

Tel Aviv University
Lester and Sally Entin Faculty of the Humanities
Linguistics Department

**Verb-specific lexical information and online sentence processing in Hebrew:
A look at prepositions and filler-gap dependency formation**

MA Thesis Submitted

by

Kayla Gold-Shalev

Prepared under the guidance of

Dr. Aya Meltzer-Asscher

May 2017

Table of contents

Abstract.....	I
1. Introduction.....	1
2. The experiment.....	11
2.1. Method.....	15
2.1.1. Participants.....	15
2.1.2. Materials.....	15
2.1.3. Fillers.....	20
2.1.4. Procedure.....	22
3. Results.....	23
4. Discussion.....	27
4.1. Filled-gap position.....	27
4.2. Main verb.....	28
4.3. Adverb.....	29
5. Conclusions.....	29
5.1. Future research.....	31
References.....	34

Abstract

This paper investigates how two types of verb-specific lexical information affect initial filler-gap dependency formation. A self-paced reading experiment with a filled-gap paradigm was used to investigate whether a mismatch in syntactic category and/or head specificity are ignored when deciding whether or not to form an object gap during the processing of filler-gap dependencies. Grammatical and plausible Hebrew sentences manipulated the filler and the verb in terms of syntactic category and head specificity. Results suggest that the parser does not ignore either mismatch during early filler-gap dependency formation. While still inconclusive, reading times at the main verb may reflect a comparison made between requirements of the verb and the filler that is provided in the sentence, rather than an association between the verb and a filler.

1. Introduction

Understanding language in real time, as it is spoken to us, otherwise known as online sentence processing, is a fast and complex process that requires the quick utilization and implementation of many different types of information. Many studies have investigated how different types of information affect processing. Much of this research has used temporary syntactic ambiguities, or garden path sentences, in order to conduct their investigations (Holmes, Stowe, and Cupples, 1989; Pickering, Traxler, and Crocker, 2000). Another structure that has received attention is filler-gap dependencies. Filler-gap dependencies are syntactic constructions that include a displaced constituent, the filler (*food*, in 1), and an empty syntactic position, the gap (underlined in 1). The filler's thematic role in the sentence is understood through its' relationship with the gap, as illustrated in (1):

- (1) The food Parker ate _ was delicious.

Research has shown that the parser wants to find the gap and close filler-gap dependencies as quickly as possible, positing the existence of a direct-object gap, even before there is conclusive evidence that it exists (Crain and Fodor, 1985; Stowe, 1986; Traxler and Pickering, 1996, among many others). For example, Stowe (1986) found that reading times at the direct-object position (*us*) were slower in (2a) compared to (2b).

- (2) a. My brother wanted to know who Ruth will bring us home to _ at Christmas.
b. My brother wanted to know if Ruth will bring us home to Mom at Christmas.

This fits well with the hypothesis that the parser constructs a direct-object gap as quickly as possible, without verifying that the position is empty. In (2a), the first possible gap site is the direct-object position, and although it is in fact occupied, upon encountering the verb, the parser

assumes it is empty and constructs a direct-object gap with the compatible filler *who* – “Ruth will bring *who*”. One word later, the parser is surprised by the true direct-object (*us*), as reflected by the longer reading times at this position, the so-called "filled-gap" position.

The current paper uses filler-gap dependencies to investigate how two types of information affect initial processing, in particular early filler-gap dependency formation: subcategorization frames and head specificity; two types of lexical information that speakers associate with a verb. Subcategorization frames (Chomsky, 1965; Grimshaw, 1979) outline the syntactic constraints that verbs impose on their complements (internal arguments taken by the verb in addition to the external argument, the subject). For example, while a verb like *bury* takes an NP (noun phrase) (3a), a verb like *insist* takes a CP (clausal complement) (3b) and a verb like *donate* takes an NP and a PP (prepositional phrase) (3c):

- (3)
- a. Parker buried [_{NP} the book].
 - b. Parker insisted [_{CP} that Zoe stay].
 - c. Parker donated [_{NP} the doll] [_{PP} to the hospital].

According to head specificity, our lexical knowledge of verbs may include not only the type of syntactic category that they select for a complement, but also the specific head of that complement. For example, not all English CP-selecting verbs take the same complementizer – some take a *that*-complementizer (4a), while others take an *if*- or a *whether*-complementizer (4b):

- (4)
- a. Parker heard that raspberries are delicious.
 - b. Parker asked if/whether the cake was burnt.

Similarly, different PP-selecting verbs select complements headed by different prepositions. For example, in English, Parker can chat *with* Lily but not *to* or *in* her, and he can complain *about* her but not *on* or *from* her.

A number of studies have tried to determine whether subcategorization information is immediately consulted during the processing of filler-gap dependencies. Tanenhaus, Boland, Garnsey, and Carlson (1989) and Boland, Tanenhaus, Garnsey, and Carlson (1995) investigated whether the number of complements a verb takes affects filler-gap dependency formation in real time. Tanenhaus, Boland, Garnsey, and Carlson (1989) used word-by-word self-paced reading accompanied by a stops-making-sense judgement task to compare filler-gap dependencies with verbs that take one complement (*call*, 5a) or two complements (*remind*, 5b). Filler phrases were either plausible or implausible (italicized) as the direct object to the verbs:

- (5) a. Which customer/*article* did the secretary call _ on the office phone?¹
b. Which child/*movie* did your brother remind _ to watch the show?

While *call*-type verbs provide only one possible gap site, the direct object (call [_]), *remind*-type verbs provide an additional gap site in their second, clausal, complement (remind [someone] [to watch _]).

For *call*-type verbs, reading times were longer and sentences were judged as making sense less often at the verb when the filler was implausible (*article*). For *remind*-type verbs, however, sentences were judged as making sense less often starting at the complementizer (*to*) and reading times were longer one word after that, at the embedded verb (*watch*), namely, after it

¹ Note that this is the only sample material provided in Tanenhaus et al. (1989). Importantly, the implausible *call*-type sentence could have a plausible continuation after the verb, e.g., which *article* did the secretary call her boss about _? Since an exhaustive list of materials is not provided, there is no way of knowing whether all *call*-type verbs suffer from this flaw. If so, the results could suggest that another relevant difference between *call*-type and *remind*-type verbs is that the former can have a second complement while the latter must, and that this may underlie the results.

was already clear that the implausible filler is indeed the direct object. The authors concluded that the argument structure of verbs guides processing by defining the possible gap sites available to the parser.

Boland, Tanenhaus, Garnsey, and Carlson (1995) conducted a similar series of word-by-word self-paced stops-making-sense judgement tasks to compare, once again, verbs that take one (6a) or two complements (6b and 6c). Here too, the filler was either plausible or implausible (italicized) as the direct object to the verbs:

- (6) a. Which teacher/*desk* did the child imitate _ for her bored classmates?
- b. Which sister/*movie* did your mother remind _ to watch the show?
- c. Which note/*bank* did the executive send _ after meeting the deadline?

Again, while *call*-type verbs provide only one possible gap site, the direct object (7a), *remind*-type and *send*-type verbs provide an additional gap site, in their second complement: an embedded infinitive complement for *remind*-type verb (7b) or a PP for *send*-type verbs (7c):

- (7) a. Lucas visited [Emma].
Whom did Lucas visit _?
- b. Zoe reminded [Noah] [to watch the movie]:
Whom did Zoe remind _ to watch the movie?
What did Zoe remind Noah to watch _?
- c. Ethan sent [a letter] [to Olivia].
What did Ethan send _ to Olivia?
Whom did Ethan send a letter to _?

Their main finding was, again, that the position in which the plausibility effect was observed depended on the verb. When the filler was implausible, the plausibility effect was

observed at the direct-object position for *call*-type verbs, but one word later for *remind*- and *send*-type verbs (*to* and *after*, respectively), when it became clear that the implausible filler is indeed the direct object. The authors concluded that the parser uses subcategorization frames when deciding whether or not to construct a direct-object gap: if the filler is implausible as the direct object and the verb provides an additional gap site, as is the case with object-control and dative verbs, the parser puts a hold on direct-object gap construction, in the hopes that the second gap might provide a better semantic fit (e.g., which *movie* did your mother remind [Oliver] [to watch _]? and which *bank* did the executive send [the intern] [to _]?).

Several things should be pointed out with regard to these studies. First, they both use globally implausible sentences, which may lead to unnatural reading and strategic processing. This should be avoided whenever attempting to examine natural, online, sentence processing. Second, they both ask participants to make stops-making-sense judgments. Asking participants to actively decide, at each and every word, whether the sentence still makes sense, could cause reading to become even less natural and make participants overly aware of how they process the sentence. Moreover, in the experiments, deciding that a sentence no longer makes sense caused that trial to terminate. Participants may have been reluctant to do so, and in an attempt to “save” as many sentences as possible, they might have postponed a “no-longer-makes-sense” response, in hopes that the sentence would eventually be resolved. This choice would affect the *remind*- and *send*-type verbs more, since they offer a second gap site. Even more problematic is the fact that the Boland et al. (1995) study relies *only* on these judgement (whereas Tanenhuas et al. (1989) rely also on word-specific reading times).

Third, both studies rely on the assumption that a disruption in reading at the verb reflects an association with an implausible filler. However, they could also reflect the opposite, i.e., that

no such association was made. Under the first assumption, the parser associates the filler with the verb, only to realize the two do not match. Under the second assumption, the parser realizes that the filler does not match the verb, and therefore does not make the association. A disruption at the verb is a reasonable outcome either way: whether the parser realizes its' mistake or is left with an open filler-gap dependency. The problem is that when a conclusion rests solely on a disruption in at the verb, it changes based on how this disruption is interpreted: when taken to indicate that an association has been made with an incompatible filler, the conclusion is that the parser does not take this incompatibility into account, but when taken to indicate that no such association was made, the conclusion is that it does take this incompatibility into account.

The Tanenhaus et al. (1989) study has two conclusions: (a) the parser treats verbs differently based on how many complements they have, and (b) the parser ignores a semantic mismatch (originally referred to as “thematic”). Their first conclusion stays the same regardless of how the authors choose to interpret long reading times at the verb: whether they indicate an association or not, this happens in two different locations, depending on whether the verb offers a second gap site. Their second conclusion, however, changes as soon as long reading times at the verb are interpreted differently. Studies that focus on reading times at the verb are susceptible to such alternative explanations, and as we will see shortly, Wagers and Phillip’s (2014) experiment 3 provides a good example for the second, non-association, interpretation and the conclusion that follows from it.

Another study which manipulated the number of complements a verb takes, reaching the opposite conclusion, is Omaki, Lau, White, Dakan, Apple and Phillips (2015). In a series of self-paced reading and eye-tracking while reading experiments they compared sentences that varied the transitivity of the verb, while maintaining plausibility of the filler:

- (8)
- a. The book that the author wrote regularly about _ was named for an explorer.
 - b. The book that the author {who wrote regularly} saw _ was named for an explorer.
 - c. The book that the author chatted regularly about _ was named for an explorer.
 - d. The book that the author {who chatted regularly} saw _ was named for an explorer.

The experiment relies on *island constraints*, syntactic constructs which constituents cannot move out of and therefore cannot contain gaps (Ross, 1976), and on the finding that the parser does not construct dependencies that cross island boundaries (e.g., Stowe, 1986, Traxler and Pickering 1996). Thus, if the first available gap position is located within an island (marked by { }), the dependency will not be formed. The authors found that reading times at the verb were longer in the non-island intransitive sentence (8c) compared to the island intransitive sentence (8d), but not in the non-island transitive sentence (8a) compared to the island transitive sentence (8b). This is understood if the parser is actively expecting a transitive verb, regardless of the true transitivity status of the verb. The authors suggest the parser creates a direct-object gap before ever coming across the verb, making it impossible for the parser to have access to or consult the verb's transitivity information prior to filler-gap dependency formation.

All the studies described above investigated only one of the two aspects related to subcategorization frames, namely the number of complements a verb has, and manipulated this factor. These studies have reached contradictory results, as well as suffered from methodological problems, in particular making use of globally implausible sentences and basing conclusions on inconclusive disruptions in reading at the verb.

The other aspect of subcategorization frames, which has to do with the syntactic category of the verb's complements, is another interesting topic to research. Pickering and Traxler (2003)

conducted a series of segment-by-segment self-paced reading² and eye-tracking while reading tasks to compare filler-gap dependencies with verbs that usually take a PP, but can also take an NP, (e.g. *worry*, 9a) and verbs that usually take an NP, but can also take a PP (e.g. *kill*, 9b).

Fillers were either plausible or implausible (italicized) as the direct object to the verbs, but all sentences resulted in a globally plausible PP-reading:

- (9) a. That's the *cat/car* that the dog worried about _ after going to the vet.
b. That's the *general/country* that the soldier killed for _ during the war.

Their main finding was that PP-preference verbs seem to act like NP-preference verbs, since they also exhibit longer reading times at the verb when the filler is implausible (*car* and *country*, respectively). The authors conclude that the parser ignores subcategorization preferences and treats both types of verbs as NPs-bias, and just like a soldier cannot kill a *country*, a dog cannot worry a *car*.

van Schijndel, Schuler and Culicover (2014) argue that these results are in fact driven by probabilities, but not only verb-specific probabilities. They suggest that just like the verb in (9a) is more likely to take a PP compared to an NP, verbs in general are more likely to be transitive compared to intransitive, and filler-gap dependencies are more likely to position a gap after a verb (*worry*) compared to a preposition (*about*). When all of these probabilities are taken into account,³ the specific PP-probability of almost all of the verbs used in Pickering and Traxler's (2003) study were overwhelmed by the other two factors, which happened to be much stronger, resulting in a parser that seems to ignore subcategorization preferences. However, given enough verbs with a strong enough PP-preference, the results would have been different.

² Participants were asked to make stops-making-sense judgments in both self-paced reading experiments, but in both cases the judgment data exhibited no main effects or interactions.

³ This conclusion is based on a mathematical calculation of the three probabilities.

Note also, that even if we accept Pickering and Traxler's (2003) conclusion, the generalization only holds for subcategorization preferences/probabilities. Namely, when a verb has multiple subcategorization options, the parser does not make use of probabilistic knowledge of the verb's preferences. This does not mean that the parser ignores categorical (rather than statistical) knowledge of subcategorization frames, namely ignores the fact that some verbs can only select a complement of a certain category.

With regard to head specificity, much less research was carried out. One study that tested whether this information is consulted during filler-gap dependency formation is Wagers and Phillips's (2014) Experiment 3, which used eye-tracking while reading with sentences in which the head of the PP was manipulated, as in (10):

- (10) The courier *to*/**from* whom the secretary warily entrusted _ the confidential business correspondence after some hesitation...

Reading times at the verb (*entrusted*) were longer when the head was a mismatch (**from*) compared to when it was a match (*to*) to the lexical specification of the verb. The authors suggested that the PP with the mismatched preposition was evaluated as an argument of the verb, but then rejected, because it failed to meet the head specificity requirements of the verb. The authors concluded that filler-gap dependency formation is sensitive to subcategorization restrictions that verbs place on the head of their PP complement.

As with Tanenhaus et al. (1989) and Boland et al. (1995), Wagers and Phillips's (2014) experiment 3 also suffers from some shortcomings. First, the authors use sentences that are globally ungrammatical, which could, once again, lead to strategic processing. In addition, the ungrammaticality of all experimental sentences is very easily detected at the same point, i.e. the

verb (*entrusted*), due to the fact that the head of the PP has been pied-piped.⁴ This means that the long reading times obtained at the verb could, at least in part, reflect processes associated with the detection of ungrammaticality or the outcome thereof. Either way, they are not limited to processes pertaining to filler-gap dependency formation.

Second, regarding Wagers and Phillips' (2014) conclusion – that head specificity is taken into account during processing – recall that this disruption at the verb could also be taken as evidence that an association *did* occur with the mismatched filler. This is, after all, how Tanenhaus et al. (1989) and Boland et al. (1995) interpreted the disruption they found at the verb. Under this assumption, Wagers and Phillips (2014) would have to conclude that head specificity requirements *are not* taken into account during processing.

Lastly, even if you were to conclude that the parser is sensitive to head specificity requirements of verbs, English would be unable to provide a full answer as to what exactly the parser is sensitive to, since it only has free prepositions. Other languages have also bound prepositions, which need to be orthographically adjacent to the noun heading their NP complement. These languages could help separate possible lexical and grammatical effects prepositions may have on processing.

The current experiment addresses the potential problems from all of the above studies by using grammatical Hebrew sentences that manipulate the filled-gap effect in a self-paced reading paradigm.

⁴ Even if the head of the PP had been stranded, i.e., *the courier whom the secretary warily entrusted the confidential business correspondence *from after some hesitation*, the sentences would still be globally ungrammatical, but this would be detected only downstream from the verb, at the stranded head.

2. The experiment

The current study investigates whether syntactic category (one aspect of subcategorization frames) and/or head specificity requirements guide parsing, in particular the formation of filler-gap dependencies. To that end, it seeks to improve on previous studies by using plausible, grammatical Hebrew sentences that manipulate the filled-gap effect. The current experiment manipulates this effect to determine whether or not the parser forms a dependency even when the filler is incompatible with the verb in terms of subcategorization or head specificity. Let us see how. In the sample materials given in Table 1, the first possible gap site is always the same main verb (*azra* ‘helped’) and the filled-gap position is always occupied by the same PP-complement (*la-saya’at* ‘to-the teacher’s aide’).

Table 1 | sample materials

<i>if</i> -clause (baseline)	Me’ir lo zoxer im ha-ganenet <i>azra</i> bi-zrizut <i>la-saya’at</i> ha-maxlifa le-hitkašer la-horim etmol ba-boker. teacher’s-aide the-substitute to-call to+the-parents yesterday in+the-morning ‘Me’ir doesn’t remember if the kindergarten teacher helped quickly the substitute teacher’s-aide call the parents yesterday morning.’
filled-gap	Me’ir lo zoxer le-eilu horim ha-ganenet <i>azra</i> bi-zrizut <i>la-saya’at</i> ha-maxlifa le-hitkašer etmol ba-boker. teacher’s-aide the-substitute to-call yesterday in+the-morning ‘Me’ir doesn’t remember which parents the kindergarten teacher helped quickly the substitute teacher’s-aide call yesterday morning.’
Subcategorization	Me’ir lo zoxer eilu horim ha-ganenet <i>azra</i> bi-zrizut <i>la-saya’at</i> ha-maxlifa le-adken etmol ba-boker. teacher’s-aide the-substitute to-update yesterday in+the-morning

	‘Me’ir doesn’t remember which parents the kindergarten teacher helped quickly the substitute teacher’s-aide update yesterday morning.’		
head specificity	Me’ir lo zoxer	be-eilu horim ha-ganenet	<i>azra</i> bi-zrizut <i>la-</i>
	Me’ir not remember in-which parents the-kindergarten-teacher helped quickly to+the- <i>saya’at</i> ha-maxlifa linzof etmol ba-boker. teacher’s-aide the-substitute reprimand yesterday in+the-morning		
	‘Me’ir doesn’t remember which parents the kindergarten teacher helped quickly the substitute teacher’s-aide reprimand yesterday morning.’		

Since the current sentences are all globally grammatical and plausible, conclusions made based on their reading times cannot be said to reflect strategic processing or identification of ungrammaticality. Furthermore, by focusing on reading times at the filled-gap position rather than the verb, we hope to avoid the alternative explanations available in previous studies since reading times at the filled-gap position provide reliable (even if indirect) evidence as to whether or not the parser formed a filler-gap dependency with incompatible fillers.

Let us now review the sentences and predictions. The filler in each sentence is ultimately determined by the embedded verb: *le-hitkašer* (‘to call’) in the filled-gap condition selects for a PP headed by the bound preposition *le-* (‘to/for’); *le-adken* (‘to update’) in the subcategorization condition selects for an NP; and *linzof* (‘to reprimand’) in the head specificity condition selects for a PP headed by the bound preposition *be-* (‘in’). Like the embedded verb in the filled-gap condition, the main verb, *azra* (‘helped’), also selects for a PP headed by the bound preposition *le-* (‘to/for’), making this first possible gap site incompatible with two of the fillers: the NP *eilu* (‘which’) and the PP *be-eilu* (‘in-which’) in the subcategorization and head specificity conditions, respectively. In all sentences a gap position that matches the filler and renders the sentence grammatical is ultimately provided, at the embedded verb. The question is – do reading

times at the filled-gap position indicate that the parser formed a dependency with the incompatible fillers?

Generally speaking, long reading times at the filled-gap position indicate that an association was made with the filler, and short reading times indicate that it was not. The *if*-clause provides a baseline which does not contain a dependency, and is therefore expected to display short reading times at its' equivalent of the filled-gap position. The filled-gap condition, in contrast, is expected to demonstrate a classic filled-gap effect: the PP-filler *le-eilu* ('to which') is compatible with the main verb in subcategorization (*azra* takes a PP) and head specificity (*azra* selects for the preposition *le-* 'to/for'). Upon encountering the verb, the parser is expected to form a dependency with the filler, only to be surprised by the true PP-complement, resulting in longer reading times at the filled-gap position.

The remaining conditions are more interesting and hopefully more revealing, due to the fact that they contain the *incompatible* fillers we are so interested in. In the subcategorization condition there is an NP-filler, *eilu* ('which'). If the parser does not ignore a mismatch in syntactic category when deciding to form a dependency, we expect short reading times at the filled-gap position: if the parser only forms dependencies with PP-fillers, as per the verb's subcategorization requirements, no dependency will be formed with the NP-filler, and the actual PP-complement will be integrated with ease. However, if the parser does ignore a mismatch in syntactic category when making this decision, we expect long reading times at the filled-gap position: if the parser forms dependencies with any filler, regardless of syntactic category, a dependency will be formed with the NP-filler and the parser will be stumped by the actual PP-complement, causing long reading times at the filled-gap position.

The same rationale lies behind the head specificity condition. If reading times at the filled-gap position are short, this will indicate that the parser does not ignore a mismatch in head specificity when forming an object gap: the PP-filler is headed by an incompatible preposition (*be- 'in'*) and therefore blocks the formation of the dependency, allowing for the integration of the PP-complement headed by a compatible preposition. However, if reading times at the filled-gap position are long, this will indicate that the parser does ignore a mismatch in head specificity: the PP-filler takes the place of the true PP-complement, despite being headed by an incompatible preposition, and this mistake is revealed at the filled-gap position.

Note that, if results indicate the parser does not ignore a certain mismatch, we will be able to say with a great deal of certainty that this type of information is taken into account in early processing, but if they indicate the parser ignores this mismatch, we will not be able to say with any certainty that it does not take this information into account – you can take your friend's nut allergy into account but decide to ignore it and make pecan pie anyway. Also, regarding short reading times at the filled-gap position, a clarification is necessary. What they really tell us is that by the time we encounter the true PP-complement, there is no association. Before that time, it is equally possible that a dependency was never formed or that it was, but it reopened very quickly. Current behavioral measures, such as reading times, cannot pinpoint exactly when a specific type of information affects processing: did it affect processing from the beginning, making sure that the dependency was never formed with the incompatible filler, or slightly after, causing the parser to reopen the dependency it has just closed? Either way, what we can say is that the specific type of information *was* taken into account, because the parser didn't need to arrive at the actual PP-complement to decide that the dependency is a mistake.

A summary of the predictions regarding incompatible fillers is provided in Table 2:

Table 2 | predictions for incompatible fillers and reading times at the filled-gap position

Long Reading Times	→	The parser ignores the mismatch in syntactic category and/or head specificity and forms the dependency despite it; the parser might still take syntactic category and/or head specificity into account during initial filler-gap dependency formation.
Short Reading Times	→	The parser does not ignore the mismatch in syntactic category and/or head specificity and therefore does not form the dependency; the parser does take syntactic category and/or head specificity into account during initial filler-gap dependency formation.

We will also examine reading times at the main verb and the adverb. Rather than reflecting an association, or the lack thereof, between the verb and a filler, the current paper suggests that reading times at the main verb may reflect a comparison drawn between the characteristics of the filler as required by the verb and the one provided in the sentence.

2.1. Method

2.1.1. Participants

Forty-four native Hebrew speakers were recruited from Tel-Aviv University. Participants received course credit or were paid twenty Shekels (approximately \$5) for participating.

Participants included thirty-three female and eleven male speakers, with ages ranging from 20-36 and an average of 24.36.

2.1.2. Materials

Thirty-two sets of Hebrew sentences were prepared. Each set contains an embedded verb that takes two complements, a PP headed by either *le-* ('to/for') or *al* ('on/about') and a CP, such as *amar le-mišehu la'asot mašehu* ('told someone to do something') or *pakad al mišehu la'asot*

mašehu (‘ordered someone to do something’). Each set contains one baseline condition (*if*-clause) and three critical conditions that contain a filler-gap dependency (filled-gap, subcategorization and head-specificity), differentiated by the type of filler: PP-filler headed by a compatible preposition in the Filled-Gap condition, NP-filler in the Subcategorization condition, and PP-filler headed by an incompatible preposition in the Head-specificity condition, as demonstrated in Table 1, repeated here as Table 3:

Table 3 | sample materials

<i>if</i> -clause (baseline)	Me’ir lo zoxer im ha-ganenet <i>azra</i> bi-zrizut <i>la-</i> Me’ir not remember if the-kindergarten-teacher helped quickly to+the- <i>saya’at</i> ha-maxlifa le-hitkašer la-horim etmol ba-boker. teacher’s-aide the-substitute to-call to+the-parents yesterday in+the-morning ‘Me’ir doesn’t remember if the kindergarten teacher helped quickly the substitute teacher’s-aide call the parents yesterday morning.’
filled-gap	Me’ir lo zoxer le-eilu horim ha-ganenet <i>azra</i> bi-zrizut <i>la-</i> Me’ir not remember to-which parents the-kindergarten-teacher helped quickly to+the- <i>saya’at</i> ha-maxlifa le-hitkašer etmol ba-boker. teacher’s-aide the-substitute to-call yesterday in+the-morning ‘Me’ir doesn’t remember which parents the kindergarten teacher helped quickly the substitute teacher’s-aide call yesterday morning.’
Subcategorization	Me’ir lo zoxer eilu horim ha-ganenet <i>azra</i> bi-zrizut <i>la-</i> Me’ir not remember which parents the-kindergarten-teacher helped quickly to+the- <i>saya’at</i> ha-maxlifa le-adken etmol ba-boker. teacher’s-aide the-substitute to-update yesterday in+the-morning ‘Me’ir doesn’t remember which parents the kindergarten teacher helped quickly the substitute teacher’s-aide update yesterday morning.’
head specificity	Me’ir lo zoxer be-eilu horim ha-ganenet <i>azra</i> bi-zrizut <i>la-</i> Me’ir not remember in-which parents the-kindergarten-teacher helped quickly to+the- <i>saya’at</i> ha-maxlifa linzof etmol ba-boker. teacher’s-aide the-substitute reprimand yesterday in+the-morning ‘Me’ir doesn’t remember which parents the kindergarten teacher helped

quickly the substitute teacher's-aide reprimand yesterday morning.'

In order to prevent a possible spill-over in reading times from the main verb (*azra* 'helped') and allow for an accurate examination of reading times at the filled-gap position, an adverb (*bi-zrizut* 'quickly') was inserted between them.

Embedded verbs were selected by the author, with the help of several native speakers. They were matched between conditions for number of letters (embedded *le-/al* verbs [baseline and filled-gap conditions] $M = 5.188$, $SD = .738$; embedded transitive verbs [subcategorization condition] $M = 5.188$, $SD = .82$; embedded *be-/im* verbs [head specificity condition] $M = 5.5$, $SD = 1.016$, $F(2, 93) = 1.389$, $p = .254$).

Two potential problems needed to be addressed when preparing the materials. The first has to do with the type of prepositions taken by different verbs, and the second has to do with the structure of the subcategorization condition.

First, Hebrew has two types of prepositions: free (*al* 'on/about', *im* 'with' etc.) and bound (*be-* 'in', *le-* 'to/for' etc.), i.e., that need to be orthographically adjacent to the noun heading their complement NP. Recall that since English has only free prepositions, Wagers and Phillips' (2014) were unable to explore whether this difference also affects processing. Hebrew allows us to separate these possible effects. But, let's imagine that all of the embedded verbs in the current experiment select for bound prepositions (e.g., *daxak be-* 'urged someone') and that the head specificity condition only featured PP-fillers that are headed by a free preposition. Now let's assume that short reading times were obtained at the filled-gap position in this condition. One might have been very happy to conclude that Hebrew never ignores a mismatch in head specificity when forming filler-gap dependencies, but this might be the result of the stark difference between the observed *free* preposition and the expected *bound* preposition. All

possible combinations should be examined in order to come to an accurate conclusion. In our example, we examine what happens (a) when the observed preposition is a different bound preposition (e.g., *me-* ‘from’) and (b) when the expected preposition is free. If reading times remain short, this would be a stronger indication that head specificity affects filler-gap dependency formation, since any deviation from the expected preposition, whether lexical or pertaining to the free/bound distinction, is important enough to block object gap formation. To that end, both types of verbs and prepositions were featured equally: half of the embedded verbs in the experiment take a bound preposition (*le-* ‘to/for’) and half take a free preposition (*al* ‘on/about’), and for each verb type, half of the sentences in the head specificity condition feature a different preposition of the same status and half feature a preposition from the other status (*be-* and *im*).

Second, one should note the possibility that the dependency could be resolved at the main verb in the subcategorization condition, if readers expect a post-verbal PP containing a resumptive pronoun (e.g., *alav* ‘on-him’). If they do, we may find longer reading at the true complement, but not for the reasons being explored in this experiment, i.e., not because the parser ignored the mismatch in syntactic category, but because it was surprised not to find the resumptive that was supposed to resolve the inconsistency, as in the following sentence:

- (11) Me’ir lo zoxer eilu horim ha-ganenet azra *la-hem* lirkod.
 Me’ir not remember which parents the-kindergarten-teacher helped to-them to-dance
 ‘Me’ir doesn’t remember which parents the kindergarten teacher helped dance.’

In general, it is agreed that resumptive pronouns are ungrammatical in Hebrew wh-questions (Sells, 1984). Nonetheless, wh-questions with resumptives do occasionally occur in natural speech. To explore whether speakers do expect sentences such as (11), an acceptability

judgments pretest was conducted. Twenty sets of sentences were prepared,⁵ each containing two sentences differentiated by whether the preposition is pied-piped (12a) or stranded and affixed to a resumptive pronoun (12b):

- (12) a. ha-katav yada *im* eizo ozeret ha-sar soxax bi-zman ha-pgiša
 the-reporter knew with which assistant the-minister spoke during the-meeting
 ‘The reporter knew which assistant the minister spoke with during the meeting.’
- b. ha-katav yada eizo ozeret ha-sar soxax *ima* bi-zman ha-pgiša.
 the-reporter knew which assistant the-minister spoke with-her during the-meeting
 ‘The reporter knew which assistant the minister spoke with her during the meeting.’

Four free and bound prepositions were distributed evenly: *al* (‘on/about’), *im* (‘with’), *be-* (‘in’), *le-* (‘to/for’), insuring that judgments would have everything to do with the structure of these sentences and nothing to do with the prepositions. Twenty filler sentences of two types were also prepared: sentences with transitive verbs that take an NP (13a) and intransitive verbs with a PP adjunct (13b):

- (13) a. ha-safranit ha-šketa atfa ‘et ha-xavila ha-gdola be-niyar atifa adom.
 the-librarian the-quiet wrapped ACC the-package the-big in-paper wrapping red
 ‘The quiet librarian wrapped the big package in red wrapping paper.’
- b. ha-yalda ha-amitca rakda ba-salon likrat bxinot ha-kabala
 the-girl the-brave danced in+the-living-room in-preparation-for tests the-admission
 ‘The brave girl danced in the living room in preparation for the auditions.’

A sentence acceptability questionnaire was constructed and distributed via Qualtrics: Online Survey Software and Insight Platform. Two lists were created, each containing twenty critical items, ten from each condition (a-b), and twenty fillers, ten from each type (i-ii), for a total of forty sentences. Questions were randomized for each participant. Twenty-eight

⁵ Of these 20 sets, one was excluded from analysis due to a typo.

participants completed the questionnaire (sixteen female, twelve male; ages ranging from 18-45, with an average of 30.14). Participants were asked to rate the sentences on a five-point scale, ranging from “sounds very bad” (1) to “sounds very good” (5). A by-participants Wilcoxon revealed a very significant main effect of PP position (pre-verbal median = 3.75, SD = 0.872, post-verbal median = 2, SD = 0.757, $Z = -4.467$, $p < .001$). A by-items Wilcoxon analysis revealed the same strong effect (pre-verbal median = 3.5, SD = 0.647, post-verbal median = 2, SD = 0.268, $Z = -3.855$, $p < .001$). In both cases, post-verbal PPs rendered sentences less acceptable than their pre-verbal counterparts.

Long reading times at the filled-gap position in the head specificity condition can therefore be interpreted in the intended spirit of the paper: that the parser ignored the mismatch in syntactic category and formed a dependency with the incompatible NP-filler *eilu* (‘which’).

2.1.3. Fillers

Twenty-four filler sentences were also prepared, to counterbalance any strategic processing the critical filler-gap sentences may trigger: in the experimental items, every time a filler appears, it turns out to be relevant to the second (embedded) verb and irrelevant to the first (main) verb. Participants could become aware of this, and cautiously decide not to form a dependency with the incompatible filler, even if this were otherwise their natural tendency. They could even decide not to form the dependency when the filler *is* compatible with the main verb, hindering our ability to replicate even the classic filled-gap effect in the filled-gap condition. In an attempt to eliminate these unnatural processing strategies, three types of filler sentences were prepared, as demonstrated in Table 4:

Table 4 | sample fillers

filled-gap filler	<p>Dror lo yod'e le-eilu našim ha-šerif ha-mekomi da'ag le'axar</p> <p>Dror not know for-which women the-sheriff the-local was-concerned after</p> <p>txilat ha-hafgana ha-hamonit.</p> <p>the-start the-demonstration the-massive</p> <p>'Dror doesn't know for which women the sheriff was concerned after the massive demonstration had begun.'</p>
Subcategorization filler	<p>Omer lo batu'ax eilu mehandesim ha-iria ha-ani'ya herši'a</p> <p>Omer not sure which engineers the-city-hall the-poor convicted</p> <p>le'axar krisat ha-binyan.</p> <p>after collapse the-building</p> <p>'Omer doesn't know which engineers the poor municipality convicted after the building collapsed.'</p>
head specificity filler	<p>Ronen lo yode'a im eilu roxvim ha-me'alefet ha-mehulelet flirteta</p> <p>Ronen not know with which riders the-trainer the-celebrated flirted</p> <p>ba-šnata'im ha-axronot.</p> <p>in+the-two-years the-last</p> <p>'Ronen doesn't know with which riders the celebrated trainer has flirted over the past two years.'</p>

Each type of filler mirrors a potentially problematic critical sentence (excluding *if* sentences) and will hopefully neutralize it by appearing just as often: in the filled-gap filler there is a PP-filler headed by *le-* ('to/for') or *al* ('on/about') that is relevant to the first verb; in the subcategorization filler there is an NP-filler that is relevant to the first verb; and in the head specificity filler there is a PP-filler headed by *be-* ('in') or *im* ('with') that is relevant to the first verb. As with the critical materials, subcategorization and head specificity fillers featured free and bound prepositions equally (*le-* 'to/for' and *al* 'on/about', and *be-* 'in' and *im* 'with', respectively).

A Latin square design was used to create four lists. Each list contained a total of fifty-six sentences: thirty-two experimental sentences, eight from each condition (a-d), and twenty-four filler sentences, eight from each type (a-c). Items were randomized automatically for each participant.

2.1.4. Procedure

A self-paced reading paradigm (Just, Carpenter and Woolley, 1982) was used to measure and compare the reading times of grammatical sentences at a true complement. The task was constructed and implemented via Linger (<http://tedlab.mit.edu/~dr/Linger/>). A word-by-word, moving window presentation was chosen. Thus, at the start of each trial, a sentence appeared as a series of dashes, such that each cluster of dashes corresponded to a specific word. After pressing the designated key (space bar) for the first time, the first cluster of dashes was replaced by the corresponding word. Each space bar press thereafter caused the previously presented word to once again be replaced by dashes, while the next word was revealed. To ensure that the participants remain focused throughout the experiment, they were required to answer a simple yes/no question after each sentence. Questions never pointed to the fact that the filler is relevant only to the embedded verb. Rather, they varied: some related to a male-female distinction relating to the sentence, some asked about a noun and its' adjective or the subject of the sentence, and others required participants to infer something from the sentence. This variation would, hopefully, require participants to process the sentences, rather than skim through them in search of an answer to a very simple, known in advance, question. Feedback regarding the accuracy of these answers was not provided.

Prior to the experiment, participants were instructed to read at a natural pace and to answer the questions as accurately as possible. Four practice items were provided, and the order of

presentation was randomized for each participant. The experiment took about twenty minutes to complete. The experiment protocol for this study was approved by the Tel-Aviv University's Ethics Committee.

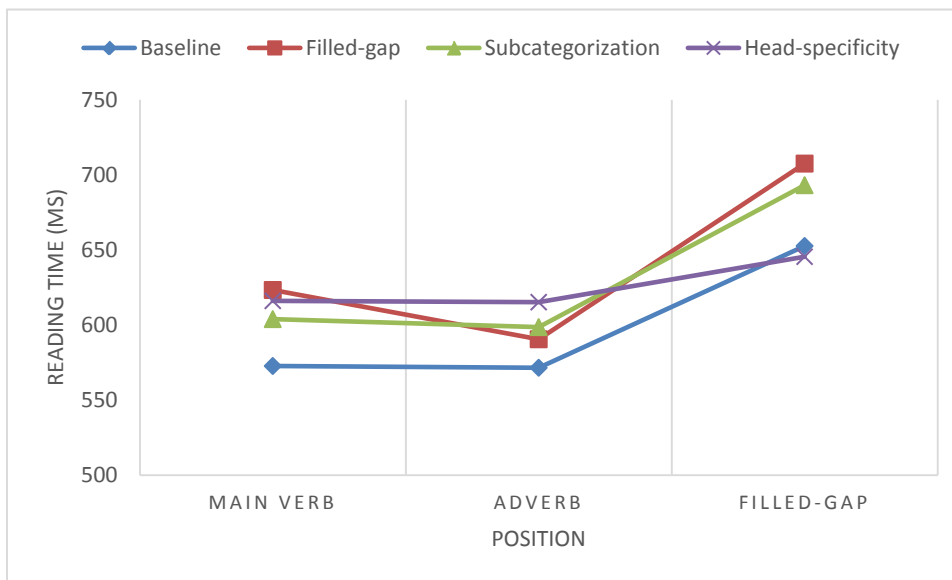
3. Results

The mean by-participants reading times (and standard deviations) for the main verb, the adverb and the filled-gap positions are provided in Table 5 and Figure 1.

Table 5 | by-participants means and SDs

Condition	Main verb position	Adverb position	Filled-gap position
Baseline	572.747 (186.161)	571.519 (180.146)	652.684 (249.274)
Filled-gap	623.289 (202.082)	590.475 (194.840)	707.522 (335.622)
Subcategorization	603.906 (183.177)	598.590 (214.690)	693.076 (289.951)
Head-specificity	616.062 (192.136)	615.235 (161.918)	645.491 (234.982)

Figure 1 | by-participants means



Reading times at three positions were subjected to a by-participants and a by-items repeated-measured ANOVA, with the within-subject factor “sentence type”. At the filled-gap position, both repeated measures ANOVAs revealed a significant main effect (by-participants: $F(3, 129) = 3.661, p = 0.014$, by-items: $F(3, 93) = 2.896, p = 0.039$). Pairwise comparisons revealed significantly longer reading times in the filled-gap condition compared to the baseline (by-participants: $F(1, 43) = 4.742, p = 0.035$; by-items: $F(1, 31) = 5.524, p = 0.025$), but not in the subcategorization or head-specificity conditions (by-participants: $F(1, 43) = 3.438, p = 0.071$; $F(1, 43) = 0.164, p = 0.688$; by-items: $F(1, 31) = 2.077, p = 0.160$; $F(1, 31) = 0.066, p = 0.799$).

At the main verb position, both repeated measures ANOVAs revealed no significant effects (by-participants: $F(3, 129) = 1.708, p = 0.169$; by-items: $F(3, 93) = 2.060, p = 0.111$). Pairwise comparisons revealed significantly longer reading times in the filled-gap and head-specificity conditions compared to the baseline (by-participants: filled-gap: $F(1, 43) = 5.135, p = 0.029$; head-specificity: $F(1, 43) = 4.755, p = 0.035$; by-items: filled gap: $F(1, 31) = 4.745, p = 0.037$; head specificity: $F(1, 31) = 4.977, p = 0.033$), but not in the subcategorization condition (by-participants: $F(1, 43) = 1.364, p = 0.249$; by-items: $F(1, 31) = 1.845, p = 0.184$).

At the adverb position, the repeated measures ANOVAs revealed no significant effects (by-participants: $F(3, 129) = 1.353, p = 0.260$; by-items: $F(3, 93) = 1.264, p = 0.291$). Pairwise comparisons revealed longer reading times in the head-specificity condition compared to the baseline in the by-items analysis, and a similar trend in the by-participants analysis (by-participants: $F(1, 43) = 3.852, p = 0.056$; by-items: $F(1, 31) = 5.386, p = 0.027$), but not in the filled-gap or subcategorization conditions (by-participants: $F(1, 43) = 0.779, p = 0.382$; $F(1, 43) = 1.278, p = 0.265$; by-items: $F(1, 31) = 0.662, p = 0.422$; $F(1, 31) = 1.361, p = 0.252$).

To further investigate the difference that was found between the filled-gap condition and the baseline at the filled-gap position, reading times were divided based on preposition type: bound preposition *le-* ('to/for') and free preposition *al* ('on/about'):

Table 6 | means and SDs: filled-gap condition, preposition type

Condition	Filled-gap position
Baseline <i>le</i>	599.363 (269.883)
Filled-gap <i>le</i>	637.409 (328.757)
Baseline <i>al</i>	706.005 (323.660)
Filled-gap <i>al</i>	777.636 (396.420)

Repeated measures ANOVAs revealed no interaction between sentence type and preposition (by participants: $F(1, 43) = 0.172, p = 0.681$, by-items: $F(1, 15) = 0.419, p = 0.527$). Paired-sample t-tests revealed significantly longer reading times in the filled-gap *al* condition compared to the baseline in the by-items analysis (by-participants: $t(43) = 1.348, p = 0.183$, by-items: $t(31) = 2.630, p = 0.013$), but not in the filled-gap *le-* condition (by-participants: $t(43) = 0.909, p = 0.367$, by-items: $t(31) = 0.966, p = 0.341$).

Similarly, it would be prudent to investigate whether the overall lack of significant difference in reading times between the subcategorization condition and the baseline at the filled-gap position is equally driven by both types of preposition. Modern Hebrew-speakers sometimes drop prepositions when they produce relative clauses (Cole, 1976; for findings from children see Friedmann, Aram, and Novogrodsky, 2011) and this seems to extend also to the production of wh-questions. Given the nature of prepositions, this may be done more easily, and therefore more often, with free prepositions. If participants treat the NP *eilu* ('which') like a PP headed by a dropped free-preposition, such as *al eilu* ('on/about which'), we may find longer reading times at the filled-gap position in the subcategorization condition compared to the baseline when the

preposition is free, but not when it is bound. This would cancel out any overall main effect and explain the final results.

Table 7 | means and SDs: subcategorization condition, preposition type

Condition	Filled-gap position
Baseline <i>le-</i>	599.363 (269.883)
Subcategorization <i>le-</i>	614.295 (277.800)
Baseline <i>al</i>	706.005 (323.660)
Subcategorization <i>al</i>	771.857 (352.966)

Repeated measures ANOVAs revealed no interaction between sentence type and preposition (by participants: $F(1, 43) = 0.680, p = 0.414$, by-items: $F(1, 15) = 1.387, p = 0.257$). Paired-sample t-tests revealed no significant difference in reading times between the subcategorization *le-* condition and the baseline (by-participants: $t(43) = 0.409, p = 0.684$, by-items: $t(31) = 0.389, p = 0.699$) or the subcategorization *al* condition and the baseline (by-participants: $t(43) = 1.685, p = 0.099$, by-items: $t(31) = 1.578, p = 0.124$).

It is also interesting to further investigate the difference that was not found between the head-specificity condition and the baseline at the filled-gap position. Recall that the head-specificity condition is made up of two sub-conditions: *same status*, in which both the expected and observed prepositions are either free or bound, and *different status*, in which one is free and the other bound. There is a possibility that one of these sub-conditions is significantly different from the baseline, despite the lack of an overall difference.

Table 8 | means and SDs: *head-specificity condition, preposition status*

Condition	Filled-gap position
Baseline <i>same status</i>	623.431 (266.982)
Head-specificity <i>same status</i>	673.596 (293.575)
Baseline <i>different status</i>	681.937 (277.129)
Head-specificity <i>different status</i>	617.386 (214.28)

Repeated measures ANOVAs revealed a significant interaction between sentence type and preposition status in the by-participants analysis (by participants: $F(1, 43) = 5.951, p = 0.019$, by-items: $F(1, 15) = 3.924, p = 0.066$). Paired-sample t-tests revealed significantly shorter reading times in the head-specificity *different status* sub-condition compared to the baseline in the by-participants analysis (by-participants: $t(43) = 2.065, p = 0.044$, by-items: $t(31) = 1.754, p = 0.089$), but not in the head specificity *same status* sub-condition (by-participants: $t(43) = 1.817, p = 0.076$, by-items: $t(31) = 1.524, p = 0.137$).

4. Discussion

4.1. Filled-gap position

Based on what is already known from the classic filled-gap effect (Crain and Fodor, 1985; Stowe, 1986), the long reading times at the filled-gap position in the filled-gap condition indicate that a dependency was formed with the filler, and the short reading times found in the subcategorization and head-specificity conditions indicate it was not.

The lack of interaction between sentence type (baseline, filled-gap condition) and preposition (*le-*, *al*) means that the significantly longer reading times found at the filled-gap position in the filled-gap condition compared to the baseline are equally driven by free and bound prepositions. In other words, participants form a filler-gap dependency with both types of

prepositions and are equally surprised when either turns out to be incorrect. This is true despite the fact that the by-items t-test revealed significantly longer reading times at the filled-gap position in in the *al* filled-gap sub-condition compared to the baseline. Going forward, I believe this difference may turn out to be inaccurate, in so much that it reflects the behavior of specific participants and not how the population as a whole the processes sentences and prepositions.

The fact that neither type of preposition causes a significant increase in reading times at the filled-gap position in the subcategorization condition compared to the baseline could, along with the fact that no interaction was found between sentence type (baseline, subcategorization condition) and preposition (*le-*, *al*), could indicate that speakers do not assume an NP is a PP headed by a dropped preposition, despite the fact that they may drop prepositions during speech. To test this hypothesis and answer whether prepositions are treated differently in production and processing (suggesting that preposition dropping has not been fully adopted in Hebrew), we would need to determine whether current participants produce sentences via preposition-dropping, perhaps with a post-test in the form of a sentence-completion questionnaire.

Lastly, the finding that the head-specificity *different status* condition has significantly *shorter* reading times than the *different status* baseline at the filled-gap position is unexpected and quite puzzling. If replicated in later studies, an explanation will be required. Note that the interaction found between sentence type (baseline, head-specificity condition) and preposition status (same, different) cancels out the above significant difference, explaining the lack of an overall difference between the head-specificity condition and the baseline.

4.2. Main verb

Results from the filled-gap position separate our three conditions into two camps: filled-gap, in which a dependency is formed vs. subcategorization and head-specificity, in which it is not.

Reading times at the verb, however, lead to a different division: filled-gap and head-specificity vs. subcategorization. I suggest that what the filled-gap and head-specificity conditions have in common – and is possibly reflected in reading times at the verb – is the nature of the comparison drawn between the filler as required by the verbs in these conditions and the filler that is provided in them: in both, the parser is required to compare within a specific syntactic category, i.e., to take into account individual prepositions. In contrast, in the subcategorization condition the parser is required to compare across syntactic categories, i.e., PP vs. NP.

4.3. Adverb

The fact that long reading times at the adverb were found only in the head-specificity condition can support conclusions drawn thus far. This result does not seem to reflect the fact that the comparison between the filler and verb is difficult in this condition, because an equally difficult comparison is also required in the filled-gap condition. Similarly, it does not seem to reflect the fact that the dependency is left open in this condition, because it is left open in the subcategorization condition as well. What is unique to head-specificity is that the parser has to perform a difficult comparison and also does not form a dependency. In other words, the parser puts a lot of effort into a hard comparison that fails to pan out, and it is this extra strain that carries on to the post-verbal adverb.

5. Conclusions

The present study set out to investigate two types of verb-specific lexical information and how they affect early filler-gap dependency formation. Another equally important goal was to explore a new way of investigating this issue without relying on inconclusive reading times obtained at the main verb.

First and foremost, the results show that the parser does not ignore a mismatch in syntactic category and head-specificity; it takes these types of information into account during early filler-gap dependency formation. The simplest conclusion would be to assume that no dependency was ever formed with the incompatible fillers in these conditions, but, as explained, there is currently no way to confirm this. Thus, we cannot rule out that a dependency was closed and very quickly reopened, prior to the filled-gap position. Either way, results do still indicate that these incompatibilities are important enough to be taken into account before the filled-gap position, and significant enough to make sure that no dependency exists when the true-complement appears.

Importantly, had the current paper relied solely on reading times at the main verb, a very different picture would have emerged from the data. As explained, reading times at the filled-gap position suggest that the head-specificity and subcategorization conditions pattern the same, while reading times at the main verb suggest that the head-specificity and filled-gap conditions pattern the same. Digging deeper, the interpretation of reading times at the main verb are, as previously demonstrated, open to alternative explanations. Based on reading times at the verb, author *a* could conclude that a dependency was formed in the filled-gap and head-specificity conditions, but not in the subcategorization condition. S/he would then have to conclude that head-specificity requirements are ignored in early sentence processing, but not subcategorization information. Unfortunately, author *b* could just as easily conclude that a dependency was not formed in the filled-gap or head-specificity conditions, and that it was formed in the subcategorization condition. S/he would then have to conclude that head-specificity requirements are not ignored in early sentence processing, but subcategorization information is – an odd conclusion in and of itself. By basing conclusions on the interpretation of the filled-gap effect,

i.e., reading times at the filled-gap position, we have gained reliable indirect evidence that (a) a dependency was formed in the filled-gap condition and (b) that subcategorization and head-specificity are in fact the conditions that pattern the same, such that neither is ignored during early filler-gap dependency formation.

Current results also call into question the assumption that long reading times at the main verb reflect an association, or lack thereof, between the verb and a filler, since they exhibit comparable effects in cases when an association was made (in the filled-gap condition) and when it was not made (in the head-specificity condition). Thus, as previously mentioned, the current paper suggests that reading times at the verb reflect the nature of the comparison made between the filler as required by the verb and the one provided. While this explanation may not be the only one available, it is simple, reasonable and fits in nicely with the reading times obtained at the adverb. The two conditions with significantly long reading times at the main verb, namely the filled-gap and head-specificity conditions, are those that demand a difficult comparison between members of the same syntactic category, i.e., two prepositions, rather than members of two different categories, i.e., preposition or noun. Of these two conditions, only head-specificity results in an incompatible match. Moving forward in this situation, the parser may find it difficult to forgo the almost-compatible preposition, explaining why there are long reading times at the adverb in this condition alone.

5.1. Future research

Several avenues of possible research stem from the current paper.

First, note that the difference in reading times at the filled-gap position between the subcategorization condition and the baseline approaches significance, in the by-participants analysis ($p = 0.071$). This could be due to the fact that in Hebrew, speakers have started omitting

prepositions and producing sentences such as “ani lo zoxer eize rehov garti” (I don’t remember which street I lived) instead of “ani lo zoxer be-eize rehov garti” (I don’t remember on which street I lived). This result is understood if participants that accept these omissions formed a dependency with the incompatible NP *eilu* only to be surprised by the true PP-complement, just like in the filled-gap condition. To investigate this possibility, a post-test, such as a sentence completion questionnaire, could be used to determine which of the original participants, if any, are susceptible to omitting prepositions, and repeat the analyses without them.

Second, regarding the possibility that long reading times at the main verb in the filled-gap and head-specificity conditions indicate that a dependency was formed in *both* conditions, suggesting that the parser initially ignores the mismatch in head-specificity, An ERP study may prove useful. The same materials would be reused on a new set of participants in search of a relevant component at the main verb in the filled-gap condition but not the subcategorization and head-specificity conditions, such as the P600 (Brouwer, Fitz and Hoeks, 2012; Gouvea, Phillips, Kazanina and Poeppel, 2010). Finding this component and isolating it to the filled-gap condition would reinforce the current theory that (a) a dependency was never formed in the subcategorization and head-specificity conditions and (b) reading times at the main verb reflect the nature of the comparison made between the filler and the verb.

Finally, other interpretations of the current results are surely possible. As explained, one could suggest that a dependency *was* formed in both the filled-gap and head-specificity conditions. Proponents of this theory would then have to assume that this error is discovered at the filled-gap position in the filled-gap condition and at the adverb in the head-specificity condition, explaining the lack of filled-gap effect in the latter condition. While no obvious reason comes to mind as to why an adverb should trigger this discovery, there might be. The important

thing to remember is that no matter the theory of early filler-gap dependency formation, it should be based on the stable interpretation of reading times at the filled-gap position, and reading times from any other position, including the main verb, must accommodate it.

References

- Boland, J., Tanenhaus, M., Garnsey, S., and Carlson, G. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, 34, 774-806.
- Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain research*, 1446, 127-143.
- Chomsky, N.A. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT press.
- Cole, Peter. 1976. "An apparent asymmetry in the formation of relative clauses in Modern Hebrew". In P. Cole (Ed.), *Studies in Modern Hebrew syntax and semantics*. Amsterdam: North Holland Publishing.
- Crain, S., and Fodor, J. D. (1985). How can grammars help parsers. In D. Dowty, D. Karttunen, and A. M. Zwicky (Eds.), *Natural language parsing: Psycholinguistic, computational, and theoretical perspectives* (pp. 94-128). Cambridge, UK: Cambridge University Press
- Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and Performance 12: The Psychology of Reading*, 559-586. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Frazier L., and Fodor, J.D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291-326.
- Frazier L., and Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178-210.
- Friedmann, N., Aram, D., and Novogrodsky, R. (2011). Definitions as a window to the acquisition of relative clauses. *Applied Psycholinguistics*, 32, 687-710.
- Gouvea, A. C., Phillips, C., Kazanina, N., and Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25, 149-188.
- Grimshaw, J. (1979). Complement selection and the lexicon. *Linguistic Inquiry* 10, 279-326.
- Holmes, V.M., Stowe, L., and Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language*, 28, 668-689.
- Just, M. A., Carpenter, P. A., and Woolley, J. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 11, 228-238.
- MacDonald, M.C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9, 157-201.
- Omaki, A., Lau, E.F., White, I.D., Dakan, M.L., Apple, A., and Phillips, C. (2015). Hyper-active gap filling. *Frontiers in psychology*, 6, 384.
- Pickering, M.J., Traxler, M. J., and Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language*, 43, 447-475.
- Pickering, M., and Traxler, M. (2003). Evidence against the use of subcategorization frequency in the processing of unbounded dependencies. *Language and Cognitive Processes*, 18, 469-503.
- Ross, J.R. (1967) *Constraints on variables in syntax*. Doctoral dissertation, Massachusetts Institute of Technology.

- Spivey-Knowlton, M., and Sedivy J.C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, 55, 227-267.
- Stowe, L.A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and cognitive processes*, 1, 227-245.
- Tanenhaus, M.K., Carlson, G.N., and Trueswell, J.C. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processing*, 4, 211-234.
- Tanenhaus, M.K., Boland, J., Garnsey, S.M., and Carlson, G.N. (1989). Lexical structure in parsing long-distance dependencies. *Journal of Psycholinguistic Research*, 18, 37-50.
- Traxler, M.J., and Pickering, M.J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study, *Journal of Memory and Language*, 35, 454-475.
- Tyler, L.K., and Marslen-Wilson, W.D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 683-692.
- van Schijndel, M., Schuler, W., and Culicover, P.W. (2014). *Frequency effects in the processing of unbounded dependencies*. Annual Meeting of the Cognitive Science Society.
- Wagers, M.W., and Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly journal of experimental psychology*, 67, 1274-1304.

תקציר

עבודה זו חוקרת כיצד שני סוגים של מידע לקסילי משפיעים על התהליך של יצירת תלות תחבירית. ניסוי קריאה בקצב אישי ניצל את הפרדיגמה המוכרת של פילד-גאפ (filled-gap effect) כדי לבדוק האם המעבד מתעלם מאי-התאמה בקטגוריה התחבירית ו/או ספציפיות הראש (head specificity) בבואו להחליט האם יש ליצור תלות תחבירית בעמדת המושא או לא. משפטים דקדוקיים וסבירים בעברית הדגימו מידות שונות של התאמה בין הפועל לפילר מבחינת קטגוריה תחבירית ומילת היחס הספציפית שעומדת בראש הפילר. התוצאות מעידות על כך שהמעבד לא מתעלם מאי-ההתאמות הללו במהלך העיבוד הראשוני של תלות תחבירית. בנוסף, על אף שזמני הקריאה בפועל העיקרי אינם חד-משמעיים, ייתכן שהם משקפים השוואה שנערכת בין הדרישות של הפועל לבין המאפיינים של הפילר שמסופק במשפט, ולא דווקא קישור ויצירת אסוציאציה בין השניים.

אוניברסיטת תל-אביב
הפקולטה למדעי הרוח ע"ש לסטר וסאלי אנטין
החוג לבלשנות

מידע לקסיקלי ועיבוד משפטים בעברית בזמן אמת: מבט על מילות יחס ויצירת תלות תחבירית

חיבור זה הוגש כעבודת גמר לקראת התואר
"מוסמך אוניברסיטה" - M.A. באוניברסיטת ת"א

על ידי

קיילה גולד-שלו

העבודה הוכנה בהדרכת

דר' איה מלצר-אשר

מאי 2017