Argument structure affects relative clause extraposition: Corpus evidence from Persian

Nasimeh Bahmanian, Goethe University Frankfurt, DE, bahmanian@em.uni-frankfurt.de

Markus Bader, Goethe University Frankfurt, DE, bader@em.uni-frankfurt.de

Sol Lago, Goethe University Frankfurt, DE, sollago@em.uni-frankfurt.de

The extraposition of a relative clause creates a discontinuous dependency between the relative clause and its host noun phrase, as in *A man just entered the bank who claimed to have a gun*. Since discontinuous dependencies are known to increase processing effort, a key question is why speakers produce them in the first place. Some factors known to affect extraposition – for example, the length of the relative clause and the main verb phrase – have received processing-based explanations, but others haven't. We focus on two factors described by previous research: verb type and the grammatical function of the noun phrase hosting the relative clause. Specifically, extraposition from grammatical subjects is more common with unaccusative and passive verbs in English; further, extraposition is more common from grammatical objects than subjects in Dutch and German. We replicate these findings using corpus data from Persian. Further, we propose that verb type and grammatical function can be linked to a single underlying notion: argument structure. We demonstrate that argument structure modulates the likelihood of extraposition in Persian. We suggest that this occurs because speakers choose to extrapose relative clauses in order to keep the main clause verb close to its internal arguments. This explanation extends previous findings in psycholinguistics on the role of argument structure in speech planning during language production.

1. Introduction

In many languages, relative clauses can be displaced from a noun-adjacent position to a clause-final position, for example ***A man*** *just entered the bank* ***who claimed to have a gun***. This phenomenon is called "relative clause extraposition" (henceforth, "extraposition"), and it creates a discontinuous dependency between a relative clause (RC) and its host noun phrase. Since discontinuous dependencies are known to increase processing effort (Futrell et al., 2020; Gibson, 2000; Levy et al., 2012), an outstanding question is why speakers produce extraposed structures in the first place. We investigate this question by relating grammatical factors that affect the likelihood of extraposition to processing constraints that affect speech production, according to the psycholinguistics literature.

In previous studies, processing-based explanations of extraposition have focused on relative differences between the length of a relative clause and the main verb phrase (Francis, 2010; Hawkins, 1994, 2004; Rasekh-Mahand et al., 2016). Specifically, extraposition from subjects is more likely when the relative clause length exceeds the length of the main verb phrase. This finding has been attributed to a processing preference to place shorter before longer constituents (Arnold et al., 2000; Wasow, 1997a, 1997b).

However, there are other factors that affect extraposition for which no processing explanation is available. We focus on two: verb type and the grammatical function of the noun phrase hosting the relative clause. With regard to verb type, extraposition from subject noun phrases is more likely when the main clause verb is unaccusative or passive, compared to active transitive or unergative (Culicover & Rochemont, 1990; Francis, 2010). With regard to grammatical function, previous research suggests that extraposition is more common from object than subject noun phrases in languages with pre-verbal objects, like German and Dutch (Bader, 2015; Perlmutter & Zaenen, 1984; Shannon, 1992). However, such a comparison has not been empirically tested, a gap that is bridged by the current study.

The next section describes previous crosslinguistic research on the role of constituent length, verb type and grammatical function in extraposition. Then, we present a novel proposal linking verb type and grammatical function to a single underlying notion: argument structure. Using corpus data, we demonstrate that argument structure modulates the likelihood of extraposition in Persian, a language that has been understudied in comparison to Germanic languages. Because argument structure has been shown to affect sentence production (Momma & Ferreira, 2021; Momma et al., 2014, 2016, 2018; Schriefers et al., 1998), we suggest that redescribing verb type and grammatical function in terms of argument structure allows for a unified processing explanation of their role in extraposition.

2. Background

2.1 The role of constituent length

Different factors have been linked to the occurrence of extraposition (for review, see Francis, 2021, on English; Strunk, 2014, on German). Here we focus on those that have been linked to processing constraints. One such factor is constituent length: it has been proposed that relative clauses are extraposed in order to reduce the distance between the host noun and the verb on which it depends. For example, in (1), the relative clause is fifteen words, while the main verb phrase (underlined) is three words long. Thus, extraposition in (1b) increases the dependency distance between the host noun *evidence* and its modifying relative clause from zero to three words. But, crucially, it reduces the distance between the noun and the verb from seventeen to two words. Thus, extraposition is advantageous in (1b), because it reduces dependency distances overall, as compared to the adjacent version in (1a).

(1)   a.   Evidence [$_{RC}$ that suggests gene trees may differ in topology from each other and the true tree] <u>has been provided</u>.
         b.   Evidence <u>has been provided</u> [$_{RC}$ that suggests gene trees may differ in topology from each other and the true tree].

The role of constituent length has been attributed to a processing bias towards local dependencies, according to which longer dependencies are harder to process than shorter dependencies (Futrell et al., 2015, 2020; Gibson, 2000; Hawkins, 2004; Temperley, 2007). Under such accounts, sentences like (1b) represent a more computationally efficient ordering that benefits both producers and comprehenders (Hawkins, 1994, 2004). Similarly, for language production, Arnold et al. (2000) and Wasow (1997a, 1997b) have suggested that the tendency to place shorter constituents before longer ones – known as the "short-before-long" or "end-weight" preference – reflects a processing mechanism that facilitates sentence planning by postponing the production of longer, more complex constituents (see Jaeger & Norcliffe, 2009, and references cited there). Together, accounts in both production and comprehension predict that extraposition should be more likely when the relative clause is long and/or the extraposition distance is short.[1]

---

[1] Note that length effects have been argued to depend on head directionality (Hawkins, 1994). For verb-final languages, empirical data suggest that the direction of length effects (short-before-long or long-before-short) depends on whether the ordering involves the pre-verbal or post-verbal domain (Wasow, 2013). According to Hawkins, in addition to the sentence head, the head direction of the noun phrase plays a role as well. As evidence, he presents relative clause extraposition in German, a verb-final language with post-nominal relatives that shows a short-before-long preference in ordering relative clauses. The language examined in this study, Persian, similar to German, has mixed head directions and has shown a short-before-long preference for relative clause extraposition, which involves the post-verbal domain (Rasekh-Mahand et al., 2016; for a detailed discussion on Persian word order, see Faghiri & Samvelian, 2020).

The prediction above is empirically well supported by corpus and experimental comprehension and production studies. Corpus studies have shown that relative clause length and extraposition distance influence the likelihood of extraposition in English (Francis, 2010; Francis & Michaelis, 2014), German (Strunk, 2014; Uszkoreit et al., 1998), and Persian (Rasekh-Mahand et al., 2016). In comprehension, acceptability studies in English and German have shown that even though adjacent sentences are usually preferred, compared to extraposed ones, the acceptability of extraposed sentences improves with increasing relative clause length and decreasing extraposition distance (Francis, 2010; Francis & Michaelis, 2017; Konieczny, 2000; Uszkoreit et al., 1998). Similar results are found in elicited production experiments in English and German (Bader, 2014; Francis & Michaelis, 2017). For example, Francis & Michaelis (2017) conducted a production experiment in which English participants had to build sentences by combining a noun phrase, a relative clause modifier and a verb. The results showed that extraposed relative clauses occurred most often when the relative clause was long, and the extraposition distance was short. In German, Bader (2014) conducted two production-from-memory experiments in which participants could place a relative clause in whatever position seemed most natural. The results showed that sentences with an extraposed relative clause occurred most often at short extraposition distances.

2.2 The role of verb type

With regard to verb type, previous work has focused on whether extraposition depends on the syntactic and semantic properties of the main clause verb. Concerning syntactic properties, verbs can be copular, transitive, or intransitive. Copular verbs link a subject to a predicate, while transitive verbs have an obligatory object argument in addition to a subject argument. Intransitive verbs have only a subject argument, and they can be further subdivided into *unaccusative* and *unergative* verbs, depending on the thematic role of their sole argument. Unaccusatives have a theme/patient subject, while unergatives have an agent subject (Perlmutter, 1978). Differences between unaccusative and unergative verbs might play a role in extraposition. For example, it has been claimed that in neutral contexts, extraposition with an unaccusative verb (2a) is more acceptable than extraposition with an unergative verb (2b) (Culicover & Rochemont, 1990; Guéron, 1980).

(2)   a.   A man arrived who wasn't wearing any clothes.
        b.   ??A man screamed who wasn't wearing any clothes.

These informal reports are supported by a corpus study in English (Francis, 2010), which found that extraposition rates were overall low (15%), but increased with unaccusatives and passives (37%) compared to other verb types (6%). It was argued that this finding was in line with Culicover & Rochemont (1990), because unaccusatives and passives can be classified together,

based on their syntactic and semantic properties: both have a single theme/patient argument. Further, a corpus study with 757 relative clauses in Persian found that for subject-modifying relatives, extraposed clauses occurred more often with copular than with transitive verbs – in fact, relative clauses with copular verbs were more likely to appear extraposed than in an adjacent position (Rasekh-Mahand et al., 2016).

Finally, there is evidence that the semantic properties of the verb also play a role in extraposition. Specifically, extraposition has been proposed to be more acceptable with so-called *verbs of appearance* – a subset of unaccusative verbs – such as *appear* and *arrive* (Culicover & Rochemont, 1990; for a similar claim for extraposition of prepositional phrases, see Guéron, 1980). There are also reasons to assume the existence of such a category from a lexico-semantic point of view. For example, Levin & Rappaport Hovav (1995) suggest that appearance verbs are special cases of intransitive (unaccusative) verbs that lack a causative argument in their lexical semantic representation.

The distinction between appearance and non-appearance verbs is supported by an acceptability study in English (Walker, 2013) that compared acceptability judgments of sentences with an extraposed relative clause that contained appearance vs. non-appearance verbs: *appear*, *enter*, *arrive* vs. *faint*, *stumble*, and *smile*, respectively. The results showed that extraposed sentences were more acceptable with appearance than non-appearance verbs, supporting a role for information about verb semantics. Unfortunately, this study did not elicit judgments to adjacent relative clauses, and it also collapsed across different syntactic classes within the non-appearance category (e.g., *faint* and *stumble* are unaccusative, but *smile* is unergative). The current study addresses this issue by jointly considering the semantic and syntactic properties of verbs.

2.3 The role of the grammatical function of the host noun phrase

A final relevant factor for our study concerns the grammatical function of the noun phrase hosting the relative clause. Previous research largely focused on extraposition from grammatical subjects, probably because historically, these studies started by investigating English extraposition. Although extraposition from non-subjects is possible in English (e.g., *I saw a man in the bank who claimed to have a gun*), extraposition across a main clause verb is not possible, because non-subjects in English are post-verbal. In this regard, languages that allow preverbal subjects and objects – e.g., German, Dutch and Persian – are better suited for an investigation of the role of the grammatical function of the host noun phrase.

In the literature on German, an asymmetry in extraposition from object vs. subject nouns was noted, but it was linked to focus-related factors. Based on a corpus of 1,200 German relative clauses, Shannon (1992) noted that the antecedent of extraposed relative clauses often occurred sentence-medially in the typical focus position in German, directly in front of the clause-final

verb, but not in sentence-initial position. Specifically, extraposition seemed very frequent from predicate nominals (3) and, to a lesser extent, from direct objects (4) – both with antecedents in sentence-medial positions. By contrast, grammatical subjects in sentence-initial position seemed resistant to extraposition: (5) is one of the few cases found by Shannon (1992). Thus, although Shannon (1992) did not report rates of extraposition from different grammatical functions, his observations suggest that in German, extraposition from subjects is less common than from other grammatical functions.

(3)   *Extraposition from a predicate nominal*
       Diese Freiwilligen werden die besten Männer sein, [die wir haben].
       these volunteers   will   the best   men   be   who we   have
       'These volunteers will be the best men that we have.'

(4)   *Extraposition from a grammatical object*
       Jeder Schritt löste   Schweiß aus, [der   sofort   mit   Bier   ersetzt
       each step   brought   sweat   out, which immediately   with   beer   replaced
       werden mußte].
       become must-PST
       'Each step caused sweat, which immediately had to be replaced with beer.'

(5)   *Extraposition from a grammatical subject*
       Neue Magazine waren gegründet worden,   [die   Astounding die
       new   magazines were   founded   become-PTCP-PRF which Astounding the
       Führungsposition streitig   machten].
       leading position controversial made
       'New magazines had been founded which rivaled with Astounding for the lead.'

Further, Bader (2015) observed that extraposition from objects was preferred when only verbal material intervened (compatible with acceptability data in Konieczny, 2000). He contrasted this with extraposition from subjects of transitive verbs, in which there was at least an intervening object noun phrase. He provided length-, discourse- and prosody-based explanations for this pattern. However, he did not quantify extraposition from subjects of intransitives or passives, in which no non-verbal material intervened. This comparison is relevant, because it could reveal whether grammatical subjects and objects – depending on verb type – behave differently with respect to the effect of length in extraposition. Our study bridges this gap.

    Our study proposes that the role of the grammatical function of the host noun phrase should be considered together with verb type, thereby suggesting a role for argument structure in extraposition. Before we introduce our proposal, we describe previous accounts of extraposition which have attributed the effects of verb type and grammatical function to discourse structure and/or prosodic constraints.

2.4 The role of discourse and prosody

Some accounts have tried to explain the occurrence of extraposition in terms of discourse and pragmatic factors (Culicover & Rochemont, 1990; Guéron, 1980; Lee, 2019; Poschmann & Wagner, 2016; Takami, 1999). For example, it has been proposed that extraposition functions as a device to introduce a new subject into the discourse, with the intervening verb phrase conveying old or backgrounded information. This claim is supported by the observation that extraposition is more frequent from indefinite noun phrases, which typically introduce new referents (for related syntactic proposals, see Baltin, 2006; Guéron & May, 1984; Kayne, 1994), or more acceptable from focused noun phrases (Bolinger, 1992; Huck & Na, 1990). Discourse-based accounts can also explain the increased likelihood of extraposition with verbs of appearance, because such verbs often introduce a new referent.

When discourse-based accounts are evaluated with empirical data, the results are mixed. For example, using a corpus of 345 English sentences containing subject-modifying relative clauses, Francis (2010) and Francis & Michaelis (2014) showed that extraposition was more likely when the subject was new and the verb phrase was contextually accessible – when a semantically related verb phrase had been mentioned in the previous context. However, Rasekh-Mahand et al. (2016) did not find that the likelihood of extraposition in Persian was affected by the givenness or accessibility of the subject or the verb phrase. Further, extraposition is not always limited to indefinite subjects: a corpus study in German with 2,555 relative clauses found that approximately 35% of extraposed relatives had a definite antecedent (Strunk, 2014). Another concern with accounts based on givenness/newness is that they cannot explain extraposition from noun phrases that serve other grammatical functions, such as object, which are part of the verb phrase. Therefore, it is unclear how they generalize to languages like Dutch, German, and Persian, in which extraposition from non-subject nouns is more frequent.

Discourse-based accounts of extraposition have also been proposed for Dutch and German, but based on different observations. Perlmutter & Zaenen (1984) suggested that extraposition in these languages was possible only if the antecedent was a non-subject and occurred in a non-initial sentence position. Corpus data from Shannon (1992) confirmed that sentence-initial arguments do not usually show extraposition in German, but it suggested that it was not subjecthood – but rather initial position – that was resistant to extraposition. Specifically, although extraposition from sentence-initial positions was rare in German, it was not limited to subject nouns: it was attested with other elements, such as direct and indirect objects, prepositional phrases, and adverbs. Based on this observation and observations about definiteness, Shannon proposed a pragmatic account according to which extraposition is motivated by sentence focus. The claim was that extraposition was only possible with focused antecedents, which are new and/or more important information and thus tend to occur in a non-initial sentence position.

While the focus-based theory of Shannon (1992) can account for extraposition from non-subject positions, its empirical support is mixed. When Shannon asked a German native speaker to read aloud sentences from his corpus, the noun modified by the extraposed relative clause was not always focused and did not always carry the sentence stress, contrary to his expectations. An effect of focus was reported in an acceptability experiment conducted by Poschmann & Wagner (2016). In this experiment, participants listened to a question, read aloud a predetermined answer and then provided an acceptability judgment. The results showed a higher acceptability of extraposition in response to subject focus questions, as compared with object focus and wide focus questions. However, there was no evidence that accent on the antecedent of the relative clause affected extraposition. In the domain of production, Bader (2024) reported a small effect of context on the extraposition rate from object nouns. In a constrained production experiment in German, participants silently read a context scenario followed by a set of phrases in a jumbled order. Their task was to arrange the phrases into a sentence and then utter it. In the focus condition, the main verb was mentioned in the context, while the noun phrase and relative clause in the target sentence were new. In the topic condition, the main verb was new, while the noun phrase and relative clause were previously mentioned in the context. Participants extraposed the relative clause more often in the focus condition than in the topic condition, although the effect was small (56% vs. 47%).

Finally, other accounts have tried to explain extraposition based on prosodic considerations (Féry, 2015; Göbbel, 2013; Hartmann, 2013, 2017; Truckenbrodt, 1995). For example, Hartmann (2013, 2017) suggested that extraposition was used as a re-ordering strategy to avoid ill-formed prosodic structures. In this account, information structure is only indirectly related to extraposition, because the alleged prominence on the host noun, which is sometimes realized as focus, results from the prosodic restructuring that happens due to extraposition.

The prosody of the intervening material has also been argued to affect extraposition, such that prominent interveners block extraposition more than unaccented ones. For example, extraposition is more likely over a locational or directional adverbial compared to a noun phrase (Kathol & Pollard, 1995), or over verbs of appearance compared to other verbs (Guéron & May, 1984). In both cases, the former element is typically assumed to be unaccented (Bolinger, 1992; Poschmann & Wagner, 2016). Others have argued against this idea (see Féry, 2015, for a review). Empirical data, has not provided much support for these hypotheses. For example, Bader (2014) showed in a production-from-memory experiment that even an unaccented indefinite pronoun (*etwas*, 'something') decreased the likelihood of extraposition from subjects when the pronoun was the only intervening material before the verb.

In sum, the available empirical evidence in English and German suggests that neither prosody nor discourse and pragmatic factors can fully explain the occurrence of extraposition across different verb types and grammatical functions. Thus, it is useful to explore additional

explanations. The next section outlines a processing-based account of extraposition in terms of argument structure.

3. The present study

We propose a processing-based explanation of two factors that affect extraposition: verb type and the grammatical function of the host noun phrase. We suggest that these factors might follow from a single factor: verb argument structure.

Argument structure refers to a labeled list of arguments that a lexical item can have (Williams, 1981). For example, the argument structure of the English verb *hit* is: (Agent, Theme). These two arguments are instantiated as *Bill* and *the button* in the sentence: *Bill hit the button*. A relevant notion for describing argument structure is that of *maximal projection* (Chomsky, 1970, 1986; Jackendoff, 1977). Maximal projection refers to a node in a syntactic tree where an item's lexical feature no longer projects upward. For example, the node VP (verb phrase) is the maximal projection of a verb. Syntactic theories posit a special status for an argument located outside of a lexical item's maximal projection and corresponding to an NP (noun phrase) of which the maximal projection of that item is predicated (Williams, 1981). In the example above, *Bill* is an agent argument, of which the verb phrase *hit the button* is predicated. The argument *Bill* is called *external*: it is located outside the verb's maximal projection and it is distinct from other arguments, called *internal*, which are realized inside the verb's maximal projection (e.g., *the button*). Crucially, this implies that internal arguments have a closer syntactic and semantic relationship to the verb than its external argument.

Under the definitions above, objects of active transitive verbs, and subjects of passive as well as unaccusative verbs, can be grouped under the label of *patients/themes* which are considered internal arguments. Thus, they contrast with subjects of transitive and unergative verbs, which function as thematic agents of a verb and are considered external arguments (Dowty, 1989; Grimshaw, 1990; Williams, 1981). Our proposal is that relative clauses modifying internal arguments are more likely to be extraposed than relative clauses modifying external arguments. This means that speakers might extrapose not only based on discourse or pragmatic factors or length. Extraposition might also be influenced by sentence-internal factors, more specifically, by the argument structure of the main verb in a sentence.

An advantage of referring to argument structure – as opposed to separately mentioning verb type and the grammatical function of the host noun phrase –is that it suggests a processing reason for why these two factors affect extraposition. Argument structure has been shown to influence sentence processing in tasks like verb recognition and sentence comprehension (Bever & Sanz, 1997; Shetreet et al., 2010), lexical priming (Friedmann et al., 2008), lexical decision (Kauschke & Stenneken, 2008), and action naming (Kauschke & von Frankenberg, 2008; for

review, see Heinzova et al., 2022). More specifically, research on sentence production has demonstrated that the syntactic relationship between verbs and their arguments influences sentence planning: Speakers tend to plan the verb before its preverbal internal arguments (patients/themes), but no such tendency is found with external arguments (Momma & Ferreira, 2021; Momma et al., 2014, 2016, 2018; Schriefers et al., 1998).

Production studies have used the so-called picture-word interference paradigm to investigate the time-course of verb planning in verb-final structures in English, Japanese and German. This paradigm takes advantage of the *semantic interference effect*: naming a picture of a dog takes longer when the word is presented together with a semantically similar word (e.g., *fish*), as compared to a dissimilar word (e.g. *tree*; Lupker, 1979; Roelofs, 1992; Schriefers et al., 1990). This effect is attributed to interference in the process of selecting the syntactic and semantic representation of a word – typically called its *lemma*.

Recent studies have used an extended version of this paradigm to examine whether a verb's lemma is retrieved before the utterance onset in verb-final structures – an indicator of early verb planning. They found that the interference effect depended on the external vs. internal argument status of the verb's arguments. Specifically, there was a verb interference effect before sentence-initial objects, but not subjects, in Japanese transitive structures (Momma et al., 2016), before sentence-initial subjects in English passive, but not active structures (Momma et al., 2014), and before sentence-initial subjects in unaccusative, but not unergative structures (Momma & Ferreira, 2021; Momma et al., 2018; Schriefers et al., 1998).

Based on previous findings as well as the results of six picture-word interference experiments, Momma & Ferreira (2021) proposed that speakers retrieve sentence-final verbs before the articulation of their sentence-initial patient/theme arguments, but not agent arguments, and before retrieving sentence-medial nouns inside propositional modifiers. They suggested that the time-course of sentence planning reflects hierarchically-defined dependency relationships over and above linear structure. They provided several possibilities for why internal arguments have a closer relation to the verb during planning, including closer semantic, syntactic, and/or cognitive relations.

Our proposal applies this logic to relative clause extraposition. We assume that a sentence-final verb is more likely to be planned before producing its internal arguments, but a relative clause does not need to be planned early – similarly to an intervening prepositional phrase in Momma & Ferreira (2021). As a result, when planning a relative clause modifying an internal argument, speakers need to keep the dependency relation between the argument and the verb lemma activated in memory, which can affect their choice to either postpone the relative clause (and produce the already planned verb), or to keep the verb in memory while planning the relative clause.

Planning a relative clause while keeping the main clause verb in memory should be costly. Thus, it would be less costly to produce the verb as close as possible to its preverbal internal argument if there is interfering material. Crucially, this suggests that extraposition might be more likely from internal arguments – like the subject of passive or unaccusative verbs – because it allows speakers to keep such noun phrases linearly close to the verb in speech. By contrast, there might be less processing pressure to keep a verb and a noun phrase close when the noun phrase instantiates an external argument – for example, the subject of a transitive or unergative verb – thus reducing the likelihood of extraposition. Therefore, showing that verb argument structure affects extraposition provides a unified production-based explanation for why extraposition is more common with certain verb types and from certain grammatical functions.

This article addresses the research question of whether argument structure affects relative clause extraposition in Persian. Persian is an SOV language with post-nominal relative clauses (Mahootian, 2002). Persian is especially interesting to study because, unlike English, extraposition occurs frequently from non-subject positions, such as objects, as in (6), and predicate nominals, as in (7). Thus, Persian allows a direct comparison of extraposition rates from subject and object hosts. This makes it possible to test hypotheses about the role of argument structure, thus bridging a gap in previous studies, which quantified the extraposition rates from subjects and objects separately (e.g., across different languages or only non-quantitatively).

(6)   *Extraposition from a grammatical object in Persian*
      dar   ānjā   faqat   bimārān-i   rā   mipazirand [ke   atebbā   ānhā rā
      in   there only   patients-INDEF DOM accept-3PL   that   doctors they DOM
      javāb   karde   bāšand].
      answer have   done
      'At that place, they only accept patients on whom the doctors have given up hope.'

(7)   *Extraposition from a predicate nominal in Persian*
      dozdi-e   daryā'i padide-i   nist   [ke   dowrān-e ān   digar be   sar
      theft-EZ   marine   phenomenon-INDEF   is not   that   era-EZ   it   other to   head
      āmade   bāšad].
      has   come
      'Piracy is not a phenomenon whose time has come to an end.'

                                                                    PerUDT corpus (Rasooli et al., 2020)

To date, there is one previous corpus study on Persian, but this study only examined extraposition from grammatical subjects (Rasekh-Mahand et al., 2016). The results replicated findings from English: extraposition from subjects was more likely when the relative clause was longer than the verb phrase. Concerning the role of verb type, a higher rate of extraposition was found with copular verbs (44%) and unaccusatives (24%) compared to transitives (15%), unergatives

(3%) and passives (9%). The current study seeks to re-examine these findings on the role of constituent length and verb type, and it additionally investigates extraposition from non-subject noun phrases, and the role of verb semantics and argument structure.

With regard to the expected patterns of extraposition in Persian, our proposal regarding argument structure makes several predictions, summarized in (8). Specifically, we expect higher extraposition rates from internal arguments than external arguments, which leads to the prediction that extraposition should be more likely from objects and from subjects of unaccusatives and passives, than from subjects of unergatives and transitives. With regard to predicate nominals, our proposal does not make a clear prediction, because predicative phrases cannot be classified as internal or external to a copular verb. However, based on previous findings in German, which showed the highest rates of extraposition with nominal copular predicates (Shannon, 1992), we might expect a similar pattern in our Persian data.

(8)   *Predicted extraposition rates according to argument structure*
        predicate nominals > objects of transitive verbs, subjects of unaccusative and passive
        verbs > subjects of unergative and transitive verbs

4. Corpus study

We used two dependency treebank corpora of Persian: Seraji Persian UD (Seraji, 2015) and PerUDT (Rasooli et al., 2020). These corpora contain 6,000 and 29,000 sentences, respectively, annotated in the CoNLL-U format, which provides information such as word order, part of speech or POS tags, dependency relations between words within a sentence and their syntactic type (https://universaldependencies.org/format.html). The data used in this study was extracted from the corpora and then manually checked as described below.

4.1 Data preparation and coding

In the first step, 4,020 sentences containing a relative clause were extracted from the two corpora. The sentences were automatically coded for relative clause position (adjacent/extraposed), relative clause length (number of words in the clause), (potential) extraposition distance, grammatical function of the host noun phrase (subject/object/nominal copular predicate), verb voice (active/passive), and whether the object noun had direct object marking (DOM). Note that (potential) extraposition distance refers to the number of words between the antecedent of the relative clause and the relativizer *ke* in the extraposed version as shown in (9b) (from PerUDT corpus). In the adjacent version (9a) (constructed from the original extraposed sentence), the potential distance for extraposition was the number of words in the post-relative clause region of the sentence.

(9)   *Extraposition from grammatical subject*

a.   qavānin   va   moqarrarāt-e   mote'added   [ke ba'zan   bā   yekdigar
     laws   and regulations-EZ numerous   that sometimes   with each other
     ta'āroz   dāšte-and]   barāye   vāgozāri-ye   sahām-e   dowlati   tasvib
     conflict   have had   for   delegation-EZ share-EZ   governmental approval
     šode ast.
     has   become
     'Many rules and regulations which sometimes have conflicted with each other have
     been passed to delegate the government shares.'

b.   qavānin   va   moqarrarāt-e   mote'added   barāye   vāgozāri-ye   sahām-e
     laws   and regulations-EZ numerous   for   delegation-EZ share-EZ
     dowlati   tasvib   šode ast   [ke ba'zan   bā   yekdigar   ta'āroz
     governmental approval has   become   that sometimes   with each other   conflict
     dāšte-and].
     have   had
     'Many rules and regulations have been passed to delegate the government shares,
     which sometimes have conflicted with each other.'

In the second step, a list of unique main clause verbs was created from the extracted data. These
verbs were manually classified as transitive active, transitive passive, unergative, unaccusative
appearance, and unaccusative non-appearance (see the next section for definitions of verb
classes). Copulas and verbs that did not fall into any of these categories were classified as: (i)
BE-HAVE stative, which comprised all adjectival copular constructions, such as *šād budan* 'be
happy', and light verb constructions with the light verb *dāštan* 'have', such as *qarār dāštan* 'be
located', (ii) existential, which comprised *budan* and *vojud dāštan* 'exist', and (iii) copular with a
predicate nominals (henceforth, *copula*).

During the second step, all sentences with a copula were manually checked to ensure that the
verb type was correctly categorized. Additionally, sentences with a main clause verb that could
potentially take a complement clause, such as *tasmim gereftan* 'decision make', were manually
checked. Sentences were discarded if they did not include a relative clause, but rather a wrongly
coded complement or appositive clause that modified a proper noun. Other discarded cases
included incomplete or incomprehensible sentences, sentences with an ambiguous relative clause
attachment site, and sentences in which the noun phrase hosting the relative clause was in a
non-canonical order (e.g., topicalized). Automatic annotations were also manually checked if
they included unusually long relative clauses and/or extraposition distances, or incompatibilities
such as a noun phrase being coded as an object when the main clause verb was not transitive.
This resulted in the manual checking of 1,986 sentences (out of 4,020) and the exclusion of

990 sentences. From the remaining 2,034 sentences, a random subset of 102 tokens (5% of the data) was manually checked to evaluate the accuracy of the automatic annotations. The rate of exclusion based on the criteria described above was 15% (15 out of 102 tokens). Additionally, of the 87 included tokens, only four relative clause lengths and eight extraposition distances had to be manually corrected. This means that the accuracy of automatic annotation for relative clause length and extraposition distance was 95% and 91%, respectively.

Finally, relative clauses were coded for the argument type associated with their host noun phrase: internal vs. external. Existentials, nominal copular constructions, and BE-HAVE statives were coded as lacking argument structure (Dechaine, 1993; Hazout, 2004). Note that adjectival copular predicates are sometimes considered to have an argument structure, because it is argued that the adjective behaves like a verb and can take complements (e.g., Chomsky, 1981; Haegeman, 1999; Stowell, 1983). Because this proposal is controversial in the theoretical literature (Pesetsky, 1982; Rothstein, 1999), we chose not to code adjectival copular predicates as having an argument structure.

4.1.1 Classification of unaccusative and unergative verbs

The classification of intransitive verbs into unaccusative and unergative classes was done based on two tests of unaccusativity in Persian (Karimi-Doostan, 1997): the formation of subject nominals vs. adjectival past participles, and whether the verb had a transitive/causative counterpart. According to the first test, unergatives can form subject nominals but not adjectival past participles (10), while unaccusatives cannot form subject nominals but can form adjectival participles (11). This is because the derivation of subject nominals requires an external argument, while the derivation of adjectival past participles requires an internal argument (Karimi-Doostan, 1997, p. 145).

(10) a. davidan 'run' → davande 'runner'
      b. davidan 'run' → *davide 'run'

(11) a. mordan 'die' → *mirande 'dier'
      b. mordan 'die' → morde   'dead'

The second test is only applicable to light verb constructions, which are more common than simple verbs in Persian. According to Karimi-Doostan (1997), any unaccusative light verb construction should have a transitive counterpart. This is shown in (12), where an unaccusative verb is formed with the light verb *xordan* and a transitive counterpart is formed with the light verb *dādan*. In our data, when it was unclear whether the subject nominal was unacceptable, this second test was applied. If a transitive counterpart was available, the verb was classified as unaccusative. Unaccusative verbs were further classified into verbs of appearance and non-appearance, based on the semantic properties of appearance verbs listed by Levin (1993).

(12) *unaccusative*   *transitive*

ŝekast xordan → ŝekast dādan

defeat collide defeat give

'to be defeated' 'to defeat'

4.2 Analysis

The statistical analysis was performed in R (R Core Team, 2023). Logistic regression models were used (Jaeger, 2008), because the dependent variable, *RC Position*, was binomial: whether a relative clause was adjacent (0) or extraposed (1).

The analysis was conducted as follows. First, we addressed our research question about the role of argument structure by evaluating whether a model with the predictor *Argument Type* provided a better fit to the data as compared to a model that included only the factors previously studied in the extraposition literature: *RC Length*, *[extraposition] Distance*, *Host NP Function* (subject/object), and *Verb Type* (transitive/unergative/unaccusative). Thus, two models were compared: a model with these four factors vs. a model that included these four factors plus *Argument Type* (external/internal). Model comparisons were performed using Chi-Square tests and the Akaike information criterion, or AIC (Akaike, 1998). AIC quantifies the change in the goodness of a model fit as a result of changing the number of predictors: a lower AIC indicates a better model fit (Navarro, 2019).

The model comparison revealed that the model with the predictor *Argument Type* provided a better fit to the data. Thus, its output is presented as the main model in 4.3.3. In this model, all categorical predictors were sum-coded. Numeric predictors were centered and coded as continuous. Note that due to the inclusion of the factor *Argument Type*, the main model could only be run on the corpus sentences that had a main clause verb with a clear argument structure (i.e., sentences with BE-HAVE statives, existentials or nominal copular verbs as main clause were excluded). Random effects were not used, because there was no relevant information – such as text style – in the corpora.

Since the first model included only the subset of the data whose argument structure could be unambiguously determined, a second or exploratory model was built to include the data without a clear argument structure – i.e., predicate nominals, BE-HAVE statives, and existentials – as well as to diagnose the potential contrast between appearance and non-appearance verbs. All predictors of the main model were kept, except for *Argument Type*. In the exploratory model, the factor *Verb Type* had the levels 'transitive/unergative/unaccusative/existential/stative'. The levels 'copula' for *Verb Type* and 'nominal predicate' for *Host NP Function* had to be removed, due to convergence issues. This was due to a "complete separation" or "quasi-complete separation" issue in the data, which happens when an outcome is absent or very rare in the data (Albert & Anderson, 1984). In our case, for sentences with a nominal copular predicate, the subject

noun phrase never appeared with an extraposed relative clause (0 out of 118 tokens), while the predicate nominal only appeared in 0.5% of cases with an adjacent relative clause (2 out of 428 tokens).

Additionally, the factors *Verb Voice* (active/passive) and *Semantic Type* (non-appearance/appearance) were entered into the exploratory model. Note that verb voice and semantic type are meaningful only for transitive and unaccusative verbs, respectively. Therefore, they were entered into the model as nested variables for the relevant verb types. The data and analysis code are publicly available at https://osf.io/6fmtz.

4.3 Results

Out of 3,015 relative clauses, 1,942 (64%) were adjacent, and 1,073 (36%) were extraposed. The results of the main and exploratory models are described below. For ease of interpretation, the figures show descriptive summaries of the data in the percentage scale.

4.3.1 Main model

To evaluate the usefulness of *Argument Type* as a predictor of extraposition likelihood, we compared a full model with all the predictors described in 4.2 with a model that differed only in that it lacked the factor *Argument Type*. The model with the factor *Argument Type* substantially lowered AIC, showing that it provided a better fit to the data: 1984.1 vs. 1991.8. An analysis of variance using a Chi-Square test confirmed that the model with argument type was significantly better (*Deviance* = 9.652, *Df* = 1, *p* = 0.002). Thus, this model was used as the main model. It examined the likelihood of extraposition as a function of argument structure in addition to grammatical function, verb type, and length-related factors (**Table 1**). The results showed a main effect of *RC Length* and a negative main effect of *Distance*: increasing *RC Length* and decreasing *Distance* increased the likelihood of extraposition. These effects match well with the descriptive results: extraposed relative clauses were, on average, 2.7 words longer than adjacent relative clauses. The (potential extraposition) distance was on average 3.7 words shorter for extraposed relative clauses than for adjacent relative clauses (**Figure 1**).

Crucially, the main model showed a main effect of *Argument Type*, with a higher likelihood of extraposition from internal arguments than external arguments. This demonstrates that argument structure affected extraposition rates. **Figure 2** shows that for verbs with an argument structure, rates of extraposition increased for internal arguments (subjects of passive and unaccusative verbs, as well as objects of transitive verbs) compared to external arguments (subjects of active transitive and unergative verbs). No significant effect of *Verb Type* or *Host NP Function* was found. This means that once *Argument Type* was entered into the model, there was no evidence that the likelihood of extraposition was further modulated by either of these factors.
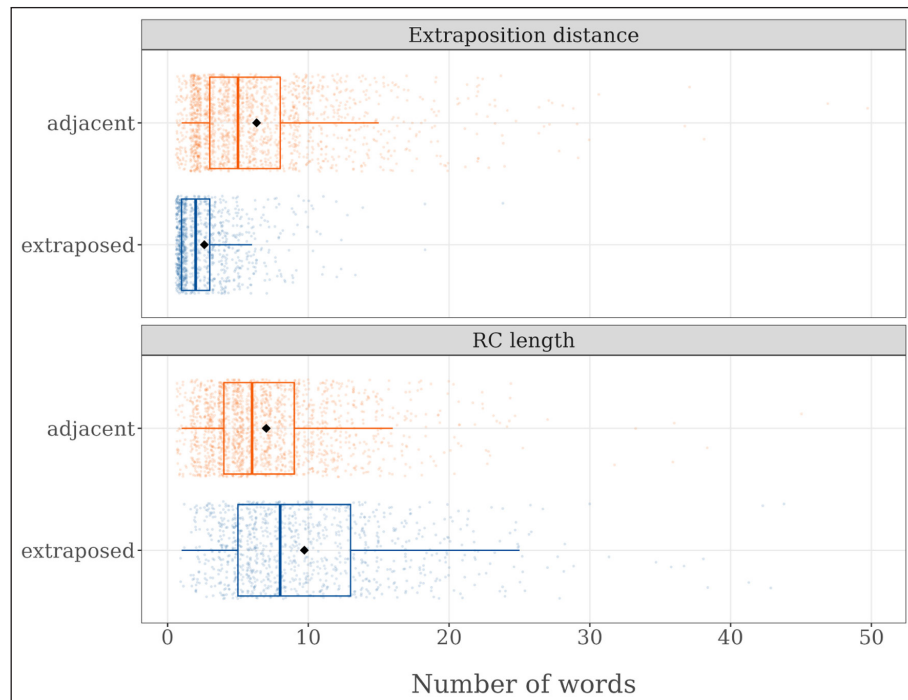
Figure 1: Descriptive summary of the distribution of extraposition distances and relative clause lengths in adjacent vs. extraposed clauses. Diamonds display mean values. Each dot represents a corpus token.

Table 1: Estimates of the main model in log odds. The intercept represents the likelihood of extraposition at the average relative clause length and extraposition distance for verbs that have an argument structure – this corresponds approximately to a probability of 12.56%. Positive estimates reflect an increase in the likelihood of extraposition. Model formula: *RC Position ~ RC Length + Distance + Host NP Function + Verb Type + Argument Type.*

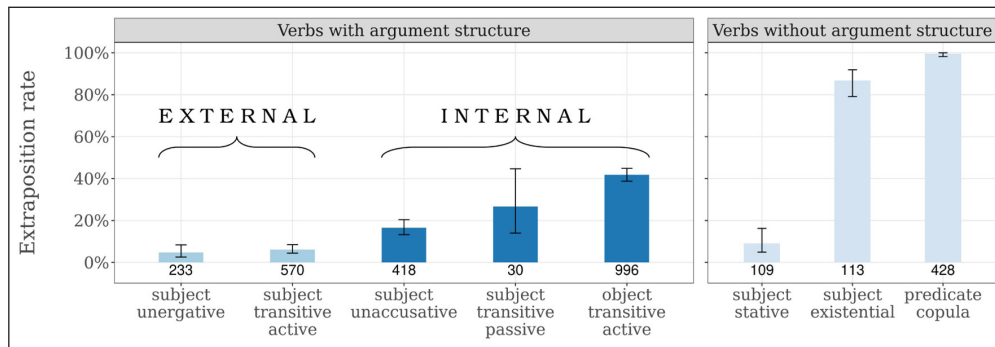| Coefficient	Estimate	Std. Error	z value	Pr($>$|z|)
(Intercept)	-1.940	0.191	-10.144	<0.001*
RC Length	0.117	0.012	9.881	<0.001*
Distance	-0.113	0.019	-6.006	<0.001*
Host NP Function object–subject	0.688	0.439	1.565	0.118
Verb Type unaccusative–unergative	-0.784	0.976	-0.804	0.422
Verb Type transitive–unergative	0.860	0.581	1.480	0.139
Argument Type internal–external	1.986	0.174	11.429	<0.001*

Figure 2: Descriptive summary of extraposition rates from different argument types. The numbers below the bars indicate the number of corpus tokens. Error bars show binomial confidence intervals. Extraposition rates from subjects of copulas are not shown, because they were absent from the corpus.

4.3.2 Exploratory model

The exploratory model was run to additionally include corpus sentences that could not be assigned a clear argument structure – i.e., predicate nominals, BE-HAVE statives, and existentials – and to test for potential differences between appearance and non-appearance verbs. The results of the exploratory model showed that increasing *RC Length* and decreasing *Distance* increased the likelihood of extraposition, similar to the results of the main model (**Table 2**).

Table 2: Estimates of the exploratory model in log odds. The intercept represents the average likelihood of extraposition at the mean relative clause length and extraposition distance for subject nouns of active transitive verbs – this corresponds approximately to a probability of 6.04%. Positive estimates reflect an increase in the likelihood of extraposition. Model formula: *RC Position ~ RC Length + Distance + Host NP function + Verb Type + Verb Type: Verb Voice + Verb Type: Semantic Type.*

Coefficient	Estimate	Std. Error	z value	Pr($>$|z|)
(Intercept)	-2.770	0.182	-15.226	<0.001*
RC Length	0.120	0.012	10.372	<0.001*
Distance	-0.123	0.019	-6.550	<0.001*
Verb Type unergative	-0.311	0.365	-0.852	0.394
Verb Type unaccusative	0.762	0.244	3.119	0.002*

(Contd.)

Coefficient	Estimate	Std. Error	z value	Pr($>$|z|)
Verb Type stative	0.284	0.390	0.730	0.465
Verb Type existential	4.161	0.342	12.182	$<0.001*$
Host NP Function object	2.269	0.197	11.502	$<0.001*$
Verb Type transitive: Verb Voice passive	1.593	0.469	3.398	0.001*
Verb Type unaccusative: Semantic Type appearance	1.204	0.309	3.898	$<0.001*$

With regard to the *Host NP Function*, extraposition rates were the highest from nominal predicates of copular constructions, followed by objects and subjects (**Figure 3**). The exploratory model showed that the likelihood of extraposition increased when the noun hosting the relative clause was a grammatical object compared to a grammatical subject.

Concerning *Verb Type*, the descriptive results suggested that extraposition from subjects was modulated by the type of predicate: the highest rate of extraposition was found with existentials, followed by passives, unaccusatives, BE-HAVE statives, active transitives, and, finally, unergatives (**Figure 3**). The model showed that the likelihood of extraposition from subjects increased when the verb was unaccusative non-appearance, unaccusative appearance, existential, or passive, as compared to when it was active transitive.

Finally, extraposition was more likely with appearance than non-appearance verbs: 32% vs. 13%, respectively. Nested comparisons between these two semantic classes supported an increase in the likelihood of extraposition for appearance verbs.
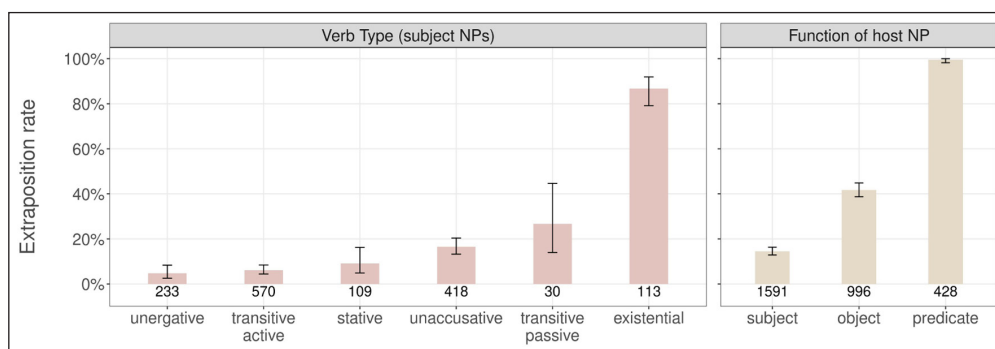

Figure 3: Descriptive summary of rates of extraposition as a function of the grammatical function of the host noun phrase and the verb type. The numbers below the bars indicate the number of corpus tokens. Error bars show binomial confidence intervals. Extraposition rates from subjects of copulas are not shown because they were absent from the corpus.

4.3.3 Overview of the results

The results of the main and exploratory models can be summarized as follows. The main model and its comparison to a model lacking argument structure as a predictor revealed three key findings: (i) extraposition was more likely with increasing relative clause length and decreasing extraposition distance; (ii) argument structure significantly affected the likelihood of extraposition; (iii) including argument structure as a predictor improved the model fit to the data. The exploratory model, which included verbs with and without a clear argument structure, additionally revealed that: (iv) predicate nominals always appeared with extraposition; (v) extraposition was more likely with existentials and appearance verbs than with non-appearance verbs; (vi) subjects of BE-HAVE statives – together with subjects of unergatives and active transitives – showed the lowest rates of extraposition.

5. General discussion

This study examined the factors affecting the likelihood of relative clause extraposition using corpus data from Persian. Our contributions are twofold. First, we replicated previous results from English and German (Bader, 2014; Francis, 2010; Francis & Michaelis, 2014; Shannon, 1992; Strunk, 2014; Uszkoreit et al., 1998): extraposition rates increased with longer relative clauses and shorter extraposition distances. Second, we provided empirical support for a novel idea: that argument structure affects speakers' tendency to extrapose relative clauses. This can potentially provide a unified processing explanation for the effects of grammatical function and verb type attested in the extraposition literature. We demonstrated higher extraposition rates from internal arguments (subjects of unaccusatives and passives, objects of transitives) than from external arguments (subjects of unergatives and active transitives). The subsections below discuss the theoretical implications of these results.

5.1 The role of dependency length: A processing preference

We reported results from two statistical models. Both models found that the rate of extraposition increased with increasing relative clause length and decreasing extraposition distance. Following previous proposals, we assume that the role of length-related factors can be explained by an end-weight processing preference (Arnold et al., 2000; Wasow, 1997a, 1997b). Specifically, longer sentence constituents might be more difficult to process, and thus speakers might opt to postpone them when their grammar allows it. This processing preference has been consistently found with other constructions, such as verb particle movement (Lohse et al., 2004; Wasow, 1997a) and heavy NP shift (Arnold et al., 2000; Wasow, 1997a).

In the case of relative clauses, extraposition might be advantageous because it allows speakers to keep the verb and its arguments close while postponing the planning of another clause to a later point in time. Thus, delaying the relative clause enables speakers to more quickly

integrate the verb and its arguments, which frees up memory resources for the later planning and production of the relative clause. A processing-based explanation of the end-weight preference predicts that the longer the relative clause – and, thus the higher memory burden – and/or the less pre-verbal material to be integrated with the verb, the higher the processing advantage of extraposing the relative clause.

5.2 The role of verb type and grammatical function

In the corpus, relative clauses were extraposed more frequently from grammatical objects than subjects (on average, 42% vs. 15%), as shown by an effect of the factor *Host NP Function object* in the exploratory model. This result is compatible with findings from German (Shannon, 1992). We failed to find an effect of verb type on the likelihood of extraposition from subjects of unergatives and active transitives, but we found a higher likelihood of extraposition with unaccusatives and passives compared to transitive actives. This is also compatible with findings from English (Culicover & Rochemont, 1990; Francis, 2010).

Past research has attributed these patterns to discourse factors, such as givenness/newness and sentence focus. We propose that the emerging pattern for the role of verb type and grammatical function might instead reflect a role of verb argument structure: extraposition is more likely from internal arguments (subjects of passives and unaccusatives and objects) than external arguments (subjects of active transitives and unergatives).

5.3 Argument structure plays a role in extraposition

The evidence supporting our hypothesis that argument structure affects relative clause extraposition comes from the main model, which showed an effect of argument type: internal arguments were more likely than external arguments to host a noun modified by an extraposed relative clause. Furthermore, the model including argument structure as a predictor provided a better fit to the data compared to the model without it.

We propose a production-based processing explanation for this result. This explanation is based on previous studies, which have shown that internal and external arguments are planned differently during language production (Momma & Ferreira, 2021; Momma et al., 2014, 2016, 2018; Schriefers et al., 1998). These studies suggest that internal arguments are planned together with verbs, due to their closer semantic and syntactic relation. Meanwhile, external arguments (as well as other material whose encoding does not depend on the verb) might be planned independently of the verb, at a different point in time. This assumption aligns well with the theoretical literature, in which an external argument is given a special status and sometimes not considered a true argument of the verb (Grimshaw, 1990; Kratzer, 1996; Williams, 1981). By contrast, internal arguments are considered true arguments. This implies that internal arguments have a closer relationship with the verb than external arguments.

We extend the processing explanations above to verb-final structures containing a relative clause. Our account of relative clause extraposition is as follows. First, we assume that verbs and their internal arguments are planned together in speech, thus forming a syntactic dependency. In verb-final structures – where verbs are uttered at the end of a clause – this dependency has to be retained in memory while intervening material is planned and produced – e.g., the relative clause, which can be syntactically complex, because its verb and argument structure can differ from those of the main clause.

Crucially, producing the relative clause in a noun-adjacent position should be more disruptive when the noun phrase is an internal argument, as both the verb and noun were planned jointly and thus are already activated in speakers' memory, as in example (13). In such situations, speakers might opt to first utter the already planned verb, freeing up memory resources for relative clause planning. By contrast, this process is less likely to happen when the relative clause attaches to an external argument, as in example (14), because external arguments do not need to be planned together with the verb. Therefore, extraposition should occur preferentially from internal as opposed to external arguments.

(13) *A RC modifying an internal argument in a Persian verb-final structure*
      mard-i   [_{RC} ke   hič lebās-i   be tan   nadāšt] resid.
      man-INDEF   that no   cloth-INDEF on body had-NEG arrived
      'A man who wasn't wearing any clothes arrived.'

(14) *A RC modifying an external argument in a Persian verb-final structure*
      mard-i   [_{RC} ke   hič lebās-i   be tan   nadāšt] jiq   kešid.
      man-INDEF   that no   cloth-INDEF on body had-NEG scream   pulled
      'A man who wasn't wearing any clothes screamed.'

A testable prediction of this hypothesis is that the effect of argument structure should be greater in relative clause extraposition than in prepositional phrase extraposition. This is because, as argued by Momma & Ferreira (2021), planning a relative clause, which contains a verb, should be more costly than planning a prepositional modifier, which lacks verbal material. Future studies could test this prediction by directly comparing the role of argument structure in the extraposition of relative clauses vs. prepositional modifiers.

5.4 Extraposition in structures with an unclear argument structure

We discuss two categories in this subsection: copular constructions with a predicate nominal and BE-HAVE statives. As for copular constructions with a predicate nominal, the predicate nominal in our data always appeared with an extraposed relative clause, similarly to what Shannon (1992) found in German. Copular constructions, which are often assumed to lack a thematic

structure, due to the reduced semantic content of the copula (Löbel, 2000), show a different pattern of extraposition that seems to be independent of relative clause length, but might be related to the short extraposition distance. Furthermore, the subject of copulas in our data always appeared with an adjacent relative clause. Since these constructions are composed of two noun phrases followed by the linking copula (e.g., *That guy a professor is*), a post-verbal relative clause can seemingly attach only to the closer noun phrase, which is the predicative one, and not the subject. We conjecture that this serves the purpose of ambiguity avoidance. To sum up, copular verbs with a predicate nominal appear to lack argument structure, so argument structure does not seem to play a role in the extraposition of these constructions.

With BE-HAVE statives, our data showed a low rate of extraposition from subjects. Further, there was no evidence that this rate differed from subjects of active transitives and unergatives. Since the argument structure of statives is debated in the literature (Rothmayr, 2009), they were not included in the main model that assessed the role of argument structure. However, we suggest that subjects of stative verbs could be seen as agent-like arguments, and that this is probably why they showed a very low rate of extraposition. This position is similar to the proposal by Levin & Rappaport Hovav (2005) and Rappaport Hovav & Levin (1998), who suggested analyzing stative verbs similarly to unergative verbs such as *run* and *whistle*, which they treat as activities. In this analysis, states and activities are events with a single structural argument: the 'holder' of the state. This argument is similar to the 'actor' of an activity and thus analogous to an agent argument.

5.5 Semantic factors affect extraposition in Persian: Existentials and appearance verbs

Previous studies have shown that extraposition from subjects is more acceptable and more common with unaccusative appearance and existential verbs (Culicover & Rochemont, 1990; Francis, 2010; Francis & Michaelis, 2017; Guéron, 1980). The exploratory model showed that these verbs do show the highest rate of extraposition from subjects. Specifically, existentials showed the highest rate of extraposition from subjects (87%), and appearance verbs showed the second highest rate (32%). This pattern has been associated with the discourse function of extraposition, that is, introducing the subject to the 'world of discourse' (Guéron, 1980, p. 654).

These results underscore the key role that discourse-related factors play in extraposition (Culicover & Rochemont, 1990; Francis, 2010; Francis & Michaelis, 2017; Guéron, 1980; Lee, 2019; Takami, 1999). Our proposal that argument structure affects extraposition should not be interpreted as claiming that syntactic and semantic factors can solely explain how and when speakers decide to extrapose a relative clause in speech. Instead, we believe that an appropriate cognitive model of such decisions should take into account different factors, including: (a)

length-related and prosodic variables; (b) sentence-internal syntactic and semantic relationships; and (c) discourse-related factors, which include the discourse status of the sentence constituents, as well as the communicative intentions of a speaker. In our view, processing considerations related to argument structure are only one piece of the puzzle.

5.6 Limitations and future directions

Our study has limitations, and it also leaves some open questions for future research. First, one factor that we did not examine is definiteness, which features prominently in research on extraposition in other languages. Specifically, extraposition rates and acceptability have been found to correlate with the indefiniteness of the antecedent of the relative clause (Culicover & Rochemont, 1990; Guéron & May, 1984; Strunk, 2014; Walker, 2013). This factor could not be properly assessed in our study, because written Persian does not have definite articles, and also because the indefinite marker – the suffix -i – is homophonous with the restrictive relative clause marker. Therefore, the morphological markers in our data did not unambiguously indicate whether a noun was definite or indefinite. But as a preliminary assessment, we examined the distribution of -rā in our data, which is a definite direct object marker DOM. We found only 30% of DOM-marked nouns modified by an extraposed relative clause, while the pattern was reversed for DOM-unmarked nouns, with a 70% rate of extraposition. This result is compatible with previous findings, but to rigorously assess the role of definiteness in Persian, future research should use experimental data, which allows us to construct stimuli controlled for definiteness.

Other limitations of our study relate to the use of written data. It was not possible to explore the role of focus or prosody, which have been reported to play a role in relative clause extraposition (Francis, 2010; Francis & Michaelis, 2017; Shannon, 1992; Takami, 1999). More critically, written corpus data cannot be used to assess the real-time choices of speakers during sentence processing. In addition to the potentially unlimited time available to speakers when writing, they may follow stylistic rules to form constructions that are more frequent in written than spoken language. For example, it has been reported that extraposition is sometimes used as a stylistic strategy to avoid center-embedded relative clauses with two consecutive verbs in sentence-final position (Najafi, 1992). Therefore, the conclusions drawn from our study need additional support from experimentally controlled studies.

Abbreviations

DOM = direct object marker, EZ = Persian *ezafe* construction, INDEF = indefinite marker, NEG = negative, NP = noun phrase, PL = plural, PRF = perfective, PST = past, PTCP = participle, RC = relative clause, SG = singular

Data accessibility statement

The data extracted from the corpus and the analysis code are publicly available at https://osf.io/6fmtz.

Acknowledgements

Competing interests

The authors have no competing interests to declare.

Authors' contributions

Nasimeh Bahmanian: conceptualization, data curation, formal analysis (lead), methodology, software, visualization, writing – original draft, writing – review & editing (lead). **Markus Bader**: supervision, writing – review & editing (supporting). **Sol Lago**: supervision, formal analysis (supporting), writing – review & editing (supporting).

Author Affiliations

Nasimeh Bahmanian orcid.org/0000-0002-2561-176X

Markus Bader orcid.org/0000-0002-9765-8970

Sol Lago orcid.org/0000-0002-4966-1913

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). Springer. DOI: https://doi.org/10.1007/978-1-4612-1694-0_15

Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika, 71*(1), 1–10. DOI: https://doi.org/10.1093/biomet/71.1.1

Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language, 76*(1), 28–55. DOI: https://doi.org/10.1353/lan.2000.0045

Bader, M. (2014). Defining distance in language production: Extraposition of relative clauses in German. *Cognitive Processing, 15*(1), S81–S84.

Bader, M. (2015). How prosody constrains first-pass parsing during reading. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor* (pp. 193–216, Vol. 46). Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-12961-7_11

Bader, M. (2024). Relative clause extraposition and information structure. In A. Himmelreich, D. Hole & J. Mursell (Eds.), *To the left, to the right, and much in between: A Festschrift for Katharina Hartmann* (pp. 205–216). Goethe University Frankfurt.

Baltin, M. (2006, January). Extraposition. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax* (pp. 237–271). Blackwell Publishing. DOI: https://doi.org/10.1002/9780470996591.ch25

Bever, T. G., & Sanz, M. (1997). Empty categories access their antecedents during comprehension: Unaccusatives in Spanish. *Linguistic Inquiry, 28*(1), 69–91.

Bolinger, D. (1992). The role of accent in extraposition and focus. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language", 16*(2), 265–324. DOI: https://doi.org/10.1075/sl.16.2.03bol

Chomsky, N. (1970). Remarks on nominalization. In R. Jacobs & P. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 184–221). Ginn.

Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures*. Foris Publications.

Chomsky, N. (1986). *Barriers*. MIT Press.

Culicover, P. W., & Rochemont, M. S. (1990). Extraposition and the complement principle. *Linguistic Inquiry, 21*(1), 23–47.

Dechaine, R.-M. A. (1993). *Predicates across categories: Towards a category-neutral syntax*.

Dowty, D. R. (1989). On the semantic content of the notion of 'thematic role'. In G. Chierchia, B. H. Partee & R. Turner (Eds.), *Properties, types and meaning. Volume II: Semantic issues* (pp. 69–129). Springer Netherlands. DOI: https://doi.org/10.1007/978-94-009-2723-0_3

Faghiri, P., & Samvelian, P. (2020). Word order preferences and the effect of phrasal length in SOV languages: Evidence from sentence production in Persian. *Glossa: A Journal of General Linguistics, 5*(1), 86. DOI: https://doi.org/10.5334/gjgl.1078

Féry, C. (2015). Extraposition and prosodic monsters in German. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor* (Vol. 46). Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-12961-7_2

Francis, E. J. (2010). Grammatical weight and relative clause extraposition in English. *Cognitive Linguistics, 21*(1). DOI: https://doi.org/10.1515/cogl.2010.002

Francis, E. J. (2021, December). Relative clause extraposition and PP extraposition in English and German. In E. J. Francis (Ed.), *Gradient acceptabilit and linguistic theory* (p. 0). Oxford University Press. DOI: https://doi.org/10.1093/oso/9780192898944.003.0006

Francis, E. J., & Michaelis, L. A. (2014, October). Why move? How weight and discourse factors combine to predict relative clause extraposition in English. In B. MacWhinney, A. Malchukov & E. Moravcsik (Eds.), *Competing motivations in grammar and usage* (pp. 70–87). Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780198709848.003.0005

Francis, E. J., & Michaelis, L. A. (2017). When relative clause extraposition is the right choice, it's easier. *Language and Cognition*, *9*(2), 332–370. DOI: https://doi.org/10.1017/langcog.2016.21

Friedmann, N., Taranto, G., Shapiro, L. P., & Swinney, D. (2008). The leaf fell (the leaf): The online processing of unaccusatives. *Linguistic Inquiry*, *39*(3), 355–377. DOI: https://doi.org/10.1162/ling.2008.39.3.355

Futrell, R., Levy, R. P., & Gibson, E. (2020). Dependency locality as an explanatory principle for word order. *Language*, *96*(2), 371–412. DOI: https://doi.org/10.1353/lan.2020.0024

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341. DOI: https://doi.org/10.1073/pnas.1502134112

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain. Papers from the first Mind Articulation Project symposium* (pp. 95–126). MIT Press. DOI: https://doi.org/10.7551/mitpress/3654.003.0008

Göbbel, E. (2013). Extraposition of relative clauses: Phonological solutions. *Lingua*, *136*, 77–102. DOI: https://doi.org/10.1016/j.lingua.2013.07.010

Grimshaw, J. B. (1990). *Argument structure*. MIT Press.

Guéron, J. (1980). On the syntax and semantics of PP extraposition. *Linguistic Inquiry*, *11*(4), 637–678.

Guéron, J., & May, R. (1984). Extraposition and logical form. *Linguistic Inquiry*, *15*(1), 1–31.

Haegeman, L. M. V. (1999). *English grammar: A generative perspective*. Blackwell.

Hartmann, K. (2013, July). Prosodic constraints on extraposition in German. In G. Webelhuth, M. Sailer & H. Walker (Eds.), *Rightward movement in a comparative perspective* (pp. 439–472). John Benjamins Publishing Company. DOI: https://doi.org/10.1075/la.200.16har

Hartmann, K. (2017). PP-extraposition and nominal pitch in German. In C. Mayr & E. Williams (Eds.), *Festschrift für Martin Prinzhorn* (pp. 99–107). Universität Wien: Wiener Linguistische Gazette.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511554285

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199252695.001.0001

Hazout, I. (2004). The syntax of existential constructions. *Linguistic Inquiry*, *35*(3), 393–430. DOI: https://doi.org/10.1162/0024389041402616

Heinzova, P., Carreiras, M., & Mancini, S. (2022). Processing argument structure complexity in Basque-Spanish bilinguals. *Language, Cognition and Neuroscience*, *38*(5), 745–763. DOI: https://doi.org/10.1080/23273798.2022.2154370

Huck, G. J., & Na, Y. (1990). Extraposition and focus. *Language*, *66*(1), 51–77. DOI: https://doi.org/10.2307/415279f

Jackendoff, R. (1977). *X̄-syntax: A study of phrase structure*. MIT Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. DOI: https://doi.org/10.1016/j.jml.2007.11.007

Jaeger, T. F., & Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass*, *3*(4), 866–887. DOI: https://doi.org/10.1111/j.1749-818X.2009.00147.x

Karimi-Doostan, G. (1997). *Light verb constructions in Persian*.

Kathol, A., & Pollard, C. (1995). Extraposition via complex domain formation. *Proceedings of the 33rd annual meeting of the Association for Computational Linguistics*, 174–180. DOI: https://doi.org/10.3115/981658.981682

Kauschke, C., & Stenneken, P. (2008). Differences in noun and verb processing in lexical decision cannot be attributed to word form and morphological complexity alone. *Journal of Psycholinguistic Research*, *37*(6), 443–452. DOI: https://doi.org/10.1007/s10936-008-9073-3

Kauschke, C., & von Frankenberg, J. (2008). The differential influence of lexical parameters on naming latencies in German. A study on noun and verb picture naming. *Journal of Psycholinguistic Research*, *37*(4), 243–257. DOI: https://doi.org/10.1007/s10936-007-9068-5

Kayne, R. S. (1994, December). *The antisymmetry of syntax*. MIT Press.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627–645. DOI: https://doi.org/10.1023/A:1026528912821

Kratzer, A. (1996). Severing the external argument from its verb. In J. Rooryck & L. Zaring (Eds.), *Phrase structure and the lexicon* (pp. 109–137). Springer Netherlands. DOI: https://doi.org/10.1007/978-94-015-8617-7_5

Lee, S. H. (2019). Why is English relative clause extraposed? A discourse-based statistical approach. *Linguistic Research*, *36*(2), 213–240. DOI: https://doi.org/10.17250/khisli.36.2.201906.003

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.

Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax-lexical semantics interface*. MIT Press.

Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511610479

Levy, R., Fedorenko, E., Breen, M., & Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, *122*(1), 12–36. DOI: https://doi.org/10.1016/j.cognition.2011.07.012

Löbel, E. (2000). Copular verbs and argument structure: Participant vs. non-participant roles. *Theoretical Linguistics*, *26*(3), 229–258. DOI: https://doi.org/10.1515/thli.2000.26.3.229

Lohse, B., Hawkins, J. A., & Wasow, T. (2004). Domain minimization in English verb-particle constructions. *Language*, *80*(2), 238–261. DOI: https://doi.org/10.1353/lan.2004.0089

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*(6), 485–495. DOI: https://doi.org/10.3758/BF03198265

Mahootian, S. (2002). *Persian*. Taylor & Francis. DOI: https://doi.org/10.4324/9780203192887

Momma, S., & Ferreira, V. S. (2021). Beyond linear order: The role of argument structure in speaking. *Cognitive Psychology*, 101397. DOI: https://doi.org/10.1016/j.cogpsych.2021.101397

Momma, S., Slevc, L. R., & Phillips, C. (2016). The timing of verb selection in Japanese sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 813–824. DOI: https://doi.org/10.1037/xlm0000195

Momma, S., Slevc, L. R., & Phillips, C. (2018). Unaccusativity in sentence production. *Linguistic Inquiry*, *49*(1), 181–194. DOI: https://doi.org/10.1162/LING_a_00271

Momma, S., Slevc, R., & Phillips, C. (2014). The timing of verb selection in English active and passive sentences. *Proceedings of MAPLL: Mental Architecture for Processing and Learning of Language*.

Najafi, A. (1992). *Ghalat nanevisim [let's not write incorrectly]*. Markaz-e Nashr-e Daneshgahi.

Navarro, D. (2019). *Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6.1)*. https://learningstatisticswithr.com/book/

Perlmutter, D. M. (1978). Impersonal passives and the Unaccusative Hypothesis. *Annual Meeting of the Berkeley Linguistics Society*, *4*(0), 157–190. DOI: https://doi.org/10.3765/bls.v4i0.2198

Perlmutter, D. M., & Zaenen, A. (1984). The indefinite extraposition construction in Dutch and German. In D. M. Perlmutter & C. Rosen (Eds.), *Studies in relational grammar 2* (pp. 171–216). University of Chicago Press.

Pesetsky, D. M. (1982). *Paths and categories*. MIT Working Papers in Linguistics.

Poschmann, C., & Wagner, M. (2016). Relative clause extraposition and prosody in German. *Natural Language & Linguistic Theory*, *34*(3), 1021–1066. DOI: https://doi.org/10.1007/s11049-015-9314-8

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rappaport Hovav, M., & Levin, B. (1998). Building verb meanings. In M. Butt & W. Geuder (Eds.), *The projection of arguments: Lexical and compositional factors* (pp. 97–134). CSLI Publications.

Rasekh-Mahand, M., Alizadeh-Sahraie, M., & Izadifar, R. (2016). A corpus-based analysis of relative clause extraposition in Persian. *Ampersand, 3*, 21–31. DOI: https://doi.org/10.1016/j.amper.2016.02.001

Rasooli, M. S., Safari, P., Moloodi, A., & Nourian, A. (2020). The Persian dependency treebank made universal. https://arxiv.org/abs/2009.10205

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*(1), 107–142. DOI: https://doi.org/10.1016/0010-0277(92)90041-F

Rothmayr, A. (2009). *The structure of stative verbs*. John Benjamins Publishing Company. DOI: https://doi.org/10.1075/la.143

Rothstein, S. (1999). Fine-grained structure in the eventuality domain: The semantics of predicative adjective phrases and *be. Natural Language Semantics, 7*(4), 347–420. DOI: https://doi.org/10.1023/A:1008397810024

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*(1), 86–102. DOI: https://doi.org/10.1016/0749-596X(90)90011-N

Schriefers, H., Teruel, E., & Meinshausen, R. M. (1998). Producing simple sentences: Results from picture–word interference experiments. *Journal of Memory and Language, 39*(4), 609–632. DOI: https://doi.org/10.1006/jmla.1998.2578

Seraji, M. (2015). *Morphosyntactic corpora and tools for Persian*.

Shannon, T. F. (1992). Toward an adequate characterization of relative clause extraposition in modern German. In I. Rauch, G. F. Carr & R. L. Kyes (Eds.), *On Germanic linguistics: Issues and methods* (pp. 253–282). De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110856446.253

Shetreet, E., Friedmann, N., & Hadar, U. (2010). The neural correlates of linguistic distinctions: Unaccusative and unergative verbs. *Journal of Cognitive Neuroscience, 22*(10), 2306–2315. DOI: https://doi.org/10.1162/jocn.2009.21371

Stowell, T. (1983). Subjects across categories. *The Linguistic Review, 2*(3), 285–312. DOI: https://doi.org/10.1515/tlir-1983-020305

Strunk, J. (2014, October). A statistical model of competing motivations affecting relative clause extraposition in German. In B. MacWhinney, A. Malchukov, & E. Moravcsik (Eds.), *Competing motivations in grammar and usage* (pp. 88–106). Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780198709848.003.0006

Takami, K.-i. (1999). A functional constraint on extraposition from NP. In K.-i. Takami & A. Kamio (Eds.), *Function and structure* (pp. 23–56). John Benjamins Publishing Company. DOI: https://doi.org/10.1075/pbns.59.04tak

Temperley, D. (2007). Minimization of dependency length in written English. *Cognition, 105*(2), 300–333. DOI: https://doi.org/10.1016/j.cognition.2006.09.011

Truckenbrodt, H. (1995). Extraposition from NP and prosodic structure. *North East Linguistics Society*, *25*(1). https://scholarworks.umass.edu/nels/vol25/iss1/34

Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S., & Skut, W. (1998). Studien zur performanzorientierten Linguistik: Aspekte der Relativsatzextraposition im Deutschen. *Kognitionswissenschaft*, *7*(3), 129–133. DOI: https://doi.org/10.1007/s001970050065

Walker, H. (2013). Constraints on relative clause extraposition in English: An experimental investigation. In G. Webelhuth, M. Sailer & H. Walker (Eds.), *Rightward movement in a comparative perspective* (pp. 145–172, Vol. 200). John Benjamins Publishing Company. DOI: https://doi.org/10.1075/la.200.05wal

Wasow, T. (1997a). Remarks on grammatical weight. *Language Variation and Change*, *9*(1), 81–105. DOI: https://doi.org/10.1017/S0954394500001800

Wasow, T. (1997b). End-weight from the speaker's perspective. *Journal of Psycholinguistic Research*, *26*(3), 347–361. DOI: https://doi.org/10.1023/A:1025080709112

Wasow, T. (2013). The appeal of the PDC program. *Frontiers in Psychology*, *4*. DOI: https://doi.org/10.3389/fpsyg.2013.00236

Williams, E. (1981). Argument structure and morphology. *The Linguistic Review*, *1*(1), 81–114. DOI: https://doi.org/10.1515/tlir.1981.1.1.81