is the end region. Notice that the end region is added after the sluicing remnant, which constitutes the critical region here and would appear clause-finally without the end region. The end region is of interest in eye movement studies, because of spillover effects, which do not occur until after the critical region whose position is clause-final. That is, processing difficulty associated with such a clause-final critical region will only be observed when words following that region are being processed.

(17) [Terje har syklet] [over noe], men jeg vet ikke [over hva/hva] [og jeg er likegyldig].

What the data show is that there is no inherent processing difficulty that can be associated with remnants with prepositions compared to remnants without prepositions, or the reverse. This finding is supported by three fixation measures (first fixation duration, first pass time, and total fixation duration), which do not reveal reliable differences between remnants with prepositions and rem-

[5] For more information on this study, see Nykiel (2014c).

nants without prepositions. These measures typically indicate some temporary processing difficulty if longer fixations are observed on particular words or regions, which are not accompanied by regressive eye movements (Staub, 2010). Another finding is that the go past time measure and the proportion of trials with regressive eye movements point to a parallel with both the corpus data and the acceptability judgment data. More regressive eye movements are launched out of the remnant region, and correspondingly, the go past time increases, for remnants with prepositions than for remnants without prepositions, if the relevant V–P combinations exhibit semantic dependence. This is the context in which remnants with prepositions are unexpected from the point of view of language production.

Regressive eye movements are known to reflect processing difficulty which results from violated expectations regarding the incoming material (Staub, 2010). If the incoming material does not fit into the structure constructed on the basis of the input parsed so far, a regressive eye movement will be initiated. Regressive eye movements have been reported in studies of English heavy NP shift constructions and relative clauses (Staub et al., 2006; Staub, 2010). For instance, if something else than the direct object follows a verb with a transitivity bias in a heavy NP shift construction, as in (18), then this postverbal phrase induces increased regressions, but not necessarily longer fixations.

(18) The teacher corrected immediately the unusual answer the student had given. (cf. The teacher corrected the unusual answer the student had given immediately.)

Staub et al. (2006) propose that the postverbal adverbial phrase conflicts with the verb's subcategorization frame, and hence does not fit into the structure built so far. This conflict causes comprehenders to program a regression, possibly in order that they may resolve the conflict by rereading the verb.

With respect to ellipsis alternation, increased regressive eye movements are launched out of remnants with prepositions in exactly the contexts in which prepositions do not typically appear in remnants. It is entirely possible that these prepositions violate comprehenders' expectations about what material should come next. Regressions are thus programmed so that comprehenders can reread the antecedent before proceeding. The pattern of results seen in the eye movement data closely tracks that seen in both the corpus data and the acceptability judgment data. This means that language production behavior and language comprehension behavior are aligned, and further that the speaker's perspective is aligned with the comprehender's in the sense that the comprehender expects to encounter the structure that frequently appears in their own linguistic output and in the output of other speakers. The results reviewed here

thus provide support for the proposal, articulated by MacDonald (1999, 2013), that language production and language comprehension are guided by similar principles.

4. Summary

In this chapter, I have provided an overview of one constraint on ellipsis alternation from the perspective of language production and language comprehension. It is fairly clear that the presence or absence of semantic dependency between V–P combinations does a good job of accounting for why prepositions are omitted from remnants in Norwegian. We have reviewed evidence that this constraint has an impact on ellipsis alternation across three sources of data, showing striking correspondences between sentences that speakers produce and the way they evaluate and comprehend sentences produced by other speakers. The eye tracking methodology, with all the measures of processing difficulty that can be collected, is particularly useful in verifying whether some remnants are inherently harder to process than other remnants or whether they are unexpected.

References

Arregi, K. (2010). Ellipsis in split questions. *Natural Language and Linguistic Theory 28*, 539–592.

Culicover, P., & Jackendoff, R. (2005). *Simpler syntax*. New York: Oxford University Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension. *Cognitive Psychology 14*, 178–210.

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities. *Journal of Memory and Language 26*, 505–526.

Hawkins, J. (2000). The relative order of prepositional phrases in English: Going beyond Manner–Place–Time. *Language Variation and Change 11*, 231–266.

Hawkins, J. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research 29*(6), 627–645.

MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The*

emergence of language (pp. 177–196). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology 4*, Article 226.
Merchant, J. (2001). *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.
Nykiel, J. (2013). Clefts and preposition omission in sluicing. *Lingua 123*, 74–117.
Nykiel, J. (2014a). The ellipsis alternation: Remnants with and without prepositions. Manuscript under review. Retrieved from http://uranos.cto.us.edu.pl/~jnykiel/Projects_and_downloadable_papers_files/paper.pdf.
Nykiel, J. (2014b). Ellipsis alternation in Norwegian: Language production shapes language comprehension. Manuscript under review. Retrieved from http://uranos.cto.us.edu.pl/~jnykiel/Projects_and_downloadable_papers_files/Norwegian.pdf.
Nykiel, J. (2015). Constraints on ellipsis alternation: A view from the history of English. *Language Variation and Change* 27(2), 227–254.
Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin 85*, 618–660.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin 124*, 372–422.
Rayner, K., Binder, K. S., Ashby J., & Pollatsek, A. (2001). Eye movement control in reading. *Vision Research 41*, 943–954.
Staub, A., Clifton, C. Jr., & Frazier, L. (2006). Heavy NP shift is the parser's last resort: Evidence from eye movements. *Journal of Memory and Language 54*(3), 389–406.
Staub, A., 2010. Eye movements and processing difficulty in object relative clauses. *Cognition 116*, 71–86.
Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 327–342). Oxford: Oxford University Press.

Joanna Nykiel

Produkcja języka i rozumienie języka online

Streszczenie

Niniejszy artykuł przedstawia trzy rodzaje studiów nad zjawiskiem *ellipsis alternation* w języku norweskim. Pierwsze z nich dotyczy produkcji języka, a pozostałe dwa jego rozumienia. Autorka opisuje użycie metodologii badania ruchu gałek ocznych, które pozwala na porównanie rezultatów wszystkich trzech studiów pod względem tego, czy preferencje widoczne w produkcji języka pokrywają się z preferencjami widocznymi w rozumieniu języka.

Language production and online language comprehension behavior

Joanna Nykiel

Produktion der Sprache und Verstehen von der *on line*- Sprache

Zusammenfassung

Der vorliegende Beitrag schildert drei Arten der Studien über das Phänomen *ellipsis alternation* in der norwegischen Sprache. Erste Studie betrifft die Produktion der Sprache und die zwei nächsten deren Verstehen. Die Verfasserin beschreibt die Anwendung der Methodologie der Augapfelbewegung, die ermöglicht, die Ergebnisse aller drei Studien in Hinsicht darauf, ob sich die bei der Produktion der Sprache erkennbaren Präferenzen mit den beim Verstehen der Sprachen auftretenden Präferenzen decken, miteinander zu vergleichen.