

A usage-based account of subextraction effects

Rui P. Chaves & Adriana King
Department of Linguistics, University at Buffalo – SUNY

Abstract

The idea that conventionalized general knowledge – sometimes referred to as a *frame* (Minsky, 1975; Goffman, 1974; Fillmore, 1977a; Langacker, 1987) – guides the perception and interpretation of the world around us has long permeated various branches of cognitive science, including psychology (Schank and Abelson, 1977), linguistics (Lakoff, 1987), and artificial intelligence (Barlett, 1932; Rumelhart, 1975). In this paper we provide experimental evidence suggesting that frames also play a role in explaining certain long-distance dependency phenomena, as originally proposed by Deane (1991). We focus on a constraint that restricts the extraction of an NP from another NP, called *subextraction*, which Deane (1991) claims is ultimately a framing effect. In Experiment 1 we provide evidence showing that referents are extractable to the degree that they are deemed important for the proposition expressed by the utterance. This suggests that the world knowledge that the main verb evokes plays a key role in establishing which referents are extractable. In Experiment 2 we offer evidence suggesting that the acceptability of deep subextractions is correlated with the overall plausibility of the proposition, suggesting that complex structures can evoke complex frames as well, if sufficiently frequent and semantically coherent, and therefore more easily license deeper subextractions.

1 Introduction

The present work focuses on contrasts such as those in (1), which range from impeccable (1a), to passable (1b), to unacceptable (1c). The gradient acceptability of such extractions remains unaccounted for since they were first noted over 50 years ago (Chomsky, 1964; Ross, 1967; Horn, 1974).¹

- (1) a. What_i did you read / write / buy [a book about __i]?
b. ?What_i did you find / open [a book about __i]?
c. *What_i did you have / lose / drop [a book about __i]?

¹An anonymous reviewer notes that examples like *What would you write a book about?* are even more natural than *What did you write a book about?* This contrast serves to foreshadow the account we will defend in this work. In our view, the former example expresses a less likely proposition than the latter: most people don't write books, but anyone can entertain that goal.

We occasionally use ‘_’ to indicate the canonical *in situ* location of the fronted constituent, whenever its position is not trivial to determine, and refer to it as the ‘gap’ without making any assumptions about the existence of traces, or about the exact nature of such filler-gap dependencies.

Extraction phenomena have played a pivotal role in the development of grammatical theory for over half a century, and have rightly been the focus of intense theoretical and experimental research. Such phenomena are not only touchstones for comparing and distinguishing theories of language, but they also offer insights into the interaction of linguistic knowledge and cognition. However, it is unclear whether extraction constraints are due to general cognitive factors, linguistic constraints, or some combination thereof; see for example Sprouse et al. (2012), Hofmeister et al. (2012) and Sprouse and Hornstein (2013) for a recent debate.

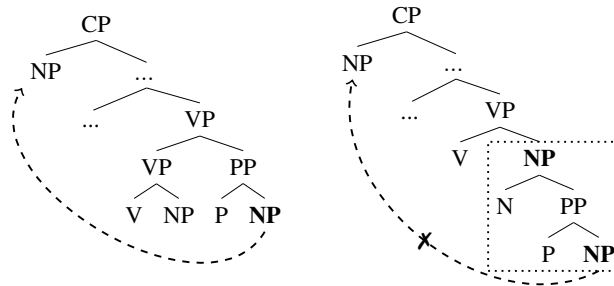
The structure of the paper is as follows. Section 2 provides an overview of previous research and of its shortcomings. One general problem that emerges is that prior studies do not examine subextraction in a systematic manner, nor pay attention to attested data. We sidestep these potential problems by using corpus data to obtain real-world examples (Section 2.3) and by running controlled experiments with larger data sets and much larger pools of naive informants (Section 3). Our overview of previous research suggests some evidence that subextraction effects are likely due to pragmatic constraints (Erteschik-Shir, 1981; Kuno, 1987), which Deane (1991) argues to be ultimately tied to framing effects. The goal of Section 3 is to contribute to this debate by testing these claims with controlled experimentation. In particular, Experiment 1 in Section 3.1 provides evidence that subextraction effects are sensitive to how relevant the extracted referent is for the proposition described by the utterance. In other words, if the extracted referent is what the proposition is about, then extraction is more felicitous pragmatically, all else being equal. This raises the question of why some NP-embedded PPs are more readily interpreted as relevant than others. Drawing from Deane (1991, 1992), our answer is that a referent is relevant to the proposition to the degree to which it is a part of the scene or frame evoked by the main predicate of the sentence. Thus, complex frames that evoke referents corresponding to NP-embedded PPs readily allow those referents to be extracted. In Experiment 2 in Section 3.2 we show that the more felicitous the mention of the NP-embedded PP is, the more acceptable the subextraction. We conjecture that propositions that are rated as highly plausible are more likely to describe routine situations, and therefore are more likely to correspond to a chunked complex background frame. This in turn more readily licenses subextraction because the event participants are more likely part of the same complex frame. As a consequence, no special constraints or subextraction need be postulated in syntax or semantics. Rather, the graded acceptability that emerges in such constructions is a result of a general pragmatic constraint on extraction and the background frame information evoked by the particular utterance.

2 Background

2.1 Syntactic accounts

Syntactic accounts of subextraction have systematically failed to explain the full range of NP Constraint phenomena. The *A-over-A constraint* (Chomsky, 1964: 930) generally banned the extraction of any phrase from another of the same kind, effectively prohibiting all subextraction. Ross (1967) rejected this constraint in light of many counterexamples, but did not propose an alternative. Others like Bach and Horn (1976) and Koster (1978) argued instead that Chomsky’s *A-over-A constraint* is correct, and that subextraction counterexamples were mere illusions. In a nutshell, if the PP-embedded NP is extractable, the PP is in fact adjoined to VP, and therefore no A-over-A constraint violation occurs. If the PP-embedded NP is not extractable, then that must mean that the PP is in fact embedded in another NP and the *A-over-A constraint* correctly rules it out. These two cases are illustrated in Figure 1, respectively. The boxed NP on the right contain a second NP, which cannot therefore be extracted out.

Figure 1: A configurational account of (apparent) *A-over-A* constraint violations



Erteschik-Shir (1981), Godard (1992) and others since have noted that such an account is not tenable. For example, given that (2a) is acceptable, it follows that the NP does not combine with the PP, which would mean that the former should be pronominalizable, contrary to fact.²

- (2) a. What did you read a book about?
 b.*You read it about chemistry.

Second, the fact that such NP PP sequences can be topicalized and conjoined, as in (3), suggests that they form a syntactic unit, by standard constituency tests.

- (3) a. [Books about chemistry], I would never read..!

²Bosque and Gallego (2014) attempt to revive the Bach and Horn account by instead proposing that the complement PP is first extraposed out of the NP, and only then extraction takes place. In other words, the PP is semantically always a dependent of the NP, but its extractability allows an escape hatch to the island constraint. But there is, to our knowledge, no theory-neutral and independently motivated reason for the NP to be extraposable in (1a) but not in (1b).

b. [Books about chemistry or books about physics], I would never read..!

The *Barriers* framework (Chomsky, 1986) prohibited extraction from crossing more than one bounding node. A bounding node is a node that does not receive a thematic role, i.e. a modifier rather than an argument. This again fails to explain the phenomena. In that theory, subextraction from a complement PP as in (1) is allowed, thus permitting too much, and subextraction from a modifier PP like (4) is not allowed, even though such extractions are acceptable.

(4) [One of the libraries in the monastery contains forbidden books]
Which library are we not supposed to read any books from?

More recently, Davies and Dubinsky (2003) propose that extraction is restricted to elements which can be linked to a participant in the lexical conceptual structure of the head, an account that draws heavily from Erteschik-Shir (1981) and Godard (1992). Thus, in Davies and Dubinsky (2003) representational nouns like *picture* allow subextraction because the depicted entity is part of the lexical semantics of the head noun *picture*. Similarly, deverbal nouns like *production* and result nouns like *victory* allow subextraction for the same reason. In contrast, concrete nominals like *cat* and *table* do not allow extraction because they lack even participants in their argument structure. There are various problems with Davies and Dubinsky (2003), however. First, if subextraction phenomena is governed by argument structure then it becomes unclear how contextualization can sometimes ameliorate illicit subextraction as seen in (5) and (6). See Erteschik-Shir (1981) for more examples.

(5) a.*Who did they destroy more pictures of?

b. [Speaker A:] Right after Chairman Mao died, they started taking pictures of Committee members off the walls.
[Speaker B:] Who did they destroy more pictures of, Chairman Mao, or Jiang Qing?
(Kuno, 1987)

(6) a.*Who did you lose a book about?

b. [Speaker A:] I am missing one of my favorite books from my collection of biographies of rock musicians. I think I lost it in the subway during rush hour.
[Speaker B:] Who did you lose a book about? It wasn't Janis Joplin right? I found a book about her yesterday, just one block away.

Second, although Davies and Dubinsky (2003: 16) consider examples like (7) ungrammatical, and take it to provide evidence that only PPs that denote event participants of the head noun can be subextracted, we have been unable to find a native speaker who shares their judgments. At the very least, Davies and Dubinsky (2003) has no way to explain such acceptability variation. In Section 3 we avoid this kind of methodological problem by resorting to controlled experimental studies with a larger pool of randomly selected speakers.

- (7) Which neighbor did Tom chain some dogs of _ to a tree?

Representational nouns like *picture* are ambiguous between the representational (information) sense and the concrete (container) sense, and according to Davies and Dubinsky (2003) the former allows extraction – as in *What did they read / edit books about* – precisely because the book topic is a participant in lexical conceptual structure of this sense of ‘book’, whereas the latter does not – as in **What did they stack / drop books about?* We find this account unconvincing. First, our informants find (7) and (8) to be acceptable, even without contextualization. In such examples it is clear that the relevant sense for the representational nouns is the concrete sense, and therefore Davies and Dubinsky (2003) would predict these subextractions to be impossible.

- (8) a. Which politician did you burn the picture of _ on national TV?
b. What did you say you found a cool book about _ in the attic?
c. What did the Nazis burn books about _ during WW2?

Second, Davies and Dubinsky’s claim that representational nouns like *picture* and *book* are ambiguous between abstract and concrete senses is in fact quite problematic. As Pustejovsky (1995), Copestake and Briscoe (1995), Antunes and Chaves (2003) and Asher (2011) note, both senses exist simultaneously, in the same word token. We illustrate this with (9), where the same occurrence of the two senses of the noun ‘book’ are predicated simultaneously. We conclude that Davies and Dubinsky (2003) raises more problems than it solves.

- (9) a. The book you spilled coffee on is absolutely hilarious.
b. Mia loves that boring and dirty math book.
c. That book is filled with typos and coffee stains.

2.2 Processing accounts

Marcus (1980), Pritchett (1991), Kluender and Kutas (1993), Hofmeister and Sag (2010) and others propose that subadjacency effects – which include subextraction effects – stem from cognitive limitations. But as Fodor (1985) points out, such misparsing phenomena cannot explain subextraction effects given that such sentences are rather short and therefore unlikely to overtax the language processor to such extremes. Speakers routinely recover from failed parses during language comprehension.

That said, the strongest evidence for the role of processing limitations in subextraction comes from ‘deep’ subextraction violations like those in (10), where NP extraction crosses not one but two NP nodes. The reported judgements are the original authors’.

- (10) a. *Who did you hear [stories about [pictures of _]]?
(Chomsky, 1973)
b. *Who did you write [articles about [pictures of _]]?
(Bach and Horn, 1976: 274)

- c. *??Who did you take [a photograph of [a statue of _]]?
(Fodor, 1983: 190)
- d.*Who did you see [enemies of [friends of _]]?
(Bošković, 2017)

The ban on deep subextraction is not robust either (Ross, 1967; Chomsky, 1973). Compare the oddness of (10d) the more acceptable counterpart in (11). In our view, the main difference between the two examples is that seeing enemies of friends of *X* is a far less plausible situation than phoning a friend of a friend of *X*. The former is not something that happens routinely in the real world, but the latter is.

- (11) ?Who did you phone [a friend [of a friend of _]]?

A sample of counterexamples from Deane (1992) is shown in (12), to which we add the attested spoken utterance in (13). In Section 3.2 we provide experimental evidence confirming that some deep subextractions can be deemed acceptable.

- (12) a. Who did you write [a book about [the impeachment of _]]? ?
- b. Which recent president did the House of Representatives lack [enough votes for [the impeachment of _]]?
- c. Which newspaper did you say that the editor exercises [strict control over [the publication of _]]?
- d. Which committee were you offered [an appointment to [the chairmanship of _]]?
(Deane, 1992)

- (13) Which segment do you think it is [time for [another edition of _]]?
[*The Tonight Show Starring Jimmy Fallon*; 2014]

Although it is possible that the recursive NP embedding constructions and long-distance dependencies interact to create processing difficulty, it is unclear how the range of acceptability contrasts discussed so far can be explained by processing limitations. Moreover, it is unclear how a processing account can explain the sensitivity to the verb's semantics as in (1) and to the discourse context as in (5) and (6).

2.3 Semantic/Pragmatic accounts

Bolinger (1972), Cattell (1979), and Kuno (1987) argue that subextraction requires the head noun of the object to denote an attribute, characteristic or meronym of the extracted PP-embedded nominal. Thus, (14a) is licit because 'name' is an attribute of 'person', whereas (14b) is illicit because 'person' is not an attribute of 'name'.

- (14) a. Which person have you forgotten the name of?
- b.*What name have you forgotten a person with?

But upon a closer inspection it is unclear how this constraint can be maintained. For example, in what sense is ‘book’ an attribute or characteristic of its content in a sentence like *who did you read a book about?* In (15) we provide corpus evidence showing that a much wider range of relations are allowed in subextraction, drawn from COCA (Davies, 2008). We conclude that something else must be responsible for the contrast in (14).

- (15) a. (...) extended the Internet tax moratorium, which I was [a fan of _].
 b. (...) a creative photo-retouching program that Gurney was just starting to get [the hang of _].
 c. (...) Lucky for you I’m not the kind who minds being made [a fool of _].
 d. (...) it also helps me practice my patience, which I don’t have [a lot of _].
 e. (...) possibility of contaminated oxygen being in the plane. What would that have been [the result of _]?
 f. (...) Andrew Cuomo is a leading rival because of who he is [the son of _].

Various authors have also argued that general pragmatic constraints restrict subextraction. For example, Erteschik-Shir and Lappin (1979) and Erteschik-Shir (1981) propose that an extractable phrase must be ‘dominant’:

- (16) DOMINANCE: “a constituent *c* of a sentence *S* is dominant in *S* if and only if the speaker intends to direct the hearer’s attention to the semantic content of *c*, by uttering *S*.”
 (Erteschik-Shir and Lappin, 1979: 43)

Analogously, Kuno (1987: 23) proposes that ‘only those constituents in a sentence that qualify as the topic of the sentence can undergo extraction’. Consider for illustration the contrast in (17). As Kuno (1987: 23,121) points out the content of a book is more important for the (volitional) act of writing a book than it is for the (non-volitional) act of losing a book.

- (17) a. Who did you write a book about?
 b.*Who did you lose a book about?

Indeed, even non-extracted examples like (18) already show some acceptability contrast precisely when the extracted nominal is irrelevant for the situation type described by the main predication (Kuno, 1987: 23,121).³

- (18) a. Speaking of Napoleon, I just read a book about him.
 b. ?Speaking of Napoleon, I just dropped a book about him.
 c. Speaking of fur, Sam was looking for a dog with SHORT FUR.

³Takami (1992) claims the *Speaking of X* test suffers from some shortcomings, however.

d. ?Speaking of fur, Sam was running from a dog with SHORT FUR.

Takami (1992) formulates a complementary constraint in terms of informational focus, i.e. information that is new or important. For Takami, extraction is restricted to the part of the utterance that carries new or more important information. That is, extraction targets a phrase denoting a referent x such that the speaker assumes the hearer cannot predict that x would have the particular role that x has in the proposition.⁴ In sum, the consensus between these authors is that only referents that are important (i.e. highly relevant, or ‘topical’ in Kuno’s terminology) for the proposition are extractable. But the question becomes: why are some referents more important/topical than others? For example, why is the object-embedded NP in examples like (14a) and (17a) more topical for the predication than (14b) or (17b)? Deane (1991) argues that the answer lies in framing effects, to which we now turn.

2.4 Framing effects

Deane’s (1991,1992) goal is to explain why certain phrases are more dominant or topical than others. Following the aforementioned pragmatic accounts, Deane assumes that subextraction is licit when (i) the extracted NP is potentially topical (and therefore commands attention); (ii) the matrix phrase for extraction is the information focus (hence, commands attention); and (iii) the remainder of the utterance corresponds to presupposed or given information. For Deane (1992), egocentric, agentive, and concrete referents are more prototypically entrenched than indefinites, abstract referents, or events, and therefore are unlikely to qualify as intrinsic information foci, especially when combined with PP, VP or clausal sisters. The latter is what blocks subextraction in general. Topics and information foci attract more attention, and therefore, are more extractable.

According to Deane (1991: 47) and Deane (1992: 247), subextractions are licensed because of framing effects. Frames (Minsky, 1975; Goffman, 1974; Fillmore, 1977a, 1982; Langacker, 1987), are conventionalized knowledge about interrelated signs that provides a prototypical description of a given scene and guides the perception and interpretation of the world around us. Frames are sometimes also referred to *schemata* (Barlett, 1932; Rumelhart, 1975), *idealized cognitive models* (Lakoff, 1987), or *scripts* (Schank and Abelson, 1977). For Deane, subextraction is possible when the relevant semantic frames are cued before the extraction structure is processed. Such subextractions are licensed because the portions of the sentence that are supposed to express presupposed or given information correspond to information given in the background frame. Thus, subextraction is acceptable to the extent that the background frame recapitulates information about the topic. For example, a claim is the sort of thing that one makes concerning certain topics, and therefore (19) is in a sense redundant. As comprehenders process *make a claim* they are also activating, through inference, the existence of a claim, and therefore the subextraction is facilitated.

(19) Who did you make a claim about?

⁴See also Grosu (1981: ch.5) for a comparable account based on ‘illocutionary unit’.

If the verb and the noun are not as intimately associated with each other, subextractions are less acceptable, as in (20). Clearly, claims are usually heard, but the link between *hear* and *claim* is not as automatic as the link between *make* and *claim*.

(20) ?*Who did you hear a claim about?

Similarly, Deane (1992: 245) argues that long-distance extractions like (21a) are more felicitous than those in (21b) plausibly because the verbs *say* and *think* are quite abstract and high frequency, and therefore the information focus can shift to the complement clause. In contrast, *deny* or *doubt* are less abstract and low frequency, and therefore compete with the embedded clause for the status of informational focus.⁵

- (21) a. Who did you say / think / hear [that Robin was writing [a book about _]]?
b. ?*Who did you deny / doubt / hiss [that Robin was writing [a book about _]]?

Other aspects of Deane's account are more problematic, however. Deane claims that his account predicts spreading activation in visual schemas, in the parietal lobe, during language processing. However, Kemmerer (1998) finds no neurolinguistic supporting evidence for that. More recently, Sprouse (2007: 63–68) measured the effect of supporting contextualization in the acceptability of sentences from Deane (1991), and found it to make no difference. Surprisingly, all items, including those starred by Deane, were deemed highly acceptable. To be sure, Kemmerer and Sprouse's null effects may be due to experimental design issues or too course-grained methodology.⁶ In the remainder of the present work, we focus on framing effects and set aside other aspects of Deane's account. We argue that (sub)extractions are acceptable to the degree that they recapitulate information from the frame evoked by the predication, with acceptability decreasing as the sentence evokes information less central to the frame. In this view, the frame itself is key for explaining the phenomena, as only the referents that are referenced by the frame are potential topics and therefore extractable.

Referents that are not part of the frame are by definition not relevant to the situation that the frame corresponds to, and therefore not possible topics, unless context succeeds to establish them as such. Such referents are more circumstantial for the overall assertion, and their very mention may constitute a Gricean violation, regardless of extraction. For example, (22) is infelicitous because the location *x* that is nearby the possessum bears no obvious relation to the action of selling that the main verb describes.

(22) *Which house did you sell the car near?

⁵For a similar stance see also Van Valin (1995), and for related semantic approaches to such 'bridge' verb phenomena see also Erteschik-Shir (1973), Kuno (1987), Kuno and Takami (1993), Culicover and Jackendoff (2005: 335), and Goldberg (2006: ch.7).

⁶For example, Sprouse (2007) did not norm the experimental items to ensure that the critical items in the supporting context condition did in fact cause the extracted referent to be interpreted as being more topical than the critical items in the non-supporting context condition.

Contrary to Deane’s, the present account has little to do with spreading activation and entrenchment. Rather, the emphasis is placed on the frames themselves. As we shall see, this leads to a different analysis of deep subextraction than Deane’s, as it predicts that for deep subextractions to occur, very complex frames spanning multiple syntactic levels of embedding are present in comprehenders’ grammars.

In the remainder of this work we explore this topic/frame-based account more closely, using controlled experimentation. In Section 3.1 we provide experimental evidence suggesting that the more topical/important the referent is taken to be by speakers the more acceptable its subextraction. This is consistent with the proposal that extraction requires that the fronted expression is what the proposition is about. In Section 3.2 we use sentence plausibility ratings as a proxy for how cohesive the verb and its dependents are. The intuition is that the more plausible a complex proposition is, the more likely it is an instantiation of a complex scene that is frequent in real world experience, and therefore more likely to be a cohesive mental frame. In other words, the assumption is that information which belongs together conceptually is likely to be used together linguistically. We then provide experimental evidence suggesting that the more plausible the proposition is the more acceptable subextraction is as well. The evidence suggests that (deep) subextraction is more acceptable when the construction has been ‘chunked’ as a complex unit of form and meaning and corresponds to a complex frame that is part of speakers’ world knowledge.

3 Experimental evidence

The question to which we now turn is the following. If subextraction is pragmatically licit only to the extent that the fronted element can be interpreted as what the sentence is about, then why are certain object subextractions more licit than others? Drawing from Deane’s work, the answer we explore here is that only referents that are part of the frame evoked by the predication can be potential topics, since only such referents are participants in the conventionalized scene that has been memorized by speakers. The mention of other referents is circumstantial, and therefore not only constitutes a Gricean violation, but cannot be understood as what the proposition is about. If a suitable context is provided that establishes why the otherwise circumstantial referent is relevant to the main assertion, then its subextraction is more felicitous since it is now a more suitable candidate for information focus. As a consequence, however, deeper subextractions entail more complex frames, with complex syntactic and semantic structure. In what follows we offer experimental evidence for this account of subextraction effects.

3.1 Topicality

If indeed some referents are more likely than others to be the topic, then the degree to which they are recognized as such should correlate with the acceptability of extracting such referents. In what follows we examine this prediction experimentally. The overall logic of the experiments detailed below is as follows. We constructed 20 pairs of

declarative sentences like those in (23), and asked informants to rate the acceptability of such sentences using a 1 – 5 Likert scale.

- (23) a. Kayla booked a trip to Paris.
b. Kayla cancelled a trip to Paris.

According to our own intuitions, the PP-embedded NP in (23a) is more important/topical for the overall proposition than that of (23b) because trips are generally booked because of the destination, whereas trips are routinely cancelled for reasons having nothing to do with the destination. Hence, (23a) should be labeled as the +Topic condition and (23b) as -Topic condition. The goal of the (norming) experiment above is to crucially ensure that the sentence pairs in (23) were equally acceptable without extraction, so that any acceptability contrasts created in their subextracted counterparts come from the extraction itself.

Next, in order to verify that the PP-embedded NPs in the +Topic condition were indeed more topical than those in the -Topic condition we used our 20 item pairs to create questionnaire counterpart of those same sentences, as in (24).

- (24) a. How much does the destination of a trip matter when booking a trip?
b. How much does the destination of a trip matter when cancelling a trip?

A different group of informants was asked to answer these questions using a 1 – 5 Likert scale. We interpreted individuals' responses to such questions as an approximate reflection of how important the referent is in the overall proposition. This way, we validate the $-/+$ Topic labelings in (23) but also obtain estimates for the importance/topicality of the PP-embedded referent in the given propositions.⁷

Finally, we created the subextracted counterparts of our 20 items, as (25) illustrates, and asked a third, unique group of informants to rate the acceptability of the sentences. As we shall see, the acceptability ratings obtained for (25) correlate very closely with the topicality ratings obtained for (24). suggesting that topicality is crucial for (sub)extractability.

- (25) a. Where did Kayla book a trip to? (+Topic)
b. Where did Kayla cancel a trip to? (-Topic)

To be clear, we did not use the sentences in (23 – 25) in our experiment, and instead restricted ourselves to speech-act relational nouns like *book*, *story*, *pamphlet*, *conversation*, etc. which select *about*-PPs as (26) illustrates.

- (26) a. Kayla posted / misread a comment about the verdict.
[Declarative sentence; acceptability norming experiment]

⁷The advantage of using the questionnaire format in (24) as opposed to so-called 'topicality tests' like *Speaking of X* discussed in Section 2.3 above is that it arguably offers a more direct measure of the overall importance that comprehenders assign to the referent in question, and sidesteps the concerns raised by Takami (1992) about Kuno's topicality tests.

- b. How much does the topic of a comment matter when posting / misreading a comment?
[Topicality questionnaire; topicality norming experiment]
- c. What did Kayla post / misread a comment about?
[Subextraction sentence; sentence acceptability experiment]

We did this for two reasons. First, *about* is the least ambiguous of the prepositions and therefore avoids potential confounds introduced by other prepositions like *to*, *for* or *on*, in particular, parses where the PP is a VP modifier. Second, the questionnaire (24) counterparts can become awkward and difficult to understand with certain prepositions. For example, although *Who did Robin forget the name of?* is perfectly well-formed, it is unclear what the counterpart *How much does the name of a person matter when forgetting a name?* means. Similarly, *Who did you see a picture of?* is impeccable, but the questionnaire counterpart *How much does the content of a photo matter when seeing a photo?* elicited floor effects in early pilot studies we conducted because the questionnaire sentences were difficult for naive informants to interpret.

3.1.1 Experiment 1

3.1.1.1 Participants

We analyzed data provided by 60 participants with IP addresses originating from the United States that were recruited through Amazon.com’s Mechanical Turk (AMT) crowdsourcing marketplace. For evidence that sentence acceptability data obtained via AMT parallel data obtained in the laboratory see Gibson et al. (2011), Melnick et al. (2011), and Sprouse (2011). By the end of the experiment, all participants self-reported as native speakers of English, and had accuracy levels of at least 75% in comprehension questions, with a mean accuracy level of 95%. Two additional participants whose accuracy scores were lower than the 75% threshold, the data of which was discarded.

3.1.1.2 Design and materials

Using introspection – and drawing some inspiration from verbal entries from FrameNet (Fillmore et al., 2003; Ruppenhofer et al., 2006) – we constructed 20 pairs of experimental items, each of which had two versions, as seen below. The items in the -Topic condition all share in common the fact that the extracted referent is relatively incidental for the predication in the utterance, and contrariwise for the items in the +Topic condition. The complete list of pairs is listed in the Appendix.

- (27) a. What did Kayla post a comment about? (+Topic)
- b. What did Kayla misread a comment about? (–Topic)
- (28) a. What did John direct a documentary about? (+Topic)
- b. What did John narrate a documentary about? (–Topic)
- (29) a. What did Katherine file a complaint about? (+Topic)

- b. What did Katherine overlook a complaint about? (–Topic)

To ensure that any acceptability difference between the above item pairs was caused by extraction rather than by semantic or pragmatic differences, we conducted a norming experiment that measured the acceptability of the declarative counterparts of our 20 items, as illustrated in (30). This task was thus designed to ensure that all non-extracted counterparts of the items were equally highly acceptable to begin with.

- (30) a. Kayla posted/misread a comment about the verdict.
b. John directed/narrated a documentary about drugs.
c. Katherine filed/overlooked a complaint about the company.

Accordingly, a different group of 40 English speakers were asked to rate how ‘natural’ the 20 declarative counterpart items sampled in (30) sounded to them, using a 5-point Likert scale. As before, the items were counterbalanced across two lists, interspersed with 40 distractor sentences, and pseudo-randomized. Half of the distractor sentences were ungrammatical as illustrated in the sample in (31) and the other half was immediately followed by a comprehension question. Each sentence in the experiment was displayed in isolation, one at a time, including comprehension question sentences. Of course, none of the distractors contained the prefix ‘*’ in the actual experiment.

- (31) a.*Tom told around a lie to a friend.
b.*Sam flew Africa to hunt a giraffe.
c. Ashley developed an app for banking.
Q: Ashley probably knows very little about software. [True/False]
d. Bob hit a car with his scooter.
Q: Bob probably lost control of his vehicle. [True/False]

The response accuracy to the comprehension questions ranged from 77% to 100% (91.5% mean accuracy). The rating results indicate that both types of sentences were deemed highly acceptable: for each experimental item, the mean was above 4, with an overall mean of 4.5. The overall mean for +Topic items was 4.7, and the overall mean for -Topic items was 4.2. Pairwise t-tests were used to verify that no significant difference existed between ‘+/- Topic’ item pairs.

Finally, to ensure that the +Topic condition items were indeed more topical than the -Topic condition items, we conducted a second norming experiment in which participants were asked to use a 5-point Likert scale to answer questions like those in (32), again created from our 20 original experimental items.

- (32) a. How much does the topic of a comment matter when posting / misreading a comment?
b. How much does the topic of a documentary matter when directing / narrating a documentary.?

- c. How much does the topic of a complaint matter when filing / overlooking a complaint?

A different set of 60 participants was recruited, and given 20 pairs of items counterbalanced across two lists in a Latin Square design so that each participant only responded to one version of each experimental item. The items were pseudo-randomized with 40 distractor sentences, and displayed one at a time. A sample of distractors is given in (33). Half of distractors concerned ‘goal’ and the other half concerned ‘name’. Some involved *when* subordination, others *for*.

- (33) a. How much does the name of a dessert matter for ordering a dessert?
- b. How much does the name of a pet matter when buying food for a pet?
- c. How much does the goal of a paper matter when writing a paper?
- d. How much does the goal of an errand matter for liking an errand?

An overall t-test revealed that the items in the +Topic condition received higher ratings than those in the -Topic condition ($t = -20.765, p < 0.001$). The overall mean rating for the former was 4.37 (SD = 0.92) and the overall mean for the latter was 2.94 (SD = 1.43). A by-item t-test analysis revealed that all item pairs were statistically different in the same direction, so that no -Topic item received an average rating that was equal or superior to its +Topic counterpart.

3.1.1.3 Procedure

The 20 subextraction items were counterbalanced across two lists using a Latin Square design so that each participant only responded to one version of each experimental item, as in the above experiments. The 20 experimental items were interspersed among 40 distractor items, a sample of which is seen in (34). Half of the distractors were ungrammatical because the post-auxiliary verb was either in the present or in the preterite form. Participants were asked to judge how natural each sentence was, by giving it a number from 1 (very unnatural) to 5 (very natural). In order to ensure that comprehenders attended to the structure and meaning of the experimental items, half of the grammatical distractors were immediately followed by a True/False comprehension question as illustrated in (34c,d).

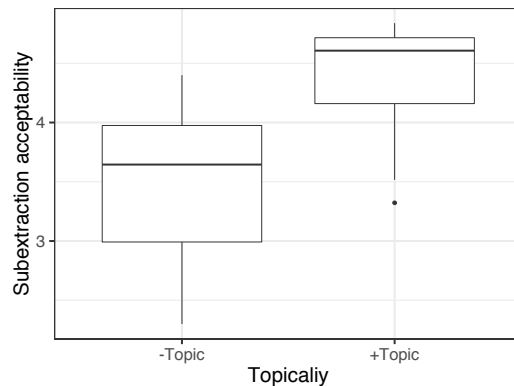
- (34) a.*What did Sue placed his pie on?
- b.*What did Phil listens to the organizer of?
- c. What did Amanda strum a rock ballad on?
Q: Amanda can play a string instrument. [True/False]
- d. Who did Tim expect a follow-up question from?
Q: Tim thought that the in-class discussion was done. [True/False]

?

3.1.1.4 Results

The mean response for the +Topic condition was 4.37 (SD = 0.95), and for the -Topic condition it was 3.5 (SD = 1.27). The mean response for the grammatical distractors was 4.38 (SD = 0.93), and 2.07 (SD = 1.31) for the ungrammatical distractors. A Linear mixed-effect regression (LMER) model with the nominal +/-Topic condition as a fixed factor and acceptability as a predictor was fit using the `lme4` package version 1.1–12 (Bates et al., 2014) in R version 3.3. In all LMER reported in this work, the intercept was allowed to be adjusted by items, subjects, lists, presentation order, and reading time in order to account for random effects. The p -values were calculated by Satterthwaite approximation, using the `lmerTest` package version 2.0-30 (Kuznetsova et al., 2015). The LMER model revealed a main effect of the +/-Topic variable in which -Topic subextractions were rated lower than +Topic subextractions ($\beta = 0.86, t = 15.18, SE = 0.05, p < 0.001$). The results are illustrated in Figure 2.

Figure 2: Object subextraction acceptability according to Topicality



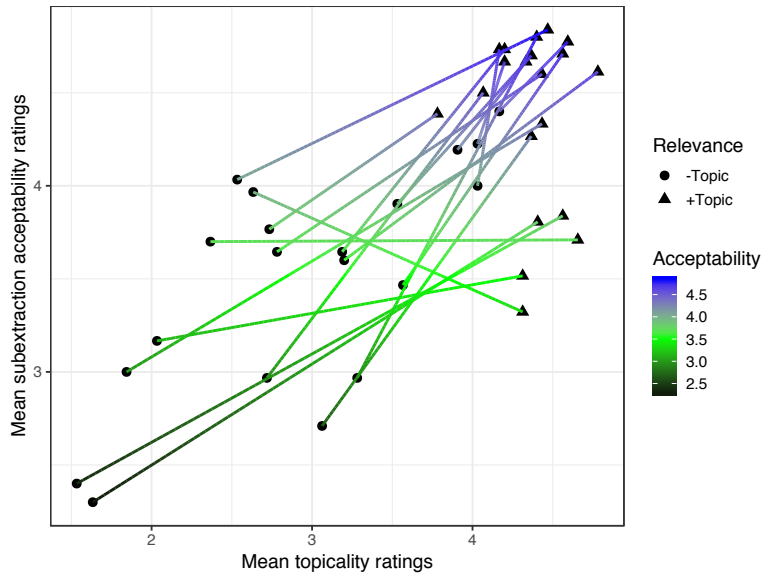
To control for frequency effects, we counted the occurrences of each verb and object noun (V+N) combinations in the experimental items in COCA.⁸ An LMER model with topicality and frequency as fixed effects revealed that the former was significant ($\beta = 0.91, t = 10.24, p < 0.0001$) but not the latter ($\beta = -0.02, t = -0.51, p = 0.6$).

Next, we fitted a simple regression model with the mean subextraction acceptability ratings per item as a dependent variable and the mean topicality ratings per item from the questionnaire experiment as the independent variable. The model suggests that the topicality ratings were a strong linear predictor of the subextraction acceptability ratings ($\beta = 0.56, SD = 0.08, t = 7.04, p < 0.0001$), with an adjusted R^2 value of 0.55. A Pearson test was used to confirm the existence of a strong positive correlation between the topicality ratings and the subextraction acceptability ($r = 0.75, p < 0.0001$). In Figure 3 the mean +/-Topic ratings for each of the sentences is on the x-axis, and the

⁸This was done by running 40 grep searches (2 per item pair) over all the V + N combinations. For example, `\bask(s|ed|ing)? [a-z]* questions?\b` was used to count all occurrences of 'ask / asks / asked / asking' followed by (at most) one word by 'question(s)'.

mean acceptability ratings for the subextraction counterpart sentence is on the y-axis.⁹

Figure 3: By-item acceptability ratings vs. topicality ratings



Ratings on a 5-point Likert scale are arguably ordinal in nature, and so an Ordered Regression Model was also fitted where both raw ratings were coded as ordinals, and frequency was a random effect variable, using the `Ordinal` (Christensen, 2018) R package (ver. 2019.3-9). There was again clear evidence of a correlation, as shown in Table 1 below.

Table 1: Location Coefficients (Experiment 1)

Topicality Rating	Estimate	SE	z	p
2	0.48	0.039	12.21	< 0.0001
3	0.83	0.035	23.5	< 0.0001
4	1.16	0.032	35.85	< 0.0001
5	1.5	0.030	49.03	< 0.0001

⁹The acceptability increased with topicality for all but two item pairs. In the case of *What did Mary proof-read/like a story about?* (-Topic/+Topic) the former was more acceptable than the latter perhaps because *what* is likely an inanimate object, and inanimate objects are not prototypical story topics. As for *What did Scott mistype/discuss a report about?* (-Topic/+Topic) both items received equally moderate acceptability ratings, perhaps because mistyping a report can have important consequences that are tied to the report's content.

3.1.1.5 Discussion

Our results suggest that the acceptability of object subextractions is contingent on an interplay between the verb and the extracted element, such that the more important the latter is for the proposition described by the utterance, the more acceptable the subextraction. This is consistent with extant pragmatic accounts of subextraction effects. The findings also suggest that not all subextractions are equally acceptable, ranging from very low acceptability (e.g. lowest acceptability means of 2.5) to highly acceptable (e.g. highest acceptability means of 4.8), and everything in between. It is therefore unsurprising that some subextractions can be ameliorated with contextualization, and that others less so. Finally, the fact that subextractions are graded and span almost the full range of the scale is not compatible with syntactic accounts, as nothing in them predicts gradient acceptability, but is fully compatible with a pragmatic account provided that the likelihood of an NP-embedded referent being the topic of the assertion is a matter of degree. We now turn our attention to the role of frames in subextraction.

3.2 Frames

3.2.1 Framing extraction

As Metusalem et al. (2012) recently show, general knowledge about events is activated during language processing, and likely drives the generation of linguistic expectations. This is consistent with the idea that linguistic semantics is related to encyclopedic knowledge, without which one cannot understand word meaning. For example the meaning of the word ‘sell’ cannot be understood without knowing about the situation of commercial transfer, which involves not only a seller, a buyer, goods, money, but also the relation between the money and the goods, the relations between the seller and the goods and the money, the relation between the buyer and the goods and the money and so on. Another classic example is the word *widow*, which evokes the knowledge that people marry as adults, that marriage terminates with their partner’s death, and so on (Fillmore, 1977b). Although a given word only corresponds to a small part of its respective frame, the entire frame becomes active when the word is in use. Frames are thus responsible for organizing complex phenomena into coherent, understandable categories, which usually encompass various participants.

Mentioning the content of a book is more natural in a sentence like *I read a book about chemistry* than in a *I dropped a book about chemistry* because the content of a book is relevant for understanding a book-reading scene, but not so much for understanding a book-dropping scene. The former is felicitous out-of-the-blue, but the latter much less so because information about the book content has no bearing on the action triggered by the actor. Hence, the latter arguably constitutes a violation of the Gricean maxims of Quantity (‘Be brief’) and/or Manner (‘Avoid Prolivity’); see Van Valin (2005: 288). Both sentences are well-formed, but the latter requires contextualization in order for the mention of the book content to be justified. All of this makes sense if the content of the text is part of the frame evoked by the *read* frame. As such, it can easily be the most important referent in the proposition. In contrast, there is not reason for the content of a text to be part of the frame evoked by *drop*, and

thus – without appropriate contextualization to that effect, such a referent cannot be the most important referent in the proposition. If Erteschik-Shir (1981), Kuno (1987: 23), and Takami (1992) are correct in the extracted element in a filler-gap dependency must correspond to a referent that is in some sense important for the proposition, then contrasts like (35), follow straightforwardly from the fact that in (35a) the extracted referent is part of the frame evoked by the verb, but not in (35b).

- (35) a. What did you read a book about?
b.*What did you drop a book about?

In our view, the oddness of examples like (36) from Davies and Dubinsky (2003) is precisely because whether or not an article is coffee-stained is irrelevant for whether someone is reading an article about *x*. Thus, the mere mention of coffee stains violates the Gricean maxims of Quantity ('Be brief') and/or and Manner ('Avoid Prolixity'), as they are of little relevance, without a suitable context.

- (36)*Who are you reading a coffee-stained article about?

In sum, subextraction should therefore be acceptable to the degree that the extracted referent is a participant in the background frame, with acceptability decreasing as the sentence evokes information less central to the frame. Thus, the oddness of (35b) is not beyond repair, since contextualization may provide extra information justifying why the referent is not merely worthy of mention but in fact pivotal for the proposition. For example, if it is known that the agent is prone to dropping books about certain topics then a discourse move like (35b) becomes more felicitous. Analogously, (37a) is not as acceptable as (37b) out-of-the-blue because *finish* does not evoke any specific action. Rather, *finishing a book* involves an implicit predicate (typically *read* or *write*) and therefore it is less clear how the mention of the content of the book is relevant for the described situation. If context establishes that Sam always gives up reading biographies but has finally finished one, then (37a) becomes a more felicitous discourse move.

- (37) a. ?Who did Sam just finish a book about?
b. Who did Sam just finish reading a book about?

Depictive relational nouns like *picture* and *photo* are inextricably tied to the subject they depict, and thus subextraction is licit to the extent that the frame evoked by the verb involves the subject that is depicted. For example, in the TAKE PICTURE frame evoked in (38a) the content of the picture is of crucial importance to the action, as pictures are usually taken because of the subject of the picture. Similarly, in (38b) the subextraction is acceptable because seeing a picture entails seeing the subject it depicts. Hence, the depicted subject is a highly prominent participant in the utterance. In contrast, the subextractions in (38c) are slightly less felicitous because the depicted subject is not part of the evoked frame and therefore its relevance is less clear, out-of-the-blue.

- (38) a. Who did you take a picture of?

- b. Who did you see a picture of?
- c. ?Who did you lose a picture of?

The oddness of more extreme examples like (39a) is caused by the title of the book not being a component of the frame evoked by *read*. In addition, there is a tendency to view the PP as an instrumental VP-modifier, rather than the complement of the object, causing a minor garden-path effect. Similarly, (39b) is not licit because the frame evoked by *forget* does not predicate attributes of the theme. That is, people don't usually forget people because of their names. Furthermore, the preposition is again biased towards the VP-modifier use, increasing the difficulty in interpreting this sentence.

- (39) a.*Which title did Sam read a book with?
- b.*What name have you forgotten a person with?

The case of (40) is similar, since the title referent is not part of the frame evoked by *read*. Note in particular that not all reading material has titles, which makes it less likely that the title is a conventionalized component of the overall reading frame. Another contributing factor is that the preposition *with* introduces a VP-modifier ambiguity, of course.

- (40)*Which title did you read a book with?

Let us now focus on more complex examples like (41), involving doubly-NP-embedded extractions. The visual perception frame evoked by *see* does not involve the concepts of friends or enemies (and it is not trivial to construct a context in which this would be so), and therefore the sentence is odd and difficult to rescue via contextualization.

- (41)*Who did you see [enemies of [friends of _]]?
- (Bošković, 2017)

Consider now (42a). The action of casting votes for a given political process is a commonplace routine situation in the real world. As such, casting a vote likely entails the existence of some political process *X* (the theme of the vote), which means that the impeachment referent is central to the state-of-affairs under discussion, and therefore can be the information focus of the assertion. Thus, the *VOTE* frame evoked in (42a) has as one of its components a political act and therefore the corresponding referent can be extracted. Similarly, (42b) is acceptable because *write* evokes a topic, and *history* is a prototypical topic.

- (42) a. Who did you cast [a vote for [the impeachment of _]]?
- (Deane, 1991: 12)
- b. Which institution did you write [a paper about [the history of _]]?

This view of deep subextraction requires us to assume that frames can correspond to otherwise very complex syntactic, semantic, and pragmatic representations. Examples like (13), repeated below as (43), would have to correspond to a complex linguistic chunk (or ‘collostruction’) that evokes a proportionally complex semantic frame.

(43) Which segment do you think it is [time for [another edition of _]]?

Indeed, there is evidence that complex forms are memorized by speakers (Bybee, 2013; Traugott and Trousdale, 2014; Bybee, 2006). In particular, various authors have argued that such syntactic and semantic chunks (even if compositional, like the one above) play an central role in construction-based language acquisition and syntactic processing (Croft, 2001; Abbot-Smith and Tomasello, 2006; Goldberg, 2006). If speakers can memorize complex syntactic and semantic chunks like *write document x about topic y*, then the subextraction of otherwise syntactically embedded referents should be facilitated simply because the referents in question are part of the complex frame. A good example of this is the (attested) subextraction in (43). The sequence *it is time for another edition of* is highly collocational, with about 20,000 attestations in Google.

The same explanation must extend to even deeper subextractions like (44), where the gap is embedded in three NPs, not two. The verbal elements in these examples have very abstract semantics. Rather, it is the nouns that evoke the relevant background frames. The link between *vote*, the votes’ theme, and its patient is rather straightforward, as is the one between *appointment*, its theme *chairmanship* and its patient. These are highly plausible and frequent state-of-affairs, and therefore it is natural that speakers have mental representations of such scenarios.

- (44) a. Nixon was one president that they had [no trouble getting [votes for [the impeachment of _]]].
- b. Which committee did you have [aspirations for [an appointment to [the chairmanship of _ ?]]]
(Deane, 1991)

In what follows we attempt to investigate whether the acceptability of deep subextractions depends on how grammaticalized they are. Ideally, this task would be informed by corpus data, but the longer the construction under investigation the less likely for it to be attested, given the well-known problem of sparseness. Indeed, even large corpora are too sparse to provide a meaningful frequency estimate of complex chunks. For example, although (42b) is perfectly acceptable to us, the pattern ‘write/wrote/writes/writing (DET) paper(s) about (DET) history’ has no attestations in COCA. Thus, we use semantic plausibility ratings as a proxy for grammaticalization, by making the assumption that the more plausible a proposition is, the more likely it corresponds to a frequent and coherent complex scene that is learned and stored as part of world knowledge.

3.2.2 Experiment 2

3.2.2.1 Participants

We analyzed data provided by 54 self-reported native speakers with IP addresses originating from the United States that we recruited through Amazon.com's AMT. As in Experiment 1, the data from 8 participants with comprehension question accuracy scores below the 75% threshold were discarded.

3.2.2.2 Design, materials and procedure

In the norming experiment, 34 native speakers were asked to rate the plausibility of the situation described by sentences like those in (45), in a 1-5 Likert scale. By asking naive native speakers about the plausibility of the situations described by the sentences we hope to elicit a rating that correlates to how commonplace and prototypical the state-of-affairs intuitively is. The assumption is that sentences with more collocational components reflect more complex frames, where the collocates are integral parts of, rather than supplements. There were a total of 16 such sentences (see Appendix B), interspersed with 32 distractor sentences.

- (45) a. The kids heard stories about pictures of this artist.
- b. The farmers question the analysis of the impact of this pesticide.
- c. The editor has strict control over the publication of this content.
- d. The experts receive requests for articles about this topic.

The items were constructed using introspection, taking inspiration from attested examples. Half of the distractors consisted of sentences expressing slightly unusual situations like (46a) and the other half expressed slightly more prototypical situations like (46b). The latter were followed by comprehension questions like *The company that Courtney works could be sued [True / False]*.

- (46) a. The waiter placed his umbrella under the kitchen table.
- b. Courtney showed her employer was negligent.

Declarative items received a mean plausibility rating of 4.62 (SD = 0.73), distractors expressing unusual situations received 2.47 (SD = 1.42), and distractors expressing more prototypical situations 4.68 (SD = 0.76).

In the next experiment, we collected data from a different group of 28 native speakers, which were asked to rate how natural the sub-subextracted counterparts of (45) are, illustrated in (47), again using a 1–5 scale.

- (47) a. Who did the kids hear stories about pictures of?
- b. Who did the farmers question the analysis of the impact of?
- c. What does the editor have strict control over the publication of?

d. What did the experts receive requests for articles about?

As before, there were a total of 16 sentences interspersed with 32 distractors, half of which were ungrammatical (48) illustrates. The grammatical items were as usual followed by comprehension questions (e.g. *The administration thinks that the professor has too few lab members [True / False]*).

(48) a.*Who does the dolphin apparently seem to chasing in the water of?

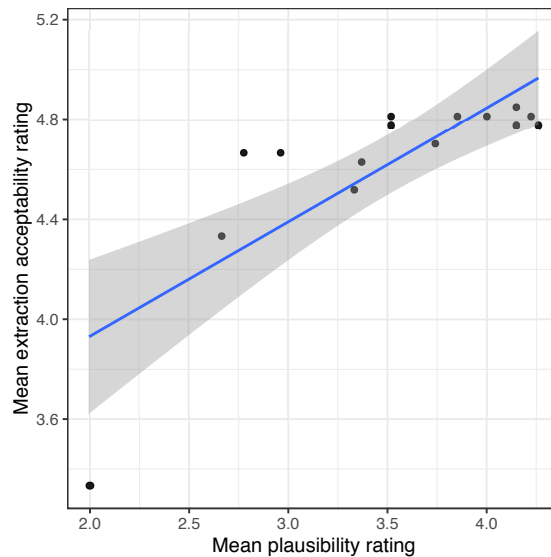
b. Who does the committee think should not hire any more research assistants?

Experimental items had a mean acceptability rating of 3.27 (SD = 1.26), grammatical distractors 4.16 (SD = 1.03), and ungrammatical distractors 2.49 (SD = 1.32).

3.2.2.3 Results

Whereas some deep subextractions like (47a,b) were deemed quite low in acceptability, with a mean of 2, others like (47c,d) are rated high, with a mean of over 4.2. Since the latter means are as high as that of grammatical distractors, this shows that at least some deep extractions are in fact acceptable. If our working hypothesis is correct, then the plausibility ratings for the declarative sentences should strongly correlate with the acceptability ratings for the extracted counterparts. The results are shown in Figure 4.

Figure 4: Subextraction acceptability v.s. Declarative plausibility



Analogously to Experiment 1, a simple regression model was fitted with the mean acceptability ratings per item as a dependent variable and the mean plausibility ratings per item from the questionnaire experiment as the independent variable. The model

suggests that the topicality ratings were a strong linear predictor of the subextraction acceptability ratings ($\beta = 0.2$, $SD = 0.04$, $t = 4.07$, $p = 0.001$), with an adjusted R^2 value of 0.54. A Pearson test further confirms that the plausibility ratings for declarative items like (45) are (highly) correlated with the acceptability ratings for the deep subextractions counterparts in (47): $r = 0.79$, $p < 0.001$. The outlier at the lower left of the graph has a minimal influence on the result, since its removal still allows for a strong correlation ($r = 0.76$, $p = 0.001$).

Like in Experiment 1, an Ordered Regression Model was also fitted with the raw ordinal value for both ratings, and again provided strong evidence for a correlation. The results are shown in Table ???. Removing the items corresponding to the outlier in the lower left corner of Figure 4 reduces the z values but does not change the significance of the results, all of which remain below $p < 0.004$ levels.

Table 2: Location Coefficients (Experiment 2)

Plausibility Rating	Estimate	SE	z	p
2	1.55	0.20	7.41	< 0.0001
3	1.21	0.16	7.43	< 0.0001
4	1.59	0.14	11.13	< 0.0001
5	2.19	0.13	15.69	< 0.0001

3.2.2.4 Discussion

The evidence suggests that there is a positive correlation between the plausibility of the proposition (as expressed by a declarative clause) and the extractability of a deeply object-embedded referent therein. This suggests that the more coherent the semantic components and the more prototypical their relations are, the more acceptable extraction of one of those components is. Thus, in a highly plausible proposition an NP-embedded PP is more likely to be a prototypical participant in the action described by the verb, and therefore, a conventionalized component of the background situational knowledge (frame) evoked by the predication. In contrast, if the NP-embedded PP is rather circumstantial and irrelevant for the action described by the verb then the proposition is not as plausible and it is less likely for the embedded referent to be part of the conventionalized component of the frame knowledge evoked by the predication.

4 Conclusion

In this work we provide novel experimental and corpus evidence in support of the view that extracted elements must come from the informational focus part of the proposition and must correspond to a referent that is deemed highly relevant/important to the proposition (Erteschik-Shir, 1981; Van Valin, 1986; Kuno, 1987; Takami, 1992). Drawing from Deane (1991), we operationalized the latter constraint in terms of the referents that belong to the frame that the main predication evokes: the more important the role that the referent plays in the frame (either directly, as a participant, or

indirectly, through contextualization) the more felicitous is its extraction. We provide supporting experimental evidence of this claim, by using sentence plausibility as a proxy for grammaticalization of complex frames. In particular, we speculated that sufficiently frequent complex syntactic structures can become constructional chunks, reifying their corresponding complex frames in the process. The latter is what enables the deep subextraction of phrases that would otherwise be too syntactically embedded for their referents to be part of the frame evoked by the main verb. Thus, no special constraints or subextraction need be postulated at the level of syntax or semantics. Rather, the graded acceptability that emerges in such constructions is a result of a general pragmatic relevance constraint on extraction (Kuno, 1987) – arguably Gricean in nature (Van Valin, 2005: 288) – and its interaction with the evoked background frame knowledge.

Acknowledgements

We wish to thank the Editors John Newman and Dagmar Divjak, and the anonymous reviewers for their invaluable assistance and feedback. We also wish to thank the Department of Linguistics at the University at Buffalo for an internal research grant that supported the experimental component of this research.

Appendix A

Below are the items used in the acceptability experiment reported in Section 3.1. The verb that appears immediately to the left of ‘/’ causes the extracted element to be more topical for the proposition, and the verb that appears immediately after ‘/’ causes the extracted element to be less topical. The norming (declarative) experimental items were systematically constructed from those below. Each item is followed by the respective COCA frequency counts. For example, the verb *like* (however inflected) is followed by the complement *story* or *stories* (with at most one word in between) 462 times; see footnote 9 above for more details. Add-one Smoothing of the raw counts to avoid zeros had no quantitative or qualitative impact on the statistical results.

- | | |
|--|------------|
| 1. What did Mary like / proof-read a story about? | (462/0) |
| 2. What did John direct / narrate a documentary about? | (34/14) |
| 3. What did Bill create / find a powerpoint about? | (0/0) |
| 4. What did Kate give / reschedule a presentation about? | (182/0) |
| 5. What did Sarah start / transcribe a conversation about? | (364/5) |
| 6. What did Steve ask / hear a question about? | (9066/350) |
| 7. What did Scott discuss / mistype a report about? | (29/0) |
| 8. What did Katherine file / overlook a complaint about? | (750/0) |

9. What did Brian criticize / rip a pamphlet about?	(0/0)
10. What did Chris make / ruin a joke about?	(847/3)
11. What did Melissa rent / rewind a movie about?	(73/3)
12. What did Rich write / see a book about?	(6228/296)
13. What did Isabella buy / misplace a magazine about?	(94/0)
14. What did Frank draft / xerox a movie script about?	(10/1)
15. What did Kim teach / miss a class about?	(702/75)
16. What did Leo read / discard a novel about?	(372/0)
17. What did Nate prepare / hear a speech about?	(43/138)
18. What did Kayla post / misread a comment about?	(484/0)
19. What did Peter create / postpone a workshop about?	(7/0)
20. What did Caitlin take / return a quiz about?	(11/0)

Appendix B

Below are the items used in the two acceptability experiments reported in Section 3.2. The norming (declarative) experimental counterparts were systematically constructed from the items listed below. No verb-noun-noun frequencies were included in the analysis because their occurrence is too sparse to appear even in large corpora like COCA.

1. Which President did pundits create blogs about the impeachment of?
2. Which actress did students write a paper about the photos of?
3. Which article did the publisher change the font in the title of?
4. Which artist did the kids hear stories about pictures of?
5. Which bridge did the Governor order a study about the condition of?
6. Which committee was the Mayor offered an appointment to the chairmanship of?
7. Which company did lobbyists advise the chairman of the board of?
8. Which content does the editor have strict control over the publication of?
9. Which disaster did the church help families of the victims of?
10. Which historical figure did visitors take photos of the statue of?

11. Which institution did researchers write a paper about the history of?
12. Which official did the Senate lack votes for the impeachment of?
13. Which pesticide did farmers question the analysis of the impact of?
14. Which reporter did the lawyers revise the wording of the email to?
15. Which suspect did the witness see enemies of friends of?
16. Which topic did the experts receive requests for articles about?

References

- Abbot-Smith, Kirsten & Tomasello, Michael. 2006. Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review* (23), 275–290.
- Antunes, Sandra & Chaves, Rui P. 2003. On the licensing conditions of co-predication. In *Poster at Second International Workshop on Generative Approaches to the Lexicon*, University of Geneva, Switzerland.
- Asher, Nicholas. 2011. *A Web of Words: Lexical Meaning in Context*. Cambridge University Press.
- Bach, Emmon & Horn, George G. 1976. Remarks on 'Conditions on transformations'. *Linguistic Inquiry* 7, 265–299.
- Barlett, Frederic. 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press, Cambridge.
- Bates, Douglas, Maechler, Martin, Bolker, Ben & Walker, Stevern. 2014. lme4: Linear mixed-effects models using 'Eigen' and S4, R package version 1.1-7.
- Bolinger, Dwight. 1972. *What did John keep the car that was in? *Linguistic Inquiry* 3, 109–114.
- Bosque, Ignacio & Gallego, Ángel J. 2014. Reconsidering Subextraction: Evidence from Spanish. *Borealis - An International Journal of Hispanic Linguistics* 3(2), 223–258.
- Bošković, Željko. 2017. Extraction from Complex NPs and Detachment. In M. Everaert and H. C. van Riemsdijk (eds.), *The Blackwell Companion to Syntax, Volume 2 (Second edition)*, Blackwell, Malden, Massachusetts.
- Bybee, J. 2006. From usage to grammar: The minds response to repetition. *Language* 82, 711–733.
- Bybee, Joan. 2013. Usage-based theory and exemplar representation. In T. Hoffmann and G. Trousdale (eds.), *The Handbook of Construction Grammar*, pages 49–69, Oxford University Press.

- Cattel, Ray. 1979. On Extractability from Quasi-NPs. *Linguistic Inquiry* 10, 168–72.
- Chomsky, Noam. 1964. *Current issues in linguistic theory*. The Hague: Mouton.
- Chomsky, Noam. 1973. Conditions on transformations. In Stephen Anderson and Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, pages 232–286, New York: Holt, Reinhart & Winston.
- Chomsky, Noam. 1986. *Barriers*. Mass.: MIT Press.
- Christensen, Rune Haubo Bojesen. 2018. Ordinal-Regression Models for Ordinal Data, R package version 2018.8-25.
- Copestake, Ann & Briscoe, Ted. 1995. Semi-Productive Polysemy and Sense Extension. *Journal of Semantics* 12(1), 15–67.
- Croft, William. 2001. *Radical Construction Grammar – Syntactic Theory in Typological Perspective*. Oxford University Press.
- Culicover, Peter & Jackendoff, Ray S. 2005. *Simpler Syntax*. Oxford University Press.
- Davies, Mark. 2008. The Corpus of Contemporary American English: 450 million words, 1990-present, available online at <http://corpus.byu.edu/coca/>.
- Davies, William D. & Dubinsky, Stanley. 2003. On Extraction from NPs. *Natural Language & Linguistic Theory* 21(1), 1–37.
- Deane, Paul. 1991. Limits to Attention: A Cognitive Theory of Island Phenomena. *Cognitive Linguistics* 2(1), 1–63.
- Deane, Paul D. 1992. *Grammar in Mind and Brain*. Mouton de Gruyter.
- Erteschik-Shir, Nomi. 1973. *On the Nature of Island Constraints*. Doctoral dissertation, Cambridge, MIT.
- Erteschik-Shir, Nomi. 1981. More on extractability from quasi-NPs. *Linguistic Inquiry* 12, 665–670.
- Erteschik-Shir, Nomi & Lappin, Shalom. 1979. Dominance and the functional explanation of island phenomena. *Theoretical Linguistics* 6, 41–86.
- Fillmore, Charles. 1977a. The case for case reopened. In P. Cole (ed.), *Grammatical Relations (Syntax and Semantics 8)*, pages 58 – 81, New York: Academic Press.
- Fillmore, Charles. 1977b. Topics in lexical semantics. In Roger Cole (ed.), *Current Issues in Linguistic Theory*, pages 76–138, Bloomington: Indiana University Press.
- Fillmore, Charles. 1982. Frame semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, pages 111–137, Seoul: Hanshin.
- Fillmore, Charles J., Johnson, Christopher R. & Petruck, Miriam R.L. 2003. Background to Framenet. *International Journal of Lexicography* 16(3), 235–250.

- Fodor, Janet Dean. 1983. Phrase Structure Parsing and the Island Constraints. *Linguistics and Philosophy* 6, 163–223.
- Fodor, Janet Dean. 1985. Deterministic parsing and subjacency. *Language and Cognitive Processes* 1(1), 3–42.
- Gibson, Edward, Piantadosi, Steve & Fedorenko, Kristina. 2011. Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass* 5(8), 509–524.
- Godard, Danièle. 1992. Extraction Out of NP in French. *Natural Language and Linguistic Theory* 10, 233–277.
- Goffman, Erving. 1974. *Frame analysis: An essay on the organization of experience*. London: Harper and Row.
- Goldberg, Adele E. 2006. *Constructions at Work: the nature of generalization in Language*. Oxford: Oxford University Press.
- Grosu, Alexander. 1981. *Approaches to island phenomena*. North-Holland.
- Hofmeister, Philip, Casasanto, Laura Staum & Sag, Ivan A. 2012. Misapplying working-memory tests: A reductio ad absurdum. *Language* 88(2), 408–409.
- Hofmeister, Philip & Sag, Ivan A. 2010. Cognitive Constraints and Island Effects. *Language* 86(2), 366–415.
- Horn, George Michael. 1974. *The noun phrase constraint*. Ph.D.thesis, University of Massachusetts, Amherst, Massachusetts.
- Kemmerer, David. 1998. Is syntax based on spatial image schemas in the inferior parietal cortex? Evidence against Deane's Parietal Hypothesis. *Cognitive Linguistics* 9(2), 180–188.
- Kluender, Robert & Kutas, Marta. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8, 573–633.
- Koster, J. 1978. Why subject sentences don't exist. In S. Jay Keyser (ed.), *Recent transformational studies in European languages*, pages 53–64, Cambridge MA: MIT Press.
- Kuno, Susumu. 1987. *Functional syntax – anaphora, discourse and empathy*. The University of Chicago Press, Chicago and London.
- Kuno, Susumu & Takami, Ken-Ichi. 1993. *Grammar and discourse principles: functional syntax and GB theory*. Chicago & London: University of Chicago Press.
- Kuznetsova, Alexandra, Brockhoff, Per Bruun & Christensen, Rune Haubo Bojesen. 2015. Package 'lmerTest'. Technical Report.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press.

- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, Volume 1: Theoretical Prerequisites*. Stanford CA: Stanford University Press.
- Marcus, M. 1980. *A Theory of Syntactic Recognition for Natural Language*. Ph.D.thesis.
- Melnick, Robin, Jaeger, T. Florian & Wasow, Thomas. 2011. Speakers employ fine-grained probabilistic knowledge, 85th Annual Meeting of the Linguistic Society of America. Pittsburgh, PA.
- Metusalem, Ross, Kutas, Marta, Urbach, Thomas, Hare, Mary, McRae, Ken & Elman, Jeffrey L. 2012. Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language* 66, 545–567.
- Minsky, Marvin. 1975. A framework for representing knowledge. In P. H. Winston (ed.), *The psychology of computer vision*, pages 211–277, New York: McGraw-Hill.
- Pritchett, Bradley L. 1991. Subjacency in a principle-based parser. In Robert C. Berwick (ed.), *Principle-based parsing: Computation and psycholinguistics*, pages 301–345, Dordrecht: Kluwer.
- Pustejovsky, James. 1995. *The Generative Lexicon*. MIT Press, Cambridge, MA, USA.
- Ross, John R. 1967. *Constraints on Variables in Syntax*. Ph.d. dissertation, MIT, Cambridge, Massachusetts. [Published in 1986 as *Infinite Syntax!* Norwood, NJ: Ablex Publishing].
- Rumelhart, David E. 1975. Notes on a schema for stories. In D. G. Brown and A. Collins (eds.), *Representation and understanding: Studies in cognitive science*, New York: Academic Press.
- Ruppenhofer, Josef, Ellsworth, Michael, Petruck, Miriam R. L., Johnson, Christopher R. & Scheffczyk, Jan. 2006. *FrameNet II: Extended Theory and Practice*.
- Schank, Roger C. & Abelson, Robert P. 1977. *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures*. Hillsdale, NJ: L. Erlbaum.
- Sprouse, Jon. 2007. *A program for experimental syntax: Finding the relationship between acceptability and grammatical knowledge*. Ph.d. dissertation, University of Maryland.
- Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43, 155–167.
- Sprouse, Jon & Hornstein, Norbert. 2013. *Experimental Syntax and island effects*. Cambridge: Cambridge University Press.
- Sprouse, Jon, Wagers, Matt & Phillips, Colin. 2012. A test of the relation between working memory capacity and syntactic island effects. *Language* 88(1), 82–123.

- Takami, Ken-ichi. 1992. *Preposition stranding: from syntactic to functional analyses*. Berlin: Mouton de Gruyter.
- Traugott, Elizabeth Closs & Trousdale, Graeme. 2014. *Constructionalization and Constructional Changes*. Oxford University Press.
- Van Valin, Robert D. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge University Press.
- Van Valin, Robert D. Jr. 1986. Pragmatics, island phenomena, and linguistic competence. In *Papers from the Parasession on Pragmatics and Grammatical Theory*, volume 22(2), pages 223 – 233, CLS.
- Van Valin, Robert D. Jr. 1995. Toward a functionalist account of so-called extraction constraints. In B. Devriendt, L. Goossens and J. van der Auwera (eds.), *Complex structures: A functionalist perspective*, pages 26–60, Berlin: Mouton de Gruyter.