

Chapter 19

Processing embedded clauses

Keir Moulton

University of Toronto

This chapter highlights key issues in sentence processing that are informed by, and inform, the study of clause embedding. Topics include center-embedding, head-final embedding structures, gap filling, islands, and relative clause processing and modifier attachment ambiguities and preferences. Much of the study of how humans parse sentences has involved embedded structures of one kind or another. Psycholinguists have exploited the complexity created by embedding to probe mechanisms of incremental parsing, including the role of memory, prediction, and language experience (usage). The overview is presented through a 'linguistic' lens through which developments in theories of human sentence processing are viewed. The chapter highlights areas where nuanced and formally-informed grammatical properties of embedding cross-linguistically have impacted our understanding of sentence processing.

1 Introduction

To understand how a system works, it is a good strategy to push it to its limits. As with other domains of inquiry, this is true in the psycholinguistic study of sentence processing. Complex sentences tax a human's parsing abilities, their comprehension, and their working memory capacity in ways that can reveal the underlying mechanisms of sentence processing. Embedded clauses, our topic, are perhaps the richest source of complexity in natural language. Like other types of embedded phrases, embedded clauses allow for sentences of arbitrary length and complexity. Embedded clauses do this better than any other kind of recursive syntactic option available because they admit the widest range of elements within them. These expansions interrupt relationships which help hold a sentence together, like subject-verb agreement and anaphoric and movement

dependencies. For this reason, embedded clauses are so routinely used as a tool to investigate sentence processing that they could be classified as the *drosophila* of the science of sentence processing.

Since embedded clauses have played such a key role in the development of sentence processing, this survey can touch on just a few issues that stand out, both for their impact on the processing literature and their interest to the linguist studying clause embedding. This chapter is not a general introduction to the field of sentence processing. For that, I can recommend, for example, the chapters in van Gompel (2013). This chapter, moreover, does not evaluate theories of sentence processing, because the significance of research on embedded clauses is too varied to serve this purpose. Rather, the purpose of this chapter is (i) to highlight some of the key debates in sentence processing that hinge intimately on the syntax and semantics of clause embedding, and (ii) to promote interaction between sentence processing research and the wealth of knowledge from deep analyses of the grammatical structures of embedding. We will see several cases where such interaction has been insightful. While psycholinguistics is a fast-moving discipline, the focus is on literature published before 2020, although pointers to more recent work are made throughout.

It is useful to start with a few key principles of sentence processing, in particular the incremental nature of the task and the information that parsing decisions might be made upon. This will constitute the first demonstration of the way psycholinguists have exploited clause embedding to probe the nature of the sentence processor. Section 3 delves into the topic of doubly center-embedded sentences. Satisfactorily capturing the difficulty of such sentences is a kind of litmus test for every sentence processing theory that has come along. So the approaches to this one phenomenon provide a good overview of some prominent theories of sentence processing. Section 4 discusses the role of clause embedding in the processing of filler-gap dependencies (e.g. *wh*-movement), in particular those filler-gap dependencies that cross clause boundaries and islands. In Section 5, the role of embedding in syntactic attachment ambiguities is discussed, both in terms of whether material is attached to matrix or embedded clauses and where embedded clauses themselves preferentially attach. Throughout, we will see that the rich syntactic and discourse-pragmatic properties of embedded clauses play a key role in how they are used by psycholinguists to investigate sentence processing.

2 Some basics of sentence processing with reference to embedding

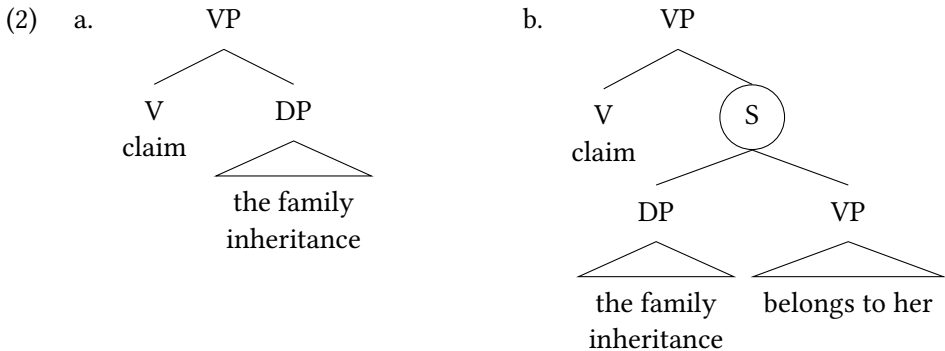
The investigation of human sentence processing (what I will simply call ‘processing’) seeks to uncover the cognitive mechanisms that give structure and meaning to incoming linguistic material, at the phrase and sentence level.¹ One of the central challenges for the processor is that material comes incrementally. This limitation can often lead to temporary ambiguities, so that the material available at one point in an utterance is compatible with more than one continuation. These are *Garden Path Sentences* (Bever 1970): sentences with a temporary ambiguity that can lead the reader to pursue a syntactic parse that is ultimately incorrect once the sentence unfolds. The classic example, not irrelevant to embedding phenomena, is one in which an embedded modifier (a reduced relative) is mistakenly taken for a matrix clause: *The horse raced past the barn fell* (See Section 5).

The abundance of research into Garden Path Sentences demonstrates that the processor does not entirely “wait” to assign an analysis at points of ambiguity (e.g. *raced*), but either commits to or privileges one of the possible analyses. What guides these commitments, and the extent to which these commitments are made or deeply encoded, and whether competitor analyses are entertained, have constituted the major topics in this area for the last 40 years. Relevant to the present chapter are temporary ambiguities of the sort shown in (1) (Frazier 1978): the verb *claim* can take either a DP complement or an embedded clause complement. These are named, after terminology of earlier times, NP/S ambiguities.

- (1) Frazier & Rayner (1982) *Minimal Attachment sentences*
 - a. The lawyers think his second wife will claim *the family inheritance*. (DP complement)
 - b. The lawyers think his second wife will claim [*the family inheritance belongs to her*]. (Embedded clause)

¹I do not mean to suggest a clear division between processing and “grammar”, a set of abstract rules separate from the implementation in comprehension and performance. See Lewis & Phillips (2015) for recent discussion and opinions. I should also note that this chapter draws from ‘on-line’ studies of real-time sentence comprehension (e.g. reading time measures) as well as acceptability judgment studies. I make no attempt to compare these methodologies, since in the studies discussed below, the research goals are varied. Moreover, the kinds of models I refer to will only be explored at a basic level, highlighting linguistic aspects primarily over foundational psycholinguistic matters, like the role of memory. For recent computational approaches to sentence processing, see for instance Vasishth & Engelmann (2021).

In a landmark study, which recorded readers' eye-movements as they read (eye-tracking while reading), Frazier & Rayner (1982) found longer reading times for the clausal complement sentences overall and, of particular interest, longer at the point of disambiguation. This suggests that readers were "garden-pathed" in (1b), pursuing an initial analysis where the DP encountered after the verb is taken to be an object, not the subject of the embedded clause. In one influential theory, the Garden Path Model (Frazier 1978, 1987a), this outcome is predicted from general structure-building principles. One of these principles is *Minimal Attachment*, which compels the processor to prefer structures with fewer nodes. The DP complement parse (which is correct for (1a)) involves fewer nodes than the embedded clause parse (1b), the latter requiring *at least* a sentential node (S) as is shown in (2b).²



At the point of processing the DP *the family inheritance* in (1), Minimal Attachment favours the structure in (2a) over (2b). When the verb *belongs* is encountered, reanalysis is required, leading to elevated reading times.³

Classic Garden Path theory prioritizes syntactic parsing principles, and in its strictest formulation, only allows other information to influence parsing in a later stage (so-called *two-stage* or *serial processing*). Other approaches allow a variety of linguistic and non-linguistic information to guide parsing, even in the earliest moments, including lexical and pragmatic information (Ford et al. 1982, Altmann & Steedman 1988, MacDonald et al. 1994). These are broadly categorized as

²This is the most conservative of phrase structures; most theories recognize a head for the S-node, which is T (Tense) or Infl (inflection). Furthermore, these sentences may contain a null complementizer, hence may in fact be embedded CPs not TPs. We return to the role of the complementizer below.

³A key feature of the garden path is the lack of an overt complementizer, which leads to other interesting predictions and manipulations (Holmes et al. 1987, Rayner & Frazier 1987, Ferreira & Henderson 1990, Trueswell et al. 1993).

constraint-based approaches, since they let a diversity of constraints be assessed at the same time. One important piece of information that the processor could use is syntactic subcategorization, such as whether a verb selects for a DP/NP or an embedded S. A number of studies from the 1990s argued that when the subcategorization preferences of verbs were taken into account, garden paths were ameliorated or disappeared. Trueswell et al. (1993) established, through an independent norming study, a collection of verbs that are biased toward embedded clausal complements ('S-biased' like *insist*, *reveal*) and those that are biased toward nominal complements ('NP-biased' like *observe*). Using a number of methodologies, including reading while eye-tracking, they found that garden path effects were present at the disambiguating predicate in the case of NP-biased verbs, but not S-biased verbs. This suggests that subcategorization information is used early (see also Mitchell & Holmes 1985, Osterhout et al. 1994, Garnsey et al. 1997, Jennings et al. 1997).

The experimental record, however, is not so clear cut. Pickering et al. (2000) examined the NP/S ambiguity but unlike previous experiments (cf. Garnsey et al. 1997) they compared manipulations just between S-biased verbs, such as *realize* in (3).⁴

(3) Pickering et al. (2000) Experiment 1 stimuli

- a. The young athlete realized her potential one day might make her a world-class sprinter.
- b. The young athlete realized her exercises one day might make her a world-class sprinter.

The crucial manipulation is the DP following *realize*. In (3a) there is a plausible NP object complement, but not in (3b) (*#She realized her exercises*). In a sentential complement analysis, both NPs are plausible. The logic of the experiment is this: if subcategorization information is available early, then readers will pursue a sentential complement analysis, and the plausibility difference will not matter since these are S-biased verbs. This is not what Pickering et al. (2000) found, however: there were increased reading times for the implausible (3b) at the post-verbal NP, suggesting readers pursued a complement clause analysis. Furthermore, downstream reading times were reversed, such that (3a) showed greater reading times than (3b) at the disambiguating modal *might*. This 'cross-over' pattern suggests a difficult reanalysis in (3a) from a plausible DP object analysis to a sentential complement analysis (see also Pickering & Traxler 1998). Pickering et al. (2000)

⁴Their stimuli also controlled for prior context and the likelihood of the matrix subject (cf. Garnsey et al. 1997).

propose a processing model that does incorporate frequency information but is not wholly guided by it. They suggest that pursuing an NP complement analysis is more readily falsifiable, since verbs and their objects are subject to pragmatic, syntactic, and semantic co-occurrence restrictions (e.g. selection) whereas verbs and the subjects within their complement clauses are not.

The question of what information is used in incremental sentence processing remains a big topic in the field, and garden path effects like the NP/S ambiguity continue to play a role in their development and formulation (see van Gompel 2013 for a recent overview). In the next section, we turn to another – perhaps *the* – embedded clause phenomenon central to the development of psycholinguistics.

3 Center-embedding

3.1 The phenomenon of center-embedding

A touchstone topic in psycholinguistics are sentences with doubly center-embedded (“nested”) relative clauses (Chomsky 1957, Yngve 1960, Chomsky & Miller 1963a,b, Miller & Isard 1964). The example in (25a) has a relative clause *who the nurse supervised* modifying a noun *intern* which is itself part of a relative clause *who the intern...bothered* modifying the matrix subject *the administrator*. Structures with nesting of two-levels such as this (compared to one-level of center embedding in (4b)) become impossible for most people to process, even if the propositional content is coherent – note for instance that (4c) has the same words and the same meaning but does not involve double center-embedding.

- (4)
- a. The administrator who the intern who the nurse supervised bothered lost the medical reports.
 - b. The administrator who the intern bothered lost the medical reports.
 - c. The intern who the nurse supervised had bothered the administrator who lost the medical reports.

The breakdown experienced in (4a) is a type of *processing overload*. It is distinct from other difficulties in sentence processing, such as classic garden paths like *The horse raced past the barn fell*, in that readers and hearers simply cannot overcome it.⁵ A skilled reader or a trained linguist, aware of the intended structure,

⁵There are certainly center-embeddings that also garden path the reader. Blumenthal (1966), for instance, pointed out that without relative pronouns or complementizers, center-embeddings give rise to a conjunction parse garden path (e.g. *The administrator, the intern, and the nurse*). But as Fodor & Garrett (1967) point out, this option is precluded with overt relative pronouns. Most of the studies reported retain the relativizers.

typically cannot reach a successful parse, although we will see how several types of manipulations can ameliorate such sentences, up to the point of acceptability and comprehensibility. In the early days of generative grammar, doubly center-embedded sentences were important for Chomsky's argument for the distinction between *competence* and *performance*. From a competence perspective, a doubly center-embedded sentence is the result of a legitimate, iterative application of the rule of relative clause formation. The problem, it was surmised, must then be with the performance systems.

Most approaches attribute the processing overload to the fact that such sentences involve suspended syntactic dependencies, which tax our working memory beyond its capacity.⁶ The term *syntactic dependency* is a theory-neutral way of describing relations that typically hold between heads and dependent phrases. One syntactic dependency is that between a predicate and its semantic and/or syntactic arguments (subject, object, etc.). Another syntactic dependency, an unbounded one, is that between a *wh*-phrase (such as a relative pronoun) and its gap. In doubly center-embedded sentences, subjects are separated from their predicates by complex material, and relative pronouns (or the associated heads of the relative clauses) are separated from their gaps by complex material. The syntactic dependencies of a doubly center-embedded sentence are diagrammed in (5); the predicate-argument dependencies are indicated by solid lines, and the dependencies between relative pronouns and their gaps (indicated with an underscore) are shown by a dashed line.

- (5) The administrator who the intern who the nurse supervised ____ bothered ____ lost

Chomsky & Miller (1963a,b), Yngve (1960) and Kimball (1973) explained the difficulty of center-embedding in terms of incomplete phrase-structure rules that must be maintained during the processing of the sentence. In the formalism of the time, subject-predicate relations were generated by a phrase structure rule of a sort where *S* expands into daughters *NP* and *VP* ($S \rightarrow NP VP$). At the point just before reaching the verb *supervised*, there are three opened *S*-rules that have not been completed with a *VP*. Keeping this many unresolved dependencies open is just too much for the processor to handle.

⁶Some accounts of center-embedding difficulty might fall outside this classification, including the idea of *perspective shift* in MacWhinney 1977, which essentially penalizes shifting the subject from one clause to the next, as happens with object relatives modifying a subject.

Hawkins (1990, 1994) emphasized how the size of subconstituents – measured in terms of word count – delays what he calls the “recognition of phrases” of the CONSTITUENCY RECOGNITION DOMAIN. The preferred order or nominal and clausal argument in (6) illustrates this (cf. Stowell 1981). Recognizing the domain of the VP in (6a) – getting evidence for each of its immediate constituents V, PP and CP – only takes three words: *to*, *everyone*, *that*. In contrast, recognizing the VP in (6b) requires parsing all the words in the CP (four) plus the head of the PP (*to*).

- (6) a. They [_{VP} said [_{PP} to everyone] [that it was raining]]
b. They [_{VP} said [that it was raining] [_{PP} to everyone]]

Complexity increases with the number of words before a constituent is “recognized.” In center-embedded sentences, sentential constituents cannot be recognized until their main predicate is encountered. The intervention of complex NPs (those with relative clauses) massively delays this and hence the complexity. Hawkins develops a far-reaching program that relates processing considerations such as these to the grammar of the world’s languages. The appeal of this program is that it recognizes that performance constraints can be “conventionalized” or “grammaticalized” to generate the distribution of word order patterns in natural languages.

Returning more concretely to center-embedding, these early characterizations just outlined turn out not to be sufficient. There are four key observations about center-embedding that a processing theory needs to capture. I take each in turn.

3.2 Locus of processing difficulty

If the difficulty of center-embedded structures arises from unresolved S-rules, we would expect that the processing overload would emerge as the opened rules “pile up” – that is, before the verbs are encountered. This is not, however, what the experimental record shows. Rather, processing breakdown occurs as the reader exits the most deeply embedded clause, at the point of the medial verb (e.g. *bothered* in (5)). This has been confirmed by various methodologies (Gibson & Thomas 1999, Grodner & Gibson 2005, Christiansen & MacDonald 2009, Vasishth et al. 2010). In fact, the difficulty at the medial verb in a doubly center-embedded sentence in English is so great that readers will perceive an ungrammatical sentence that simply omits this verb (7b) as more acceptable than its grammatical counterpart in (7a). This is known as the Missing VP effect (Frazier 1985, Gibson & Thomas 1999):

- (7) a. The apartment that the maid who the service had sent over was cleaning every week was well decorated.
- b. * The apartment that the maid who the service had sent over was well decorated.

The Missing VP effect is one piece of evidence that the processing overload occurs at the verbs that complete the subject-predicate dependencies. A theory of the processing complexity of English double center-embedding needs to capture this.⁷

3.3 The role of relative clauses

A second important fact about double center-embedding that speaks against a simple account involving unresolved subject-predicate dependencies is that a doubly center-embedded sentence involving noun complement clauses (CC), like the complements *fact* and *rumour* in (8), are relatively easier to parse (Gibson & Thomas 1997):

- (8) The fact that the rumour that the PM resigned is false should not surprise you.

If the processing difficulty were simply due to the fact that there are three open expectations for predicates, then we would expect doubly center-embedded complement clause sentences like (8) to show the same processing difficulty as doubly center-embedded relative clause sentences. This is not the case, which means that the causal nature of relative clauses is crucial to the processing overload. Of course, a relative pronoun introduces its own requirement, for a gap (see Section 4 for a fuller discussion of gap filling across clauses). We could add that as an open, unresolved expectation. This would mean that doubly center-embedded relative clauses like (5) in fact introduce five unresolved dependencies before the most deeply embedded verb is encountered. This would capture the fact that (5) is harder to process than (8).

3.4 The distance of the dependency

Still, even this is not a sufficient characterization of the difficulty of double center-embedding. There is a third important observation about center-embedding due

⁷The Missing VP effect has also been found in French (Gimenes et al. 2009) and German (Häussler & Bader 2015).

to Cowper (1976) that is a major desideratum for processing theories. Both sentences in (9) involve center-embedding to a depth of two, but while (9a) causes processing overload, (9b) does not (Gibson & Thomas 1997). What makes them different is that in (9a), the innermost clause is a relative clause (RC) while the outermost clause is a complement clause (CC) to the noun *fact*. In (9b) it is the reverse: the innermost clause is a CC embedded in an RC.

- (9) a. The fact [that the employee [who the manager hired ____] stole office supplies] worried the executive.
b. # The executive [who the fact [that the employee stole office supplies] worried ____] hired the manager.

Both sentences involve the same number of unresolved expectations just before encountering the most deeply embedded verb (*hired* in (9a), *stole* in (9b)), namely four (three subjects seeking a predicate and one relative pronoun seeking a gap). And yet (9b) causes processing difficulties while (9a) does not.

The important difference is the distance across which the relative pronoun-gap dependency spans. In (9a) it is a local *wh*-dependency, whereas in (9b) it is long-distance. The role this plays in center-embedding, then, must be captured by processing theories.

3.5 Nature of intervening DPs

A final and equally influential observation is that doubly center-embedded relatives in which the innermost relative has an indexical pronoun (*I*, *we*, *you*) as a subject are easier to process (Bever 1970, 1974, Kac 1981, Gibson 1991, Kluender 1998, Warren & Gibson 2002):

- (10) a. The reporter [everyone [*I* met] trusts] said the president won't resign yet.
b. The student [who the professor [who *I* collaborated with] had advised] copied the article.

For instance, Warren & Gibson (2002) found that a doubly center-embedded sentence where the innermost subject was an indexical pronoun was rated as more acceptable than those in which the same position contained a proper name, definite description or third person pronoun.

Perhaps the first explicit theory to capture each of the above observations about center-embedding is Gibson's *Dependency Locality Theory* (DLT) (Gibson

1998, 2000).⁸ DLT is a general processing theory with broad coverage beyond just center-embedding. DLT has two components: a storage component and an integration component. Integration involves the retrieval of the head of a dependency: an argument for a predicate or a relative pronoun (or head noun) for a gap position in a relative clause. Retrieving an element from memory involves reactivating it. The level of reactivation needed will depend on the extent to which the element's activation level has decayed, and this can be affected by the amount and complexity of intervening material. Gibson capitalizes on the fact that the kind of intervening material that seems to most tax the processor are discourse referents – entities like individuals and events that are introduced in a discourse structure (Kamp 1981, Heim 1982). It has been independently established that processing a DP depends on the nature of the referent – in particular how “accessible” (Gundel et al. 1993) it is (Haviland & Clark 1974, Halliday & Hasan 1980, Garrod & Sanford 1982, Garrod 1994). In DLT, processing new discourse referents incurs a cost.⁹ These include referential DPs, proper nouns, third person pronouns, as well as tensed verbs (which introduce reference to times, Partee 1973). Importantly, indexical pronouns are not discourse referents so they do not incur a processing cost. Integration costs, defined in (11) (Gibson 2000: 105: (9)), are dependent on the number of discourse referents that intervene between the site of integration and the target of retrieval.

(11) *DLT structural integration cost*

The structural integration cost associated with connecting the syntactic structure for a newly input head h_2 to a projection of a head h_1 that is part of the current structure for the input is dependent on the complexity of the computations that took place between h_1 and h_2 . For simplicity it is assumed that 1 EU [Energy Unit] is consumed for each new discourse referent in the intervening region.

Object relative clauses provide a good demonstration of DLT's structural integration costs. It has been overwhelmingly established that object-gap relatives as in (12a) incur greater processing difficulty than subject gap relatives (12b) (Holmes 1973, Hakes et al. 1976, Wanner & Maratsos 1978, Ford 1983, Waters et al. 1987, King & Just 1991, among many others).

⁸Gibson (1998) proposed two variants of the theory, one of which was the Syntactic Prediction Locality Theory (SPLT). I concentrate on DLT as it is reported in Gibson (2000).

⁹Gibson uses the term “new discourse referent.” Typically we use that term for expressions like indefinites that introduce a discourse referent. Definite descriptions often have familiar referents, although the role of accommodation in psycholinguistics studies is relevant here (Schwarz 2007). I think “new discourse referent” should simply be understood here as discourse referent.

- (12) a. The intern who the nurse supervised ____
b. The nurse who ____ supervised the intern.

At the gap site, the object relative pronoun *who* in (12a) is retrieved and integrated as the object of *supervised*, but there are two discourse referents within the span of this dependency (the DP *the nurse* and the verb *supervised*). The structural integration cost at the gap is therefore 2 EUs. At the gap site in a subject relative (12b), there are no integration costs since there are no discourse referents between the head and tail of this dependency.

In doubly center-embedded object relative clauses, repeated below from (5), the second gap is the site of maximal integration cost because there are five discourse referents intervening: *the intern*, *the nurse*, the second relative pronoun *who*, and the verbs *supervised* and *bothered*.

- (13) The administrator who **the intern who the nurse supervised** ____
bothered ____ lost

Gibson's account captures each of the four observations discussed above. The fact that costs are highest at the gap site, which is adjacent to the embedded verbs, predicts the location of the processing difficulty. Cowper's (1976) contrast in (9) between CCs embedded in RCs versus RCs embedded in CCs is predicted, since the number of discourse referents within the span of the relative pronoun-gap dependency is greater in the former than the latter. Finally, the ameliorating role of indexicals (10) follows from the definition of discourse referents: indexicals are, in a certain sense, "cost-free" since the speaker and hearer are always easily accessible. The presence of an indexical pronoun instead of a third person pronoun, a definite description or a proper noun, will reduce the integration cost.

The second component of DLT, the storage component, bears more similarity to earlier theories of center-embedding difficulty. There is a cost for storing a prediction needed to resolve a syntactic dependency and this can accrue. For instance, each subject requires a predicate, and this sets up an expectation which is added to working memory. Relative pronouns also set up an expectation for a gap. In (5), just before the word *supervised*, there are 5 expectations – which in DLT incur a "5-unit" storage cost. This should be compared to the non-nested relatives in (4b), where the storage costs never surpass 3 units – before reaching *bothered* there are two DP arguments requiring a predicate and one gap expectation. The contrast in (9) is also reflected in different storage costs: since a CC does not involve a relative pronoun-gap dependency, it will incur one less unit of storage costs than an RC.

Support for storage costs is less robust, but has been argued for by Chen et al. (2005) and Gibson et al. (2005). Chen et al. (2005) looked at the reading time costs associated with expectations initiated by a DP for its predicate. What makes this study particularly informative is that sentential complements were used as the intervening material, which avoids the added complexity of a relative which would itself have a gap. In (14a), the expectations of each subject DP for a predicate are fulfilled immediately. In (14b), however, the DP subject *the implication* sets up an expectation for a predicate but this is only reached after processing its sentential complement. In (14c) there are two DP subjects whose expectation for a predicate must be held while processing the innermost sentential complement (which is identical to that in (14b))

- (14) a. The employee realized that the boss implied that *the company planned a layoff* and so he sought alternative employment.
 b. The employee realized that **the implication** that *the company planned a layoff* was not just a rumour.
 c. **The realization** that **the implication** that *the company planned a layoff* was not just a rumour caused a panic.

As predicated by storage costs, reading times at the critical region *the company planned a layoff* were greater in (14b) than (14a) and greatest in (14c), where two subjects' expectations had to be held. In the section below on cross-linguistic investigations of embedded clauses, we return to the role of storage costs vs. the role of integration (or retrieval costs).

Since Gibson's seminal work on center-embedding, the retrieval component of integration has lead to the greatest interest among psycholinguistics. Over the last two decades or so, increased attention has been paid to research on memory in other domains of cognitive psychology. One prominent model holds that retrieval of items in memory involves cue-based recall from a content-addressable memory (Lewis & Vasishth 2005, Lewis et al. 2006). This is best demonstrated with pronominal anaphora: a pronoun, such as *she*, bears the features *singular* and *feminine* and these serve as a cue to recall antecedents in memory. So items in memory that bear either of these features will be activated, although the one that matches best (i.e. *the girl* vs. *the boy*) will attain the highest activation level and will be retrieved more readily. One thing that makes this model different from DLT is that a content-addressable memory does not involve a search related to a syntactic parse tree. Items in memory are not "syntactically" structured. Furthermore, elements in working memory compete for retrieval, with retrieval being easier the better and more recent the match, and therefore more

highly activated, the target. So when a gap site cues for the retrieval of a filler, other DPs constitute potential targets. This will make object gap relatives harder than subject gap relatives because there is simply a more recently activated competitor for the head noun (i.e. the embedded subject) in object gap relatives. In addition, competitors that are more similar to the target can interfere (Gordon et al. 2001, Traxler et al. 2002, Gordon et al. 2004), thus causing processing difficulty. This *similarity-based interference* straightforwardly captures why indexicals ameliorate double center-embedding: they are so distinct from the kinds of DPs that serve as heads of relative clauses (or their associated relative pronouns) that they do not compete in cue-based recall. Gordon et al. (2001, 2004) examine how DPs of different types (not just indexical pronouns) do and do not interfere in the retrieval of the head of the dependency. Lewis & Nakayama (2002) explore the effects of similarity factors, such as morphological case, in Japanese center-embeddings. Models of sentence processing that implement these observations within broader research on memory can be found in Van Dyke & Lewis (2003), Lewis & Vasishth (2005), Van Dyke & McElree (2006), Vasishth & Lewis (2006) among others. As these theories are general and applicable to a range of dependencies (not just those that arise in center-embedding) I refer to the reader to these works. Nonetheless, it can safely be said that the work by Gibson and others on center-embedding greatly motivated work that embeds sentences processing in the larger cognitive science of memory.

Yet another approach to the difficulty of center-embedding, among other processing difficulties, has been proposed within *Surprisal* theory (Hale 2001, Levy 2008). In this framework, the frequency of a structure matters to the expectations of the processor. When incoming material requires pursuing a less-frequent structure, processing difficulties arise. Object-gap relatives following a matrix subject are rare; encountering two such structures consecutively is even rarer (Gildea & Temperley 2007, Karlsson 2007). Space prevents doing justice to these approaches; for an overview of Surprisal Theory, see Levy (2013).

3.6 Embedding cross-linguistically and cross-constructionally

The preponderance of experimental work on center-embedding in English is limited to clausal dependents of nominals: relative clauses and sentential complements.¹⁰ In robustly head-final languages, center-embedding is routine. In

¹⁰These are not the only sources for center-embedding in English, adjunct clauses can be self-embedded:

- (i) Because if when the baby is crying, the mother gets upset, the father will help, the grand-mother can rest easily.

Japanese, for instance, which has SOV word order, even one level of clause-embedding constitutes center-embedding. In (15a) the embedded clause intervenes between the matrix subject and its verb. The example in (15b) involves two levels of center-embedding:

- (15) a. Bebiisittaa-ga [ani-ga imooto-o ijimeta to]
 babysitter-NOM [older.brother-NOM younger.sister-ACC bullied that
 itta
 said
 ‘The babysitter said that my older brother bullied my younger sister.’
- b. #Obasan-ga [bebiisittaa-ga [ani-ga imooto-o
 aunt-NOM [babysitter-NOM [older.brother-NOM younger.sister-ACC
 ijimeta to] itta to] omotteiru
 bullied that] said that] thinks
 ‘My aunt thinks that the babysitter said that my older brother bullied
 my younger sister.’

Babyonyshev & Gibson (1999: 424) reports that the doubly nested structure is perceived as more complex (and appended the # sign) than the singly nested structure (see also Uehara & Bradley 1996). Consistent with DLT, Babyonyshev & Gibson (1999) found an effect of transitivity, such that center-embedded sentences whose innermost clause was transitive were rated as harder to understand than minimally different sentences where the innermost clause was intransitive. This follows on Gibson (1998)’s DLT, since it reduces the number of dependencies being processed and the number of discourse referents introduced within the overarching dependencies. Center-embeddings have also been examined from the perspective of cue-based retrieval models, both in Japanese (Lewis & Nakayama 2002) and Hindi (Vasishth 2003).

A number of other studies have revealed that sometimes center-embedding can facilitate processing. These findings have been given the moniker *anti-locality effects*. Konieczny (2000) found in a self-paced reading experiment of German, a verb-final language, that the matrix verb *hingelegt* ‘laid down’ was read faster when it followed the embedded relative clause as in (16a) than when it preceded the (extraposed) relative clause in (16b):

Processing overload is similarly felt in these cases. The dependencies here are at the clausal level, rather than of the predicate-argument type in the classic center-embedded sentences discussed above. Moreover, they vary in whether there is a filler-gap dependency: *if*-clauses do not contain a gap whereas temporal *when*-clauses do (Geis 1970). To my knowledge, these embeddings have not received the same amount of attention in the recent psycholinguistics literature.

- (16) a. Er hat das Buch, [das Lisa gestern gekauft hatte], hingelegt.
 he has the book, that Lisa yesterday bought had laid.down.
 ‘He has laid down the book that Lisa had bought yesterday.’
 b. Er hat das Buch hingelegt, [das Lisa gestern gekauft hatte].
 he has the book, laid.down that Lisa yesterday bought had .
 ‘He has laid down the book that Lisa had bought yesterday.’

Konieczny (2000: 644) suggests that a relative clause “preceding the verb might just provide enough time” for arguments that are being processed to help build predictions about the verb to come (that it *is* coming and what its argument structure might be), and thus facilitates reading time when that verb is encountered (see also Konieczny & Döring 2003). Vasishth & Lewis (2006) found similar effects in Hindi doubly center-embedded constructions. In a self-paced reading experiment, they compared doubly center-embedded control constructions like (17) to similar constructions in which material was added between the object and verb of the innermost clause, indicated by “intervening material” in (17). The intervening material was of several types, including a subject-gap relative clause with a DP object.

- (17) Sita-ne Hari-ko Ravi-ko kitaab-ko *intervening material* khariid-neko
 Sita-ERG Hari-DAT Ravi-DAT book-ACC buy-INF
 bol-neko kahaa.
 tell-INF told
 ‘Sita told Hari to tell Ravi to buy the book.’

Gibson (2000)’s DLT predicts increased processing difficulty at the innermost verb *buy* when intervening material is present, particularly intervening material like the relative clause that contains discourse referents. Contrary to this prediction, Vasishth & Lewis (2006) found faster reading times on this verb in the presence of intervening material. They develop a model in which intervening material heightens the expectation for a verb, thus facilitating its integration. Ensuing research has established the existence of both locality and anti-locality effects in the same language and study (Demberg & Keller 2008, Jaeger et al. 2008, Vasishth & Drenhaus 2011, Levy & Keller 2013).

Recall that retrieval is not the only possible processing difficulty that center-embedding poses. Gibson’s DLT also included a storage cost: encountering a DP subject predicts a verb, and as such predictions stack up (as they do with center-embedding) processing difficulty might increase. This is a different prediction from retrieval approaches – that is, the integration component of DLT

or more recent cue-based retrieval models – which predict difficulty at the retrieval site (the verbs or their close-by gaps). These predictions were tested by Nakatani & Gibson (2010), which is notable as the first online processing study of Japanese center-embedding. For doubly center-embedded sentences like (18) (with a schematization just above it), a retrieval-based account predicts equal costs on the DPs but increasing costs on the verbs, with V1 being the most difficult, since it has to retrieve its subject across two other candidate NPs (NP2, NP3).

- (18) [NP1 [NP2 [NP3 V3 COMP] V2 COMP] V1]
 [syoki-ga [daigisi-ga [syusyoo-ga utatanesita
 [secretary-NOM [congressman-NOM [prime minister-NOM dozed
 to] koogisita to] hookokusita]
 COMP] protested COMP] reported]
 ‘The secretary reported that the congressman had protested that the
 prime minister had dozed.’

In contrast, storage predicts processing difficulty as NPs accrue: for (18) the most costly region would be that corresponding to NP3 because here there are the stored expectations of three verbs. We would expect difficulty to decrease as each dependency is completed. So, unlike the retrieval approach, the verbs should be increasingly easy to process moving rightward. The results of a self-paced reading experiment revealed that participants slowed down as the NPs stacked up and sped up as the verbs were encountered, supporting a storage-based view.¹¹

The overall picture, then, shows both effects of storage (the maintenance of syntactic predictions) and effects of retrieval, but in different languages. While Hindi, German, and Japanese center-embedding do not pose processing difficulty for retrieving the heads of the dependencies, there remains the evidence of retrieval costs in English center-embedding (as noted above from Grodner et al. 2005; see Santi et al. 2019 for a recent replication of the effect for single embeddings). One possibility, suggested by Nakatani & Gibson (2010: 108), is that the type of dependency matters: retrieval difficulty arises in relative clause constructions precisely because they are filler-gap configurations. The German, Hindi and Japanese dependencies, where no retrieval effects are found, involve predicate-argument dependencies. The idea then is that storage costs correlate

¹¹Nakatani & Gibson (2010) also discuss the predictions of *Surprisal Theory*, which is compatible with this result but other details of their study were more equivocal on separating the predictions of a DLT-style storage theory with *Surprisal Theory*.

with predicate-argument dependencies and retrieval costs correlate with gap-filler dependencies. This suggestion, however, is somewhat at odds with the full experimental record: both Grodner et al. (2005) and Santi et al. (2019) find elevated reading times at the matrix predicate in English sentences with relative clause embeddings, and this is presumably where the subject-predicate dependency is resolved. At present, I do not know of a resolution to the various findings of locality and anti-locality.

The value of cross-linguistic experimentation should be apparent by now. What must not be overlooked is the value of examining a range of embedded constructions, even in one language. As noted in footnote 9, relative and complement clauses are by far the most studied types of embedding, especially in terms of center-embedding. But interesting insights – both for time-course processing theories and structural analyses – arise from studying other types of embedding. A case in point are comparative clauses. Grant (2013) examined deviant sentences like (19a), in which a comparative clause modifies a matrix subject. The deviance is surprising since the version where the comparative clause is extraposed (19b) is perfectly acceptable.

- (19) a. *?/# More boys than girls ordered steak ordered salad
b. More boys ordered salad than girls ordered steak

Osborne (2009) had developed a grammatical theory to rule out (19a). Grant shows that rather than complicate the grammar, the unacceptability of (19a) can be profitably understood as a type of garden path. The comparative *than* is preferentially parsed as taking a DP sister rather than a clausal element, and so upon encountering another VP the parser aborts. When the element after *than* disambiguates, as with the proper noun in (20a) in comparison to the bare plural in (20b), processing is easier.

- (20) *Grant (2013) Experiment 1 stimuli (subset)*
a. More nurses than Percival thanked cared a great deal about their jobs.
b. More nurses than patients thanked cared a great deal about their jobs.

Structural theorists can profit from results such as these, as they both clarify and, in this case, simplify what needs to be hard-coded in the grammar. We will see an analogous situation – island phenomena – where the reduction of a perceived unacceptability to processing difficulty is more controversial.

While we are on comparatives, it is relevant to highlight another fact that Grant (2013) discovered. Unlike in the canonical relative clauses as we saw above,

subject gaps in comparative clauses are more difficult to process than object gaps. In eye-tracking studies, Grant compared subject vs. object gaps in comparatives in sentence-internal (or “base”) position (21a), (21b) with subject vs. object gaps in extraposed comparative clauses (21c), (21d):

(21) *Grant (2013) Experiment 2 stimuli*

- a. Object Extraction, Base Position
More friends than we talked to were at the party, according to the pictures on Facebook.
- b. Object Extraction, Extraposed
More friends were at the party than we talked to, according to the pictures on Facebook.
- c. Subject Extraction, Base Position
More friends than talked to us were at the party, according to the pictures on Facebook.
- d. Subject Extraction, Extraposed
More friends were at the party than talked to us, according to the pictures on Facebook.

There were faster reading times for object gap comparatives, and an overall penalty against subject gap comparatives in base position. This is quite striking since extraposed clauses tend to be more difficult (Levy et al. 2012). Grant shows that these processing differences between canonical subject-gap relatives and subject-gap comparatives relate to underlying structural differences in the way the gaps are represented: essentially, the gap in a subject comparative is *within* a subject, which following Pancheva & Tomaszewicz (2011) constitutes a type of island violation (see Section 4 on islands). The lesson here, which will be repeated a few times below, is that deep understanding of the compositional details of the syntactic and semantic structures can offer insight into processing observations.

3.7 The role of prosody in embedding

Before leaving the topic of center-embedding, one final ameliorating factor needs to be discussed. Fodor (2013) has argued that prosody plays a role in the difficulty – and amelioration – of center-embedding. Fodor starts by pointing out the tension between the syntax of a doubly center-embedded sentence – being a deeply-nested structure – and the requirements of prosodic phrasing – which is typically thought to require a flatter structure. In a sense, there is no optimal

prosody for a doubly center-embedded structure. This can pose processing difficulty because, as argued by Janet Fodor and colleagues (see Frazier & Gibson 2015 for an overview), silent reading involves implicit prosody. If there is no natural prosody that can be applied to such sentences, this could register as processing difficulty in reading studies.

Fodor (2013) and Fodor et al. (2018) lay out the prosodic options for double center-embedding. Given the length of such sentences, a prosodic break is required. The break forms two phonological phrases (see, e.g. Selkirk 1996 on distinctions among phonological and intonational phrases). Phonological phrasing is sensitive to major syntactic boundaries (e.g. between subject and predicate), as well as phonological considerations. A relevant phonological consideration is that sister prosodic phrases prefer to be of equal weight. So while a natural place for a prosodic break in English is between subject and predicate, this delivers unbalanced phonological phrases in the case of double center-embedding as in (22a) (Fodor et al. 2018: (8)). (The double pipe “||” indicates a phonological phrase break.) To create balanced phrases, the predicate needs to be longer and the RCs need to be shortened, the latter being achieved in (22b) (Fodor et al. 2018: (9)) with a pronoun (see above for a different interpretation of the pronoun manipulation).

- (22) a. The girl the man the cat scratched kicked || died.
b. The girl the man I love met || died of cholera in 1960.

Fodor et al. (2018) suggest that cases like (22a) are going to be hard to come by. So a more generally applicable solution requires dividing a double center-embedding into three phonological phrases. This will alleviate some of the pressures on the length and weight, but here too there are pitfalls. Syntactic considerations pressure a second break between the matrix subject and the RCs (23a) (Fodor et al. 2018: (10)), but in order to have phrases of equal length the subject needs to be relatively heavy and the RCs relatively short – although here even a non-pronominal innermost subject appears possible (23b) (Fodor et al. 2018: (12)).

- (23) a. The girl || that the young man I love met in Barcelona || died.
b. The elegant woman || that the man Jill loves met || moved to Barcelona.

If the RCs are any longer, however, a fourth phrase is required. But this appears not to be helpful: (24) is on the verge of a list interpretation (see footnote 5 on Blumenthal 1966).

- (24) The elegant woman || that the man Jill loves || met on a cruise ship || moved to Barcelona.

The three-phrase prosody, then, seems to be the best compromise. Prosody should then be included as yet one more factor that interacts with the processing difficulty of center-embedding. This avenue of research is relatively new, and experimental evidence is accumulating (I refer the reader to the studies cited in Fodor et al. 2018). But these prosodic manipulations have not, to my knowledge, been extensively tested with respect to the manipulations and measures discussed in the previous sections.

4 Extraction from embedded clauses

4.1 The phenomenon of extraction

We have looked at syntactic dependencies that span embedded clauses. Syntactic dependencies can also enter into embedded clauses, as in long-distance *wh*-movement.

- (25) a. Who did they report [_{CP} that the police arrested ____].
 b. This is the criminal which they reported [_{CP} that the police arrested ____].

Frazier & Clifton (1989) examine long-distance extraction, finding that *wh*-dependencies that cross clause boundaries are more difficult to process than short, mono-clausal extraction. This work builds on an important finding in the processing of filler-gap dependencies known as the *Filled Gap Effect* (FGE) (Crain & Fodor 1985, Stowe 1986). A FGE arises in (26a): the filler *who* expects to find a gap after the verb *bring*. When this gap is ‘filled’ by material, like *us*, processing difficulty is incurred in comparison to the same lexical NP in a non filler-gap sentence like (26b).

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

Frazier (1987b) and Frazier & Clifton (1989) attribute FGEs to an *Active Filler Strategy* (27) (Frazier & Clifton 1989: 95): when the processor encounters a filler (like a *wh*-phrase) it prioritizes finding a gap site, and so expects a gap as soon as possible, such as the direct object position in (26a).¹² When that position is filled, the processor’s strategy is thwarted and processing difficulty registers.

¹²See Stowe 1986 for discussion of the matrix subject in (26a).

(27) *Active Filler Hypothesis*

When a filler has been identified, rank the option of assigning it to a gap above all other options.

Frazier & Clifton (1989) tested whether sentences with extraction from an embedded clause show FGEs (at *the guests* in (28b)) just as it is expected to in the short extraction in (28a). (The sentences in (28c), (28d) are non filler-gap controls, where no FGE is expected.)

(28) Frazier & Clifton (1989) Experiment 1 stimuli

- a. Who did the housekeeper from Germany urge the guests to consider ____?
- b. Who did the housekeeper say [_{CP} she urged the guests to consider ____]?
- c. The housekeeper from Germany urged the guests to consider the new chef.
- d. The housekeeper said [_{CP} she urged the guests to consider the new chef].

In self-paced reading measures, both the short and bi-clausal extractions elicited a FGE. Nonetheless, in this and other studies they report, longer reading times were found for long-distance extraction as well as reduced acceptability at an end-of-sentence grammaticality detection task. In another study (their Experiment 4), Frazier & Clifton (1989) directly compared short (29a) vs. long extraction (29b) with short vs. long non-extraction (yes/no) questions.

(29) Frazier & Clifton (1989) Experiment 4 stimuli (subset)

- a. What did your beautifully dressed niece mutter ____ to Willy in the house?
- b. What did you think your niece muttered ____ to Willy in the house?
- c. Did your beautifully dressed niece mutter something to Willy in the house?
- d. Did you think your niece muttered something to Willy in the house?

They found an interaction in acceptability judgments, such that there was no difference between long and short non-extraction questions, but there was a difference between the long and short extractions, the former being judged less acceptable. Frazier & Clifton (1989) take these results to suggest both that fillers are active across clause boundaries, and that maintaining a filler across a clause

boundary is costly. They claim that there is nothing about the complexity of the bi-clausal structure itself that could be driving the differences they observed, given the lack of an effect with the non-extraction sentences which were also bi-clausal questions (although see below for possible differences between content questions and yes/no questions). They offer an interpretation in terms of successive cyclic movement (Chomsky 1977, 1981, 1986), whereby long-distance movement proceeds in shorter steps, moving to the edge of each (finite) CP.

(30) Who did you say [_{CP} ____ that Fred thinks [_{CP} ____ that Mary hired ____]]

Since cross-clausal extraction requires creating an additional filler-gap dependency, this creates a substantial processing load. Gibson & Warren (2004) present evidence that successively cyclic movement can actually aid in processing, as it breaks up the span of material across which the filler has to be held in memory.

In more recent work, Wagers & Phillips (2014) also examined extraction from embedded clauses. They too used FGEs as a probe for active-filling, but with a refined version of this method due to Lee (2004). Instead of comparing filler-gap sentences to non filler-gap sentences as above, fillers of two different types are compared: NPs fillers (31a) vs. PP fillers (31b):

- (31) a. The stones *which* the pilgrim toppled **the cairn** for ____.
 b. The stones *for which* the pilgrim toppled **the cairn** ____.

The unmarked word order in English requires NPs to precede PPs. So while the processor will predict a gap for the NP filler in (31a) just following the verb (and hence we expect a FGE at *the cairn*), the PP filler in (31b) makes no such prediction hence no FGE. This design has the advantage of comparing sentences with the same interpretation that both have filler-gap dependencies.

Wagers & Phillips (2014) import this design to test three types of dependency lengths: short mono-clausal (32), long mono-clausal where the subject has a modifying PP (32b), and bi-clausal (32c) where the gap is inside an embedded declarative clause. This three-level factor was crossed with whether the filler was an NP or PP (as illustrated in (31)).

- (32) Wagers & Phillips (2014), Experiment 1 stimuli
 a. ...*which/for which* the technician carefully prepared **the clean tubes** (for) ____ while wearing a mask...
 b. ...*which/for which* the technician [_{PP} at the medical research facility] carefully prepared **the clean tubes** (for) ____ while wearing a mask...

- c. ...*which/for which* the young biologist said [_{CP} that the technician carefully prepared **the clean tubes** (for) while wearing] a mask....

They indeed found elevated reading times (using a self-paced reading methodology) at the ‘filled’ direct object position for NP fillers, and they found this across all conditions. This is consistent with Frazier & Clifton (1989)’s findings, an encouraging result since the filled-gap design is more reliable here. They also report overall that the CP conditions were read more slowly, but this does not interact with filler type.

In addition to manipulating the category of the filler, Wagers & Phillips (2014) conducted a second study in which the plausibility of the filler with respect to the verb was tested (Traxler & Pickering 1996), e.g. *The acorns which the squirrels quickly crammed their small puffy cheeks with* (plausible) vs. *The cats which the squirrels quickly crammed their small puffy cheeks with*. If the processor is engaged in actively resolving the filler-gap dependency, the semantic anomaly of the implausible condition should be detected at the verb itself. This is indeed what Wagers & Phillips (2014) found for short, mono-clausal extraction but not for the bi-clausal extractions. In that case, and to some extent in the PP case, the processing difficulty registered at a later region where the gap position was unambiguously present (object of a stranded preposition). Wagers & Phillips (2014) interpret these results within a broader question of whether filler-gap dependencies are processed with the Active Filler Strategy (where the filler is maintained in memory until it finds a satisfying gap) or a “retrieval”-based approach, whereby at a gap site, fillers that have been shunted out of working memory are retrieved (and this is susceptible to decay and interference in the ways suggested in Section 3). Wagers & Phillips argue that their results show that a hybrid account is warranted, in which dependency length plays a role. Gross, category-based predictions about a gap are, as is consistent with the Active Filler Strategy, maintained in working memory over both short and long dependencies, including cross-clausal dependencies. This is what the Filled Gap Effects showed. Finer grained information about the filler like certain semantic details, however, are not actively maintained and therefore have to be retrieved. This retrieval strategy is sensitive to dependency distance, and so we do not see the same early effects in long-distance movement at the gap site. What remains to be sorted out is whether it is simply distance that regulates these memory resources or whether clausal embedding itself plays a role above and beyond distance.

4.2 Islands

One of the most well studied aspects of extraction are island phenomena (Ross 1967), where movement dependencies are blocked. Many islands are, of course, embedded clauses: *if*- and *whether*-clauses (33a), *wh*-islands (33b), complex NP islands (33c), sentential subject islands (33d) and factive islands (33e):

- (33) a. *? Who did you wonder if/whether they hired ____?
 b. *? Who did you wonder who to hire ____?
 c. * Who did you reject the claim that they should hire ____?
 d. * Who did [that you hired ____] surprise most?
 e. * How did you hate [that they played the song ____]

There is notoriously a great deal of gradation and variation in judgments about island-violating sentences, something that has inspired many to seek processing explanations for the unacceptability of extraction from islands (Kluender & Kutas 1993, Kluender 1998, Hofmeister & Sag 2010 among others). Kluender (1998) and Hofmeister & Sag (2010), argue that islands are the result of a type of processing overload, much like center-embedding. The idea is that in island-violating sentences, several factors conspire, each hard in themselves, to make such sentences unacceptable. One of those factors, they claim, is the general difficulty of long-distance extraction.¹³ The nature of the clause boundary also appears to incur a cost, independently of movement. Kluender & Kutas (1993) show that even in *yes-no* questions, in which no *wh*-element crosses a clause boundary, different types of embedded clauses are processed differently. Using Event-related potentials (ERPs, see Kutas & Federmeier 2007 for a handbook overview), they found neurophysiological responses which they interpreted as indicating a processing load to be highest with *wh*-complements, followed by *if*-complements, and lowest with *that*-complements.

- (34) a. Has she forgotten [that he dragged her to a movie on Christmas Eve]?
 b. Has she forgotten [if he dragged her to a movie on Christmas Eve]?
 c. Has she forgotten [who he dragged ____ to a movie on Christmas Eve]?

In a judgment study they also found a cline in acceptability, whereby declarative complements were judged more acceptable than embedded *if*-clauses, which

¹³ Other factors they cite include the referential complexity of the *wh*-filler, and the kinds of referents within the span of the filler-gap dependency, as with Dependency Locality Theory.

in turn were more acceptable than embedded *wh*-questions. Kluender (1998) interprets these differences in relation to the “referential” load of the complements; Hofmeister & Sag (2010) point out that question complements are semantically more complex since, on most semantic theories, they invoke sets of alternative propositions (e.g. Hamblin 1973) whereas declarative complements simply denote simple propositions. The general idea is that the processing load of embedded questions, combined with the processing load of long-distance extraction, have a “super-additive” effect – a processing load that is more than the sum of its parts. This is enough to overload the processor. In particular, both of these features tax working memory beyond its resources, leading to the effect of reduced acceptability. Hofmeister & Sag (2010) are not precise about how the semantic complexity of embedded questions as compared to embedded declaratives taxes working memory, although one could imagine that the difficulty involves maintaining in working memory the kinds of alternatives required by questions (two such alternatives in embedded polar questions, possibly many such alternatives in *wh*-questions).

The processing-based approach to islands has been tested by Sprouse et al. (2012a,b). An important component of this work is their attempt to verify the proposed connection between acceptability and working memory. They charted participants’ working memory capacity, and asked whether there was a correlation between a participant’s working memory capacity and the strength of the super-additive effects of islands from ratings. They found no such correlation (although see the reply in Hofmeister et al. 2012, challenging both the statistical analyses in Sprouse et al. 2012a and their choice of working memory tests). What interests us in the context of embedding is the proposed cost of island structures independently of extraction. Sprouse et al. (2012a) systematically constructed factorial designs that help separate the difficulty of extraction from the difficulty of structures that involve islands. Below is an example of a 2×2 factorial design from their acceptability studies, in which all sentences are *wh*-questions. One factor was island status, whether the embedded clauses is an island structure (here a *whether*-clause) as in (35c), (35d) or a non-island declarative clause (35a), (35b). The second factor was the nature of the filler-gap dependency: within a clause (35a), (35b) or cross-clausal. They tested four types of islands: *whether*-islands and Complex NP islands (shown (35) and (36)) as well as subject islands, adjunct islands, and complex NP islands (36).

(35) *whether islands* stimuli (Sprouse et al. 2012a)

- a. Who ___ thinks that John bought a car NON-ISLAND, MATRIX GAP
- b. What do you think that John bought ___ NON-ISLAND, EMBEDDED GAP

- c. Who ___ wonders whether John bought a car ISLAND, MATRIX GAP
- d. What do you wonder whether John bought ___ ISLAND, EMBEDDED GAP

(36) *Complex NP islands* stimuli (Sprouse et al. 2012a)

- a. Who ___ claimed that John bought a car NON-ISLAND, MATRIX GAP
- b. What do you claim that John bought ___ NON-ISLAND, EMBEDDED GAP
- c. Who ___ made the claim that John bought a car ISLAND, MATRIX GAP
- d. What do you make the claim that John bought ___
ISLAND, EMBEDDED GAP

Like the studies reported from Kluender & Kutas (1993), the contrast between (36a) and (36c) serves as test of whether islands in general – even if they do not involve extraction from the island – bear an independent processing load. In an acceptability judgement task, Sprouse et al. (2012a) found an independent cost for *whether*-islands as well as subject and adjunct islands, but not complex NP islands. That is, there was no significant difference in acceptability between (36a) and (36c). Importantly, the island-violating condition in (36d) was still judged just as unacceptable as the other island violation conditions in the experiment. This finding calls into question the idea that the processing load of island structures, independent of extraction, contributes to a super-additive effect that leads to the unacceptability of all islands. There has been a recent expansion of experimental work using the factorial definition of islands to different languages, examining variation within and across languages and within and across island types (Sprouse et al. 2016, Kush et al. 2018, Pañeda et al. 2020, Lu et al. 2020).

There are also accounts of islands that reduce unacceptability to discourse-pragmatic properties (Erteschik-Shir 1973). Following Goldberg (2006), the studies in Ambridge & Goldberg (2008), Goldberg (2013), Abeillé et al. (2020) pursue the idea that backgrounded or discourse-given constituents block extraction of question phrases they contain because givenness is at odds with the pragmatic function of questions as seeking new information. Many islands constructions correspond to material that is often backgrounded in the discourse, e.g. subjects and factive complements. This approach needs to address whether non-question extraction (e.g. relatives, degree, topicalization) shows similar effects, and if so why, given that the discourse properties of questions are absent with such extractions. At the moment of writing this chapter, I believe that these questions are under active investigation.

Another “reductionist” approach to island phenomena is that proposed in Liu et al. (2022). They argue that frequency can explain the perceived unacceptability of island effects in factive and manner-of-speaking verbs. These are claimed

in the formal literature to resist extraction, at least of certain elements (**How tall did John know/whisper that the mountain was?*; cf. *How tall did John think that the mountain was?*). Liu et al. (2022)'s claim is that acceptability of extraction in these sentences depends on the language user's exposure to clausal complements generally, not any grammatical or even pragmatic feature of these constructions. Their verb frame frequency account posits that the more likely a verb is to appear with a clausal complement, the higher the acceptability of extraction from the complement. Liu et al. (2022), however, examined only *wh*-object NP extraction from factive and manner-of-speaking complements. Since at least Cinque (1990), if not earlier, it has been known that these types of 'weak' islands are *selective*, allowing some kinds of *wh*-phrase to extract rather readily but not so others, including adjuncts and degree phrases (Szabolcsi & Lohndal 2017). Indeed, Potter & Carlson (2019) and Williams et al. (2022) report that when these kinds of extractees are taken into consideration, an island effect emerges. Williams et al. (2022) further show that while frequency does affect acceptability, alongside this an effect of factivity remains. Further work, however, needs to pin down the etiology of factive islands. The consensus in the formal linguistics literature is that some form of semantic or pragmatic (Abrusán 2011, Schwarz et al. 2019) deviance is at play, and this is often tied to the nature of questions. As with other accounts noted above, further experimental work could be helpful in teasing apart the role of question phrase extraction versus extraction in general. Another area of interest is possible differences among embedding verbs. Very recently, Degen & Tonhauser (2022) contend that there is no crisp category of factive verb; rather, the extent to which a factive presupposition projects varies greatly by verb and is gradient. In my opinion, experimental work on clause embedding going forward will benefit from much more attention to subtle discourse-pragmatic effects, teasing these apart from purely syntactic phenomena. For some recent experimental and corpus work on issues involving the discourse-pragmatics of embedding (including factivity and embedded verb second), see Djärv (2019).

5 Attachment

5.1 Attachment *to* embedded clauses

Some of the earliest psycholinguistic work in sentence processing was devoted to showing that the clause is privileged as a major unit of processing (Jarvella 1971, Jarvella & Herman 1972, Caplan 1972, Frazier 1978, Marslen-Wilson et al. 1978, Bever & Townsend 1979). In these early studies, much of the attention was paid to what is retained after clausal and sentential boundaries. Generally, the

gross message of a clause is retained but not its linguistic form (Sachs 1967). For instance, Jarvella (1971) had participants hear passages like those in (37).

- (37) a. With this possibility, Taylor left the capital. After *he had returned to Manhattan*, he explained the offer to his wife.
 b. Taylor did not reach a decision until after *he had returned to Manhattan*. He explained the offer to his wife.

The clause of interest is in italics above: in (37a) it is part of the second sentence; in (37b) it is part of the first sentence. Participants were better at repeating verbatim this particular clause in cases like (37a), suggesting that the linguistic form of a clause is shunted from memory after it is closed.¹⁴

A relevant, but “little known” (Frazier & Clifton 2005) study by Walker et al. (1968) cited in Fodor et al. (1974), highlights the privileged status of the matrix clause. Walker et al. (1968) used a two-word probe experiment in which participants responded to whether the probe words were part of a previously seen sentence, like (38).

- (38) The scout the Indians saw killed a buffalo.

If the probe words were from different clauses (e.g., *Indians* and *buffalo*) participants responded more slowly than if they were drawn from the same clause. Interestingly, they were faster to respond to two words from the main clause (*scout, killed*) than two words drawn from the embedded clause. Frazier & Clifton (2005) suggest that this reflects that the matrix clause typically carries the main assertion of the sentence – “most accessible or salient part of the sentence.”

The results by Walker et al. are about the “message” of a clause and therefore its informational relevance. In terms of syntactic processing, by contrast, recent and embedded clauses typically define the domain of accessible material. For example, languages have a strong locality preference in attachment ambiguities involving two VP attachment sites, as in (39) Carreiras & Clifton (1993):

¹⁴This era saw a number of interesting studies concerning main and subordinate clauses. One popular question was the linear order of main (matrix) vs. subordinate clause. Clark & Clark (1968, 1977), Fodor et al. (1974) present evidence that the main-subordinate order is preferred, and memory is better when clauses are present in that order (but see Townsend & Bever 1978, Bever & Townsend 1979 on interesting reversals in terms of subordinator type). Clark (1973) found that when children begin to use embedding structures, they use the main-subordinate order. Jou & Harris (1990) found, in a sentence recall method, a bias toward main-subordinate structures. Another set of studies suggests that the temporal (or causal) order of the events described factors into the linguistic order (Opacic 1973), including developmental studies by Johnson (1975).

- (39) Juan dijo que Bill se murió (#morirá) ayer.
Juan said that Bill died (#will die) yesterday.

There is a strong preference, in both Spanish and English, to attach the temporal adjunct *yesterday* to the embedded clause, which is also the more recent clause. This holds even when the tense of the embedded clause – which is processed in advance of the adjunct – is semantically incompatible with *yesterday*. In classic Garden Path Theory (Frazier 1978, 1987a), this parsing preference falls under the principle of *Late Closure*, which requires the processor to attach incoming material to the phrase currently being processed, which in this case would be the embedded clause. The Dependency Locality Theory also prioritizes local attachment, cf. Gibson 2000.

There is a tension, then, in the processing of complex embedded structures. The syntax likes local, low attachment, whereas discourse pressures privilege matrix clause material. Inspired by a range of findings, including those in Walker et al. (1968), Frazier & Clifton (2005) propose the *Main Assertion Hypothesis* (see also the notion of relativized relevance by Frazier 1990).

- (40) *Main Assertion Hypothesis*

Other things being equal, comprehenders prefer to relate material in a new sentence to the main assertion of the preceding sentence.

In an interesting study, Traxler & Frazier (2008) test the Main Assertion Hypothesis by comparing attachment preferences with main vs. embedded clauses. In their stimuli, shown in (41), the phrase of relevance for attachment is the PP *to the store*: it can attach to the lower verb *checked* or the higher verb *delivered*. A lower attachment is consistent with syntactic attachment preferences (such as Late Closure), while a higher attachment is consistent with the Main Assertion Hypothesis. The prediction is that the Main Assertion Hypothesis will exert greater pressure if the higher verb belongs to a matrix clause (as in (41c), (41d)) and so higher attachments will be preferred here. If the higher of the two verbs itself belongs to an embedded clause, as in (41a), (41b), that clause does not serve as the main assertion, so the Main Assertion Hypothesis will not prefer high PP attachment. In addition to this clause-type factor, the grammatical attachment site of the PP was manipulated via verb selection properties: (41a), (41c) require higher verb attachment (since *checked to the store* is ungrammatical), whereas (41b), (41d) require low attachment.

- (41) a. Before the baker **delivered** the bread that he **checked** [to the store] on the corner, the customer stopped by for some donuts.

- b. Before the baker **checked** the bread that he **delivered [to the store]** on the corner, the customer stopped by for some donuts.
- c. The baker **delivered** the bread that he **checked [to the store]** early this morning, but it was too late.
- d. The baker **checked** the bread that he **delivered [to the store]** early this morning, but it was too late.

In contrast to predictions of the Main Assertion Hypothesis, if syntactic attachment alone matters, readers should have difficulty with the sentences in (41a), (41c) since low attachment leads to an ill-formed sentence. In an eye-tracking experiment, Traxler & Frazier (2008) found difficulty for the embedded clause condition (41a) but crucially not for the matrix clause condition (41c). That is, when the PP needed to be attached high, this was easier for readers if the clause it attached to was a matrix clause. A matrix clause is more likely to serve as a main assertion, and so the Main Assertion Hypothesis acts as a countervailing pressure against low attachment.

Frazier & Clifton (2005) report evidence for the Main Assertion Hypothesis from verb phrase ellipsis. In both sentences in (42), the elided VP could correspond to the whole matrix clause (Tina laughed after she made a joke) or the embedded clause (Tina made a joke).

- (42) a. Mary laughed after she made a joke about the supervisor. Then Tina did too.
- b. After Mary laughed she made a joke about the supervisor. Then Tina did too.

They found that participants chose to resolve the ellipsis with the matrix clause antecedent over 70% of the time in both (42a), (42b).

The gist of these findings is that both the syntax and discourse-pragmatic properties of embedded clauses influence processing, often in divergent and competing ways. These insights tie into a wealth of research on how discourse properties influence sentence processing, perhaps the earliest and most prominent of which is the contention by Crain & Steedman (1985) that the famous garden path in (43) (Bever 1970) is a result of pragmatic considerations, not context-blind syntactic parsing preferences.

- (43) The horse raced past the barn fell.

The target restrictive relative (RR) parse requires the listener to not only imagine a (unique) horse that was raced past the barn, but also a contrast set of horses

which do not have the property ascribed by the relative.¹⁵ This is a complex discourse model. The more-readily pursued main verb (MV) analysis only requires the reader to allow the existence of a horse, not a contrast set. In this sense, the MV parse requires a more parsimonious or simpler discourse model. Crain & Steedman (1985) propose a Principle of Parsimony: the reading which carries fewer unsupported presuppositions¹⁶ will be favored over one that carries more. Grodner et al. (2005) found that even without ambiguous input, readers find restriction difficult to process in a context that does not motivate restriction. They compared (unambiguous) restrictive and (unambiguous) non-restrictive relatives and found that in null contexts, the restrictive relatives were read more slowly. (When supporting contexts were provided, the effect reversed direction.) They argue that the results support the Principle of Parsimony, in that the restrictive relative requires a more elaborate discourse model (evoking a contrast set) which the non-restrictive does not.¹⁷ Additionally, several studies have shown that individuals prefer unmodified NPs in several ambiguous constructions and that the Principle of Parsimony is responsible (Ni et al. 1996, Spivey-Knowlton & Sedivy 1995, Sedivy 2002). Findings such as these are used to support constraint-based models of processing, in which a variety of linguistic and extra-linguistic bits of information can guide processing. A particularly stunning demonstration of this point is the study in Schwarz (2007), which shows that presupposition triggers can guide readers to otherwise syntactically dispreferred structures.

5.2 Attachment of embedded clauses

In the previous section, we examined ambiguities that arise when material can attach to either a matrix or an embedded clause. This section looks at ambiguities about where embedded clauses themselves attach. This serves as a case study revealing the importance of careful syntactic and semantic analyses to the study of embedded clause processing.

As we saw above, the processor generally likes to attach material low in the syntactic tree, which constitutes more local or recent material. And this appears

¹⁵Modifiers do not always implicate contrast – see below – but the claim here is that readers are more likely to pursue a contrastive interpretation for the relative clause (Sedivy 2003).

¹⁶Crain & Steedman (1985) used the word *presupposition* to describe the contrast set component, but Portner (1989) and Sedivy (2003) show that the contrast set evoked by a restrictive relative is a conversational/Gricean implicature.

¹⁷The principle of parsimony relies, in turn, on a more general pressure following from the Gricean Maxim of Quantity: Speakers should say as much as needed to be informative, without saying more than is necessary (Grice 1975). That is, since the simple discourse model is preferred, any additional modification in the description is redundant.

to hold up cross-linguistically, e.g., the low attachment of the adjunct in (39) in Spanish and English. Cuetos & Mitchell (1988), however, discovered an apparent exception to this. They found that English and Spanish contrast in the preferred attachment site of the relative clause in sentences like those in (44).

- (44) a. Someone shot the maid₁ of the actress₂ [who was standing on the balcony].
 b. Alguien disparó contra la criada₁ de la actriz₂ [que estaba en el balcón].
 ‘Someone shot at the maid of the actress that was on the balcony’

The issue is the attachment of the relative clause *who was standing on the balcony*: it can grammatically attach to the higher noun phrase (NP₁, which is headed by *maid*) or to the lower NP₂ (*actress*).

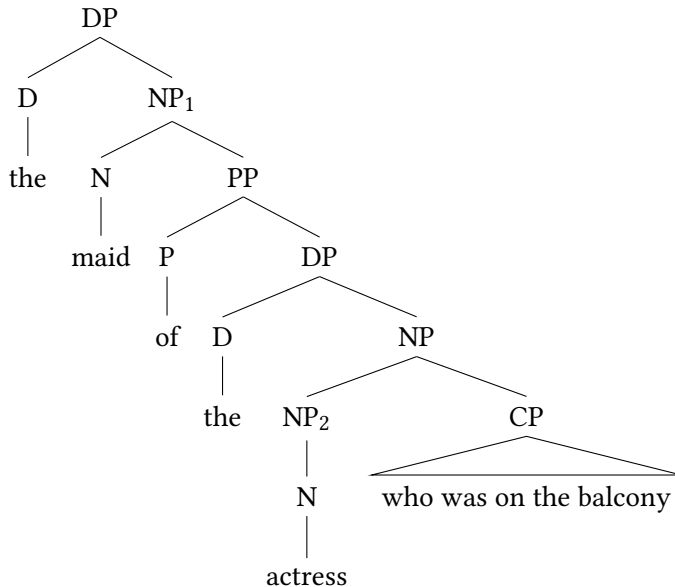


Figure 1: Low attachment

English speakers prefer low attachment, while Spanish speakers prefer high attachment (Cuetos & Mitchell 1988). Several languages show similar preferences for high attachment: French (Mitchell et al. 1990, Zagar et al. 1997, Frenck-Mestre & Pynte 2000, Colonna & Pynte 2001), Dutch (Brysbaert & Mitchell 1996, Mitchell & Brysbaert 1998, Mitchell et al. 2000, Desmet et al. 2002), Greek (Papadopoulou & Clahsen 2003), and Serbo-Croatian (Lovrić 2003). Other languages behave like

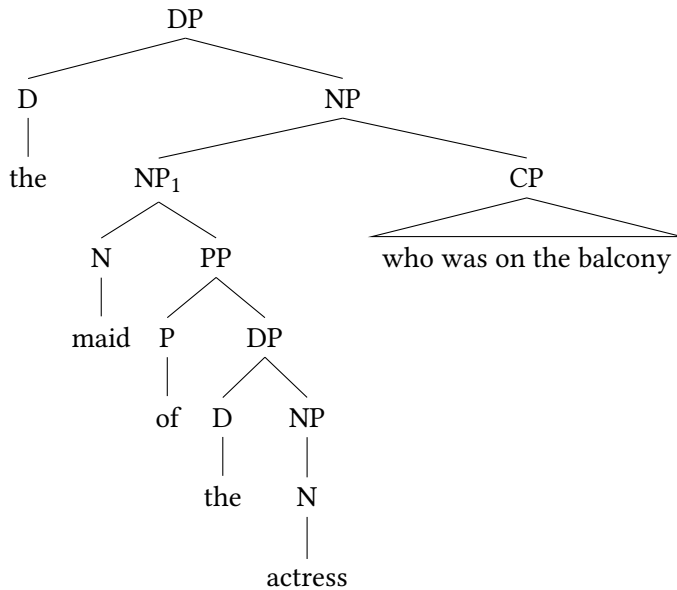


Figure 2: High attachment

English, including Basque (Gutierrez-Ziardegi et al. 2004), Romanian (Ehrlich et al. 1999), and Chinese (Shen 2006). These findings call into question the universality of parsing principles: why would some languages show preferences for high attachment and others for low attachment? Moreover, the presence of high attachment preference languages is puzzling from the perspective of locality: the attachment that requires minimal effort is NP₂. For decades, this cross-linguistic difference posed a major puzzle for processing.

In more recent work, Grillo and colleagues (Grillo & Costa 2014) have revealed that the stimuli used to test RC attachment suffer from a confound in certain languages. The problem is that languages like Spanish have a distinct construction called a pseudo-relative (PR) that can often be string-identical to a RC. Many of the sentences used to test RC attachment are actually ambiguous between a true RC structure and a PR. The PR structure, however, is only compatible, for reasons discussed below, with high attachment. If the sentence is parsed as a PR and not an RC – which Grillo shows is very often the case – then the high attachment behaviour does not have any bearing on attachment preferences.

Examples of pseudo-relatives from Spanish, Italian, French, and Greek are given below. PRs are superficially similar to relatives, since they involve noun phrases followed by a tensed clause with a gap. The easiest way to distinguish

PRs from RCs is to use a proper noun in the place of the head noun, as below. This forces either a PR parse or an appositive relative. The appositive parse, however, has a distinctive prosody, which the sentences in (45) (Grillo & Costa 2014) do not need to have.

- (45) a. He visto a **Juan que** corría. (Spanish)
 I.have seen to Juan that ran.IMPF
 ‘I saw Juan running.’
 b. Ho visto **Gianni che** correva. (Italian)
 I.have seen Gianni that ran.IMPF
 ‘I saw Gianni running.’
 c. J’ai vu **Jean qui** courait. (French)
 I.have seen Jean that ran.IMPF
 ‘I saw Jean running.’
 d. I Maria evlepe **ton Jani pu** etrexe. (Greek)
 the Mary watch.PAST.IMP the John.ACC that run.PAST.IMPF
 ‘Mary was watching John running.’

Unlike (appositive) relatives PRs are constructions in which the “head” and the finite clause are in a subject-predicate relation, and the PR as a whole describes an event, not the individual denoted by the “head.” This can be appreciated by the Italian cleft sentences in (46): the *wh*-element in (46a) is the inanimate *cio* which refers to the event of Mario crying, not Mario. If an animate *wh*-word (*chi*) is used then the post-copular phrase cannot be a pseudo-relative and must be an appositive.

- (46) a. Ciò₁ che ho visto è [Mario che piangeva]₁.
 That which I.have seen is Mario that cry.IMPF
 ‘What I saw was Mario crying’
 b. *Chi₂ ho visto è [Mario₂ che piangeva].
 Who I.have seen is Mario that cry.IMPF
 ‘Who I saw was Mario crying’¹⁸

The event-denoting nature of PRs restricts the types of predicates that can select it, which are typically event-selecting perception verbs. There are a number of other distinctions between RCs and PRs, including the fact that PRs allow

¹⁸After Radford (1977: 160 (98)). The example is o.k. under the interpretation of an appositive relative.

only subject gaps and that the tense of the PR and the matrix clause are highly restricted (they must have simultaneous interpretations).

In the above examples, the PR is shown with a proper noun head, but a common noun is possible as well. And, indeed that common noun can be a complex NP of the sort employed in the RC attachment sentences, as in (47), a French example.

- (47) a. J'ai vu le fils de l'homme qui courait.
 I.have seen the son of the.man that ran.
 'I saw the son of the man running.'

Recall that as in Spanish, sentences such as these were argued to show a high attachment preference. But (47) is in fact three-ways ambiguous. It could involve an RC attached to either of the NPs. Or it could involve a PR in which the “head” *the son of the man* serves as a PR subject whose predicate is the CP *qui courait* (Kayne 1975, Moulton & Grillo 2015, Grillo & Moulton 2016).

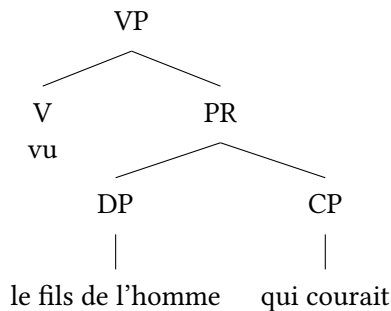


Figure 3: Pseudo-relative structure

If the PR parse is pursued for Figure 3, it must be the *son* who is running since the complex DP is the subject of the PR. It is not possible for a PR to be a complement of *fils* ‘son’ (via the preposition *de* ‘of’) since ‘son’ does not select for event-denoting phrases.¹⁹ In this sense, PRs force a “high attachment” of the clause.

¹⁹If the noun itself denotes an event, like the object language *evento* ‘event’, a PR can be introduced by a preposition.

- (i) l'evento di Maria che ballava.
 the.event of Maria that danced
 ‘the event of Maria dancing’

Grillo & Costa (2014) showed that removing the possibility of a PR parse in Italian reduced the high attachment preference. They compared sentences where the matrix verb was perceptual like *see* (which licenses either an RC or a PR) to those in which the matrix verb was a stative verb like *live with* (which does not license a PR, only an RC). For the perceptual verbs, the rate of high attachment was 78.6%. For the non-PR-taking stative verbs, the rate of high attachment went down to 24.2%. Similar results were found for other PR-languages (Fernandes 2012, Grillo & Spathas 2014, Tomaz et al. 2014, Aguilar & Grillo 2016). Pozniak et al. (2019) reach similar conclusions for French when PRs and RCs are disambiguated using tense properties. These results offer a lesson: informed linguistics is integral to understanding processing results and creating the appropriate stimuli. Attachment differences among languages lead to decades of research devoted to a finding that was, in part if not entirely, due to a simple linguistic confound between RCs and PRs.

Of course, if the presence of a PR analysis is what is responsible for the high attachment preferences found in languages like Spanish and Italian, then that means that readers are pursuing, in the face of ambiguity, a PR analysis over an RC analysis (see Pozniak et al. 2019 for discussion and ideas). Here there is an interesting connection to recent work exploring the difference between relative clause and clausal complements of nouns. We already saw, in Cowper (1976)'s examples (see (9)), that clausal complements of nouns, even if center-embedded, are easier to process than relative clauses (and their relative nesting is crucially important). Staub et al. (2018) have presented evidence that a complement clause (CC) parse (48a) is generally preferred over a relative clause (RC) structure (48b):

(48) Staub et al. (2018) Experiment 1 stimuli

- a. The information [_{CC} that the health department provided a cure]
reassured the tour operators. CC
- b. The information [_{RC} that the health department provided]
reassured the tour operators. RC

Their processing results suggest that the processor “avoids” an RC parse. They relate this to De Vincenzi's (1991) *Minimal Chain Principle* which militates against postulating filler-gap dependencies quite generally. This would favour CC analysis (which does not involve a filler-gap dependency) over the RC analysis. What is particularly interesting about Staub et al. (2018) is that the CC preference did not correlate with the general bias of the selecting noun (cf. Trueswell et al. 1993 discussed above). Staub et al. do note, however, that the role of context might modulate the differences, suggesting that a context that supports a contrast set for the

head noun (e.g. the existence of multiple pieces of information) might better motivate a RC, potentially reducing the CC preference or reversing the preference altogether. This question, I think, is particularly warranted in light of the earlier discussion in Section 5.1 of Grodner et al. (2005)’s finding that non-restrictive relatives show advantages over restrictive relatives in impoverished contexts. Complement clauses in complex noun phrases like (48b) bear a semantic relation to the head noun that, while still poorly understood, is often characterized as similar to an appositive one (see Stowell 1981, Moulton 2009, 2015, Krapova & Cinque 2016, de Cuba 2017). Additionally, it has been argued that definite complex noun phrases with complement clauses (as opposed to those with relative clauses) involve a weaker notion of definiteness (Hankamer & Mikkelsen 2021, Srinivas & Legendre 2024). These subtleties could influence the processing of CCs and RCs independently of the presence of a gap. Comparing CCs with appositive, non-restrictive relatives might help put the two embedded clauses on more equal footing for their contextual requirements. This, of course, would require a language where the appositive, non-restrictive relativizer/complementizer is identical to the CC-introducing complementizer.

More recently, Konrad et al. 2021 find further evidence in Italian and French that the parser avoids postulating gaps where possible. These studies are rather notable for their focus on syntactic ambiguities involving embedded clauses that have not been widely explored in the processing literature, including free relatives and indirect questions.

6 Conclusion

This overview has not touched on a number of important topics in clause embedding. One prominent lacuna are infinitival and other “reduced” embedded clauses. Infinitivals have played a large role in the processing of control structures (Frazier et al. 1983, Boland et al. 1990, Demestre et al. 1999, Green 2018). Another area is the processing of center-embedded infinitivals, verb clusters, and clause union phenomena (Vasishth 2003, Bayer et al. 2005), something that touches on famous differences between crossed and nested dependencies that arise in these constructions across languages (Bach et al. 1986). *That*-trace effects have also played a prominent role in the literature; for a recent overview see Cowart & McDaniel (2021). This overview has focused exclusively on comprehension rather than production; on production, the reader is referred to the interesting studies concerning multi- and embedded clause production in Ferreira & Swets (2017); for an overview of production see Momma (2021).

Embedded clauses are probably one of the psycholinguist's best tools. They are a testing ground for theories of syntactic parsing and memory maintenance and retrieval. In fact, it is rare to find psycholinguistic work on sentence processing that is squarely about embedding itself, rather than using embedded clauses as a tool to probe the architecture of the processor or the memory systems deployed for language comprehension. It is indeed the *drosophila* of sentence processing. This is as it should be: we do not choose the object of scientific study because it's our favourite specimen or construction, but because it offers a window into the things we want to know. But a word of caution: the linguistic landscape of embedding is complex and nuanced and there are many false friends. The con-found between relative clauses and pseudo-relatives is one good lesson, so are the nuanced theories of center-embedding that arise from taking into account Cowper's (1976) classic contrast between complement and relative clause nesting and center-embedding in head-final languages. As noted throughout this review, especially in connection with island effects (Section 4.2) and the processing of relative versus complement clauses (Section 5.1), some of the interesting questions going forward involve the discourse-pragmatic status of embedded clauses. At the same time, cross-linguistic coverage of work on processing embedding remains limited. Of course, investigating the processing of different types of embedding systems should be motivated by theoretical questions, be they psycholinguistic or formal linguistic.

If this chapter can offer anything, it is the advice to look at a broader range of embedded clause structures in sentence processing. This strategy has a good track record, when it is paired with precise questions about sentence processing. It profits both theories of processing and theories of syntactic and semantic structure. So maybe you can choose your favourite embedded structure to look at, but find a processing friend to explore it with.

References

- Abeillé, Anne, Barbara Hemforth, Elodie Winckel & Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204(104293). 1–23.
- Abrusán, Márta. 2011. Predicting the presuppositions of soft triggers. *Linguistics and Philosophy* 34. 491–535.
- Aguilar, Miriam & Nino Grillo. 2016. Testing the effect of pseudo relatives on relative clause attachment in Spanish. In *AMLap 22 (Architectures and mechanisms for language processing)*. Bilbao.

- Altmann, Gerry & Mark Steedman. 1988. Interaction with context during human sentence processing. *Cognition* 30. 191–238.
- Ambridge, Ben & Adele E. Goldberg. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19. 357–389.
- Babyonyshev, Maria & Edward Gibson. 1999. The complexity of nested structures in Japanese. *Language* 75. 423–450.
- Bach, Emmon, Colin Brown & William Marslen-Wilson. 1986. Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes* 1. 249–262.
- Bayer, Josef, Tanja Schmid & Markus Bader. 2005. Clause union and clausal position. In Marcel den Dikken & Cristina Tortora (eds.), *The function of function words and functional categories*, 79–113. Amsterdam: John Benjamins.
- Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes (ed.), *Cognition and the development of language*, 279–362. New York: Wiley.
- Bever, Thomas G. 1974. The ascent of the specious, or there's a lot we don't know about mirrors. In David Cohen (ed.), *Explaining linguistic phenomena*, 173–200. Washington: Hemisphere Press.
- Bever, Thomas G. & David J. Townsend. 1979. Perceptual mechanisms and formal properties of main and subordinate clauses. In William Cooper & Edward Walker (eds.), *Sentence processing: Psychological studies presented to Merrill Garrett*, 159–226. Hillsdale: Lawrence Erlbaum.
- Blumenthal, Arthur L. 1966. Observations with self-embedded sentences. *Psychonomic Science* 6. 453–545.
- Boland, Julie E., Michael K. Tanenhaus & Susan M. Garnsey. 1990. Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language* 29. 413–432.
- Brysbaert, Marc & Don C. Mitchell. 1996. Modifier attachment in sentence parsing: Evidence from Dutch. *The Quarterly Journal of Experimental Psychology. Section A* 49. 664–695.
- Caplan, David. 1972. Clause boundaries and recognition latencies for words in sentences. *Perception and Psychophysics* 12. 73–77.
- Carreiras, Manuel & Charles Clifton. 1993. Relative clause interpretation preferences in Spanish and English. *Language and Speech* 36. 353–372.
- Chen, Evan, Edward Gibson & Florian Wolf. 2005. Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language* 52. 144–169.
- Chomsky, Noam. 1957. *Syntactic structures*. Berlin: De Gruyter Mouton.

- Chomsky, Noam. 1977. *Conditions on transformations*. North-Holland: Elsevier. 81–162.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, Noam. 1986. *Barriers*. Cambridge: MIT Press.
- Chomsky, Noam & George A. Miller. 1963a. Finitary models of language users. In Duncan Luce, Robert Bush & Eugene Galanter (eds.), *Handbook of mathematical psychology*, vol. 2, 419–491. New York: Wiley.
- Chomsky, Noam & George A. Miller. 1963b. Introduction to the formal analysis of natural languages. In Duncan Luce, Robert Bush & Eugene Galanter (eds.), *Handbook of mathematical psychology*, vol. 2, 269–321. New York: Wiley.
- Christiansen, Morten H. & Maryellen MacDonald. 2009. A usage-based approach to recursion in sentence processing. *Language Learning* 59. 126–161.
- Cinque, Guglielmo. 1990. Ergative adjectives and the lexicalist hypothesis. *Natural Language and Linguistic Theory* 8. 1–41.
- Clark, Eve V. 1973. How children describe time and order. In Charles A. Ferguson & Dan I. Slobin (eds.), *Studies of child language development*, 586–606. New York: Holt, Rinehart & Winston.
- Clark, Herbert H. & Eve V. Clark. 1968. Semantic distinctions and memory for complex sentences. *Quarterly Journal of Experimental Psychology* 20. 129–138.
- Clark, Herbert H. & Eve V. Clark. 1977. *Psychology and language*. New York: Harcourt Brace Jovanovich.
- Colonna, Saveria & Joël Pynte. 2001. Relative clause attachment in French: The role of Fodor’s same size sister constraint. In *Paper presented at the workshop on prosody in processing, 5–6 july 2001*. Utrecht Institute of Linguistics.
- Cowart, Wayne & Dana McDaniel. 2021. The *that*-trace effect. In Grant Goodall (ed.), *The Cambridge handbook of experimental syntax*, 258–277. Cambridge: Cambridge University Press.
- Cowper, Elizabeth. 1976. *Constraints on sentence complexity: A model for syntactic processing*. Brown University. (Doctoral dissertation).
- Crain, Stephen & Janet D. Fodor. 1985. How can grammars help parsers. In David Dowty, Lauri Karttunen & Arnold Zwicky (eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives*, 94–128. Cambridge: Cambridge University Press.
- Crain, Stephen & Mark Steedman. 1985. On not being led up the garden path: The use of context by the psychological parser. In David Dowty, Lauri Karttunen & Arnold Zwicky (eds.), *Natural language parsing*, 94–128. Cambridge: Cambridge University Press.

- Cuetos, Fernando & Don C. Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition* 30. 73–105.
- de Cuba, Carlos. 2017. Noun complement clauses as referential modifiers. *Glossa* 2(1). 3. DOI: 10.5334/gjgl.53.
- De Vincenzi, Marica. 1991. *Syntactic parsing strategies in Italian: The minimal chain principle*. Dordrecht: Kluwer.
- Degen, Judith & Judith Tonhauser. 2022. Are there factive predicates? An empirical investigation. *Language* 98. 552–591.
- Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109. 193–210.
- Demestre, Josep, Sheila Meltzer, José E. García-Albea & Andreu Vigil. 1999. Identifying the null subject: Evidence from event-related brain potentials. *Journal of Psycholinguistic Research* 28. 293–312.
- Desmet, Timothy, Constantijn de Baecke & Marc Brysbaert. 2002. The influence of referential discourse context on modifier attachment in Dutch. *Memory & Cognition* 30. 150–157.
- Djäv, Kajsa. 2019. *Factive and assertive attitudes*. University of Pennsylvania. (Doctoral dissertation).
- Ehrlich, Karen, Eva M. Fernández, Janet D. Fodor, Eric Stenshoel & Mihai Vinereanu. 1999. Low attachment of relative clauses: New data from Swedish, Norwegian and Romanian. In *12th Annual CUNY Conference on Human Sentence Processing*. New York.
- Erteschik-Shir, Nomi. 1973. *On the nature of island constraints*. Cambridge: MIT. (Doctoral dissertation).
- Fernandes, Bruno. 2012. *Attachment preferences in prepositional infinitive constructions*. Centro de Linguística da Universidade Nova de Lisboa. (MA thesis).
- Ferreira, Fernanda & John M. Henderson. 1990. Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16. 555–568.
- Ferreira, Fernanda & Benjamin Swets. 2017. The production and comprehension of resumptive pronouns in relative clause ‘island contexts. In Anne Cutler (ed.), *Twenty-first century psycholinguistics: Four cornerstones*, 263–278. London: Routledge.
- Fodor, Janet D. 2013. Pronouncing and comprehending centre-embedded sentences. In Montserrat Sanz, Itziar Laka & Michael K. Tanenhaus (eds.), *Language down the garden path: The cognitive and biological basis for linguistic structures*, 206–228. Oxford: Oxford University Press.

- Fodor, Janet D., Stefanie Nickels & Esther Schott. 2018. Center-embedded sentences: What's pronounceable is comprehensible. In Roberto G. de Almeida & Lila Gleitman (eds.), *On concepts, modules, and language: Cognitive science at its core*, 139–168. Oxford: Oxford University Press.
- Fodor, Jerry A., Thomas G. Bever & Merrill F. Garrett. 1974. *The psychology of language: An introduction to psycholinguistics*. New York: McGraw-Hill.
- Fodor, Jerry A. & Merrill F. Garrett. 1967. Some syntactic determinants of sentential complexity. *Perception and Psychophysics* 2. 289–296.
- Ford, Marilyn. 1983. A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior* 22. 203–218.
- Ford, Marilyn, Joan Bresnan & Ronald M. Kaplan. 1982. A competence-based theory of syntactic closure. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 727–796. Cambridge: MIT Press.
- Frazier, Lyn. 1978. *On comprehending sentences: Syntactic parsing strategies*. University of Connecticut. (Doctoral dissertation).
- Frazier, Lyn. 1985. Syntactic complexity. In David Dowty, Lauri Karttunen & Arnold Zwicky (eds.), *Natural language parsing*, 129–189. Cambridge: Cambridge University Press.
- Frazier, Lyn. 1987a. Sentence processing: A tutorial review. In Max Coltheart (ed.), *Attention and performance XII: The psychology of reading*, 559–586. Hillsdale: Lawrence Erlbaum.
- Frazier, Lyn. 1987b. Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5. 519–559.
- Frazier, Lyn. 1990. Parsing modifiers: Special purpose routines in the human sentence processing mechanism. In David A. Balota, Giovanni B. Flores d'Arcais & Keith Rayner (eds.), *Comprehension processes in reading*, 303–330. London: Routledge.
- Frazier, Lyn & Charles Clifton. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4. 93–126.
- Frazier, Lyn, Charles Clifton & Janet Randall. 1983. Filling gaps: Decision principles and structure in sentence comprehension. *Cognition* 13. 187–222.
- Frazier, Lyn & Edward Gibson. 2015. *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor*. Heidelberg: Springer.
- Frazier, Lyn & Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14. 178–210.
- Frazier, Lynn & Charles Clifton. 2005. The syntax-discourse divide: Processing ellipsis. *Syntax* 8. 121–174.

- French-Mestre, Cheryl & Joël Pynte. 2000. 'Romancing' syntactic ambiguity: Why the French and the Italians don't see eye to eye. In Alan Kennedy, Ralph Radach, Dieter Heller & Joël Pynte (eds.), *Reading as a perceptual process*, 549–564. Amsterdam: Elsevier.
- Garnsey, Susan M., Neal J. Pearlmutter, Elizabeth Myers & Melanie A. Lotocky. 1997. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37. 58–93.
- Garrod, Simon C. 1994. Resolving pronouns and other anaphoric devices: The case for diversity in discourse processing. In Charles Clifton, Lyn Frazier & Keith Rayner (eds.), *Perspectives on sentence processing*, 339–357. Hillsdale: Lawrence Erlbaum.
- Garrod, Simon C. & Anthony J. Sanford. 1982. The mental representation of discourse in a focussed memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics* 1. 21–41.
- Geis, Michael. 1970. *Adverbial subordinate clauses*. Cambridge: MIT. (Doctoral dissertation).
- Gibson, Edward. 1991. *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Pittsburgh: Carnegie Mellon University. (Doctoral dissertation).
- Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68. 1–76.
- Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita & Wayne O'Neil (eds.), *Image, language, brain*, 95–126. Cambridge: MIT Press.
- Gibson, Edward, Timothy Desmet, Daniel Grodner, Duane Watson & Kara Ko. 2005. Reading relative clauses in English. *Cognitive Linguistics* 16. 313–353.
- Gibson, Edward & James Thomas. 1997. *The complexity of nested structures in English: Evidence for the syntactic prediction locality theory of syntactic complexity*. Tech. rep. Cambridge: MIT.
- Gibson, Edward & James Thomas. 1999. Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes* 14. 225–248.
- Gibson, Edward & Tessa Warren. 2004. Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax* 7. 55–78.
- Gildea, Daniel & David Temperley. 2007. Optimizing grammars for minimum dependency length. In *Proceedings of ACL* 45, 184–191.
- Gimenes, Manuel, François Rigalleau & Daniel Gaonac'h. 2009. When a missing verb makes a French sentence more acceptable. *Language and Cognitive Processes* 24. 440–449.

- Goldberg, Adele E. 2013. Backgrounded constituents cannot be ‘extracted’ from. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and island effects*, 221–238. Cambridge: Cambridge University Press.
- Goldberg, Adele E. 2006. *Constructions at work*. Oxford: Oxford University Press.
- Gordon, Peter C., Randall Hendrick & Marcus Johnson. 2001. Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27. 1411–1423.
- Gordon, Peter C., Randall Hendrick & Marcus Johnson. 2004. Effects of noun phrase type on sentence complexity. *Journal of Memory and Language* 51. 97–114.
- Grant, Margaret. 2013. *The parsing and interpretation of comparatives: More than meets the eye*. University of Massachusetts at Amherst. (Doctoral dissertation).
- Green, Jeffrey J. 2018. *Adjunct control: Syntax and processing*. University of Maryland. (Doctoral dissertation).
- Grice, Herbert Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Speech acts*, 41–58. New York: Academic Press.
- Grillo, Nino & João Costa. 2014. A novel argument for the universality of parsing principles. *Cognition* 133. 156–187.
- Grillo, Nino & Keir Moulton. 2016. Event kinds and the pseudo-relative. In Christopher Hammerly & Brandon Prickett (eds.), *Proceedings of NELS 16*, vol. 2, 11–20. Amherst: GLSA.
- Grillo, Nino & Giorgos Spathas. 2014. Tense and aspect modulate RC attachment: Testing the PR hypothesis in Greek. In *DGfS 36 Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft*. Oldenburg.
- Grodner, Daniel & Edward Gibson. 2005. Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science* 29. 261–290.
- Grodner, Daniel, Edward Gibson & Duane Watson. 2005. The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition* 95. 275–296.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69. 274–307.
- Gutierrez-Ziardegi, Eider, Manuel Carreiras & Itziar Laka. 2004. Bilingual sentence processing: Relative clause attachment in Basque and Spanish. In *17th CUNY Conference on Human Sentence Processing*. College Park.
- Hakes, David T., Judith S. Evans & Linda L. Brannon. 1976. Understanding sentences with relative clauses. *Memory & Cognition* 4. 283–290.
- Hale, Kenneth L. 2001. Navajo verb stem position and the bipartite structure of the Navajo conjunct sector. *Linguistic Inquiry* 32. 678–693.

- Halliday, Michael A. K. & Ruqaiya Hasan. 1980. *Cohesion in English*. London: Longman.
- Hamblin, Charles Leonard. 1973. Questions in Montague English. *Foundations of Language* 10. 41–53.
- Hankamer, Jorge & Line Mikkelsen. 2021. CP complements to D. *Linguistic Inquiry* 52. 473–518.
- Häussler, Jana & Markus Bader. 2015. An interference account of the missing-VP effect. *Frontiers in Psychology* 6. 766.
- Haviland, Susan E. & Herbert H. Clark. 1974. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior* 13. 512–521.
- Hawkins, John A. 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21. 223–262.
- Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Heim, Irene. 1982. *The semantics of definite and indefinite Noun Phrases*. University of Massachusetts at Amherst. (Doctoral dissertation).
- Hofmeister, Philip, Laura Staum Casasanto & Ivan A. Sag. 2012. How do individual cognitive differences relate to acceptability judgments? A reply to Sprouse, Wagers, and Phillips. *Language* 88. 390–400.
- Hofmeister, Philip & Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86. 366–415.
- Holmes, Virginia M. 1973. Order of main and subordinate clauses in sentence perception. *Journal of Verbal Learning and Verbal Behavior* 12. 285–293.
- Holmes, Virginia M., Alan Kennedy & Wayne S. Murray. 1987. Syntactic structure and the garden path. *The Quarterly Journal of Experimental Psychology Section A* 39. 277–293.
- Jaeger, T. Florian, Evelyn Fedorenko, Philip Hofmeister & Edward Gibson. 2008. Expectation-based syntactic processing: Anti-locality effects outside of head-final languages. In *21st annual CUNY Sentence Processing Conference*. University of North Carolina-Chapel Hill, NC.
- Jarvella, Robert J. 1971. Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior* 10. 409–416.
- Jarvella, Robert J. & Steven Herman. 1972. Clause structure of sentences and speech processing. *Perception and Psychophysics* 11. 381–384.
- Jennings, Francis, Bill Randall & Lorraine K. Tyler. 1997. Graded effects of verb subcategory preferences on parsing: Support for constraint-satisfaction models. *Language and Cognitive Processes* 12. 485–504.

- Johnson, Helen L. 1975. The meaning of *before* and *after* for preschool children. *Journal of Experimental Child Psychology* 19. 88–99.
- Jou, Jerwen & Richard J. Harris. 1990. Event order versus syntactic structure in recall of adverbial complex sentences. *Journal of Psycholinguistic Research* 19. 21–42.
- Kac, Michael B. 1981. Center-embedding revisited. In *Proceedings of the Annual Conference of the Cognitive Science Society*, vol. 3, 123–124.
- Kamp, Hans. 1981. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen & Martin Stokhof (eds.), *Formal methods in the study of language*, 277–322. Amsterdam: Mathematical Centre Tracts.
- Karlsson, Fred. 2007. Constraints on multiple centre-embedding of clauses. *Journal of Linguistics* 43. 365–392.
- Kayne, Richard. 1975. *French syntax: The transformational cycle*. Cambridge: MIT Press.
- Kimball, John. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2. 15–47.
- King, Jonathan & Marcel Adam Just. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30. 580–602.
- Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. In Peter Culicover & Louise McNally (eds.), *The limits of syntax*, 241–280. New York: Academic Press.
- Kluender, Robert & Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8. 573–633.
- Konieczny, Lars. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research* 29. 627–645.
- Konieczny, Lars & Philipp Döring. 2003. Anticipation of clause-final heads: Evidence from eye-tracking and SRNs. In *Proceedings of the 4th international conference on cognitive science and the 7th conference of the Australasian society for cognitive science*, 330–335.
- Konrad, Ingrid, Massimo Burattin, Carlo Cecchetto, Francesca Foppolo, Adrian Staub & Caterina Donati. 2021. Avoiding gaps in Romance: Evidence from Italian and French for a structural parsing principle. *Syntax* 24. 191–223.
- Krapova, Iliana & Guglielmo Cinque. 2016. On noun clausal ‘complements’ and their non-unitary nature. *Annali di Ca’Foscari* 50. 77–107.
- Kush, Dave, Terje Lohndal & Jon Sprouse. 2018. Investigating variation in island effects. *Natural Language and Linguistic Theory* 36. 743–779.

- Kutas, Marta & Kara D. Federmeier. 2007. Event-related brain potential (ERP) studies of sentence processing. In Gareth Gaskell (ed.), *The Oxford handbook of psycholinguistics*, 385–406. Oxford: Oxford University Press.
- Lee, Ming-Wei. 2004. Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research* 33. 51–73.
- Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106. 1126–1177.
- Levy, Roger. 2013. Memory and surprisal in human sentence comprehension. In Roger P. G. van Gompel (ed.), *Sentence processing*, 78–114. London: Psychology Press.
- Levy, Roger, Evelina Fedorenko, Mara Breen & Edward Gibson. 2012. The processing of extraposed structures in English. *Cognition* 122. 12–36.
- Levy, Roger & Frank Keller. 2013. Expectation and locality effects in German verb-final structures. *Journal of Memory and Language* 68. 199–222.
- Lewis, David, Shravan Vasishth & Julie A. Van Dyke. 2006. Computation principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10. 447–454.
- Lewis, Richard L. & Mineharu Nakayama. 2002. Syntactic and positional similarity effects in the processing of Japanese embeddings. In Mineharu Nakayama (ed.), *Sentence processing in East Asian languages*, 85–110. Stanford: CSLI Publications.
- Lewis, Richard L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29. 375–419.
- Lewis, Shevaun & Colin Phillips. 2015. Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research* 44. 27–46.
- Liu, Yingdong, Rachel Ryskin, Richard Futrell & Edward Gibson. 2022. A verb-frame frequency account of constraints on long-distance dependencies in English. *Cognition* 222(104902). 1–17.
- Lovrić, Nenad. 2003. *Implicit prosody in silent reading: Relative clause attachment in Croatian*. CUNY. (Doctoral dissertation).
- Lu, Jiayi, Cynthia K. Thompson & Masaya Yoshida. 2020. Chinese wh-in-situ and islands: A formal judgment study. *Linguistic Inquiry* 51. 611–623.
- MacDonald, Maryellen, Neal J. Pearlmutter & Mark Seidenberg. 1994. Lexical nature of syntactic ambiguity resolution. *Psychological Review* 101. 676–703.
- MacWhinney, Brian. 1977. Starting points. *Language* 53. 152–168.
- Marslen-Wilson, William, Lorraine K. Tyler & Mark Seidenberg. 1978. Sentence processing and the clause boundary. In Willem J. M. Levelt & Giovanni B. Flores d'Arcais (eds.), *Studies in the perception of language*, 219–246. New York: John Wiley.

- Miller, George A. & Stephen Isard. 1964. Free recall of self-embedded English sentences. *Information and Control* 7. 292–303.
- Mitchell, Don C. & Marc Brysbaert. 1998. Challenges to recent theories of crosslinguistic variation in parsing: Evidence from Dutch. In Dieter Hillert (ed.), *Sentence processing: A crosslinguistic perspective*, 313–335. New York: Academic Press.
- Mitchell, Don C., Marc Brysbaert, Stefan Grondelaers & Piet Swanepoel. 2000. Modifier attachment in Dutch: Testing aspects of construal theory. In Alan Kennedy, Ralph Radach, Dieter Heller & Joël Pynte (eds.), *Reading as a perceptual process*, 493–516. Amsterdam: Elsevier.
- Mitchell, Don C., Fernando Cuetos & Daniel Zagar. 1990. Reading in different languages: Is there a universal mechanism for parsing sentences? In David A. Balota, Giovanni B. Flores d'Arcais & Keith Rayner (eds.), *Comprehension processes in reading*, 313–335. Hillsdale: Lawrence Erlbaum.
- Mitchell, Don C. & Virginia M. Holmes. 1985. The role of specific information about the verb in parsing sentences with local structural ambiguity. *Journal of Memory and Language* 24. 542–559.
- Momma, Shota. 2021. Syntax and speaking. In Grant Goodall (ed.), *The Cambridge handbook of experimental syntax*, 714–740. Cambridge: Cambridge University Press.
- Moulton, Keir. 2009. *Natural selection and the syntax of clausal complementation*. University of Massachusetts at Amherst. (Doctoral dissertation).
- Moulton, Keir. 2015. CPs: Copies and compositionality. *Linguistic Inquiry* 46. 305–342.
- Moulton, Keir & Nino Grillo. 2015. Pseudo-relatives: Big and direct. In Thuy Bui & Denis Özyildiz (eds.), *Proceedings of NELS 45*, 193–202. Amherst: GLSA.
- Nakatani, Kentaro & Edward Gibson. 2010. An on-line study of Japanese nesting complexity. *Cognitive Science* 34. 94–112.
- Ni, Weijia, Stephen Crain & Donald Shankweiler. 1996. Sidestepping garden paths: Assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes* 11. 283–334.
- Opacic, Gordana. 1973. *Natural order in cognizing and clause order in the sentencing of conjoined expressions*. University of Illinois at Urbana-Champaign. (Doctoral dissertation).
- Osborne, Timothy. 2009. Comparative coordination vs. comparative subordination. *Natural Language and Linguistic Theory* 27. 427–454.
- Osterhout, Lee, Phillip J. Holcomb & David A. Swinney. 1994. Brain potentials elicited by garden-path sentences: Evidence of the application of verb infor-

- mation during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20. 786–803.
- Pancheva, Roumyana & Barbara Tomaszewicz. 2011. Experimental evidence for the syntax of phrasal comparatives in Polish. *University of Pennsylvania Working Papers in Linguistics* 17. 185–194.
- Pañeda, Claudia, Sol Lago, Elena Vares, João Veríssimo & Claudia Felser. 2020. Island effects in Spanish comprehension. *Glossa* 5(1). 21. DOI: 10.5334/gjgl.1058.
- Papadopoulou, Despina & Harald Clahsen. 2003. Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition* 25. 501–528.
- Partee, Barbara H. 1973. Some structural analogies between tenses and pronouns in English. *Journal of Philosophy* 70. 601–609.
- Pickering, Martin J. & Matthew J. Traxler. 1998. Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24. 940–961.
- Pickering, Martin J., Matthew J. Traxler & Matthew W. Crocker. 2000. Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language* 43. 447–475.
- Portner, Paul. 1989. Processing indefinite noun phrases in quantified sentences. *Five College Cognitive Science Program* 89-3. 78.
- Potter, David & Katy Carlson. 2019. *How we know factive predicates are islands: Experimental evidence*. Poster presented at Architectures and Mechanisms of Language Processing (AMLaP) 25.
- Pozniak, Céline, Barbara Hemforth, Yair Haendler, Andrea Santi & Nino Grillo. 2019. Seeing events vs. entities: The processing advantage of pseudo relatives over relative clauses. *Journal of Memory and Language* 107. 128–151.
- Radford, Andrew. 1977. Review of 'French syntax' by Richard Kayne. *Journal of Linguistics* 13. 118–128.
- Rayner, Keith & Lyn Frazier. 1987. Parsing temporarily ambiguous complements. *The Quarterly Journal of Experimental Psychology* 39. 657–673.
- Ross, John R. 1967. *Constraints on variables in syntax*. Cambridge: MIT. (Doctoral dissertation).
- Sachs, Jacqueline. 1967. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics* 2. 437–442.
- Santi, Andrea, Nino Grillo, Emilia Molimpakis & Michael Wagner. 2019. Processing relative clauses across comprehension and production: Similarities and differences. *Language, Cognition and Neuroscience* 34. 170–189.

- Schwarz, Bernhard, David Y. Oshima & Alexandra Simonenko. 2019. Factive islands from necessary blocking. In Katherine Blake, Forrest Davis, Kaelyn Lamp & Joseph Rhyne (eds.), *Proceedings of SALT 29*, 529–548. Linguistic Society of America.
- Schwarz, Florian. 2007. Processing presupposed content. *Journal of Semantics* 24. 373–416.
- Sedivy, Julie C. 2002. Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language* 46. 341–370.
- Sedivy, Julie C. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* 32. 3–32.
- Selkirk, Elisabeth. 1996. The prosodic structure of function words. In James L. Morgan & Katherine Demuth (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 187–214. Hillsdale: Lawrence Erlbaum.
- Shen, Xingjia. 2006. *Late assignment of syntax theory: Evidence from Chinese and English*. University of Exeter. (Doctoral dissertation).
- Spivey-Knowlton, Michael & Julie C. Sedivy. 1995. Resolving attachment ambiguities with multiple constraints. *Cognition* 55. 227–267.
- Sprouse, Jon, Ivano Caponigro, Ciro Greco & Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory* 34. 307–344.
- Sprouse, Jon, Matthew Wagers & Colin Phillips. 2012a. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88. 82–123.
- Sprouse, Jon, Matthew Wagers & Colin Phillips. 2012b. Working-memory capacity and island effects: A reminder of the issues and the facts. *Language* 88. 401–407.
- Srinivas, Sadhwi & Geraldine Legendre. 2024. Does D select the CP in light verb constructions? A reply to hankamer & mikkelsen (2021). *Linguistic Inquiry* 55. 595–621.
- Staub, Adrian, Francesca Foppolo, Caterina Donati & Carlo Cecchetto. 2018. Relative clause avoidance: Evidence for a structural parsing principle. *Journal of Memory and Language* 98. 26–44.
- Stowe, Laurie A. 1986. Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1. 227–245.
- Stowell, Timothy. 1981. *Origins of phrase structure*. Cambridge: MIT. (Doctoral dissertation).

- Szabolcsi, Anna & Terje Lohndal. 2017. Strong vs. weak islands. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell companion to syntax*, 2nd edn., 1–51. Oxford: John Wiley. DOI: 10.1002/9781118358733.wbsyncom008.
- Tomaz, Margarida, Maria do Carmo Lourenço-Gomes, Andrea Santi & Nino Grillo. 2014. A concordância de número em construções relativas e pseudorelativas em português europeu. In António Moreno, Fátima Silva, Isabel Falé, Isabel Pereira & João Veloso (eds.), *Textos selecionado: Encontro nacional da Associação Portuguesa de Linguística*, vol. 29, 519–532. Porto: APL.
- Townsend, David J. & Thomas G. Bever. 1978. Interclause relations and clausal processing. *Journal of Verbal Learning and Verbal Behavior* 17. 509–521.
- Traxler, Matthew J. & Lyn Frazier. 2008. The role of pragmatic principles in resolving attachment ambiguities: Evidence from eye movements. *Memory & Cognition* 36. 314–328.
- Traxler, Matthew J., Robin K. Morris & Rachel E. Seely. 2002. Processing subject and object relative clauses: Evidence from eye-movements. *Journal of Memory and Language* 47. 69–90.
- Traxler, Matthew J. & Martin J. Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35. 454–475.
- Trueswell, John C., Michael K. Tanenhaus & Christopher Kello. 1993. Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19. 528–553.
- Uehara, Keiko & Dianne Bradley. 1996. The effect of -ga sequences on processing Japanese multiply center-embedded sentences. In *Proceedings of the Pacific Asia Conference on Language, Information and Computation*, vol. 11, 187–196.
- Van Dyke, Julie A. & Richard L. Lewis. 2003. Distinguishing effects of structure and decay on attachment and repair: A retrieval interference theory of recovery from misanalyzed ambiguities. *Journal of Memory and Language* 49. 285–316.
- Van Dyke, Julie A. & Brian McElree. 2006. Retrieval interference in sentence comprehension. *Journal of Memory and Language* 55. 157–176.
- van Gompel, Roger P. G. 2013. *Sentence processing*. London: Psychology Press.
- Vasishth, Shravan. 2003. *Working memory in sentence comprehension: Processing Hindi center embeddings*. London: Routledge.
- Vasishth, Shravan & Heiner Drenhaus. 2011. Locality in German. *Dialogue & Discourse* 2. 59–82.
- Vasishth, Shravan & Felix Engelmann. 2021. *Sentence comprehension as a cognitive process: A computational approach*. Cambridge: Cambridge University Press.

- Vasishth, Shravan & Richard L. Lewis. 2006. Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language* 82. 767–794.
- Vasishth, Shravan, Katja Suckow, Richard L. Lewis & Sabine Kern. 2010. Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes* 25. 533–567.
- Wagers, Matthew & Colin Phillips. 2014. Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology* 67. 1274–1304.
- Walker, Edward, Paul Gough & Robert Wall. 1968. Grammatical relations and the search of sentences in immediate memory. In *Paper presented at the Midwestern Psychological Association*. Chicago.
- Wanner, Eric & Michael Maratsos. 1978. An ATN approach to comprehension. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic theory and psychological reality*, 119–161. Cambridge: MIT Press.
- Warren, Tessa & Edward Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85. 79–112.
- Waters, Gloria, David Caplan & Nancy Hildebrandt. 1987. Working memory and written sentence comprehension. In Max Coltheart (ed.), *Attention and performance XII: The psychology of reading*, 531–555. Hillsdale: Lawrence Erlbaum.
- Williams, Will, Dave Kush & Keir Moulton. 2022. *How weak are factive islands?* Poster presented at Architectures and Mechanisms of Language Processing (AMLaP) 28.
- Yngve, Victor H. 1960. A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society* 104(5). 444–466.
- Zagar, Daniel, Joël Pynte & Sylvie Rativeau. 1997. Evidence for early closure attachment on first pass reading times in French. *The Quarterly Journal of Experimental Psychology Section A* 50. 421–438.