NORTHWESTERN UNIVERSITY

Relations and Structure: A View from Dependency Resolution

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By

Kathleen Hall

EVANSTON, ILLINOIS

June 2021

ABSTRACT

This dissertation explores how items encountered in the comprehension of language are stored in memory and subsequently accessed. Processing and comprehending language frequently requires the retrieval of items in memory so that a current linguistic element can be assigned an interpretation. For example, in a sentence such as "Miles loved his father", the pronoun "his" has no intrinsic value but instead receives its interpretation from the noun "Miles". Another similar example concerns ellipsis, where phonologically silent material must select an antecedent from memory to be successfully interpreted.

In this work, I examine dependent elements such as pronouns and ellipsis in light of popular cue-based models of memory retrieval. While these models can accommodate a certain amount of syntactic information, maintaining large hierarchical structures and the relation between items in such a structure is problematic. To probe whether such structures and relations are available to the parser during dependency resolution, I employ a series of both offline tasks and eye tracking tasks. The first set of experiments uses the concept of 'parallelism' to explore whether the relation created by a pronoun and its antecedent is stored in memory, and whether such a representation can affect subsequent processing. The second set of experiments looks at sluicing type ellipsis, and challenges the claims of cue-based models by providing observations that antecedent size and complexity have an effect on processing. Finally, the last set of experiments employs verb phrase ellipsis to examine cases in which the antecedent and the interpretation of ellipsis is non-isomorphic. The overall results of these experiments provide evidence that highly detailed syntactic structure is both stored in memory and re-examined during dependency resolution, and that the content of retrieval is remarkably faithful.

Acknowledgements

What a very strange journey this has been. I imagine this is true for all who make it to this point – it's inevitable that during the course of an endeavor that spans so much time that life will happen too. Sometimes working on a graduate degree was the only constant, while people and events floated in and away and sometimes arrived with a life-changing bang. I've had the pleasure of knowing so many bright and lovely people who I hold dear, and have watched them succeed and fail, fall in and out of love, grow in ways large and small. I've done those things too.

I never would have made it through that first Chicago winter, let alone the next 7, without the constant warmth and support of my husband Sean Esler. We had only known each other for 4 months when he took a gamble and packed up a moving truck with me. It has paid off in the most marvelous of ways. We have both changed on so many dimensions since the day we hit the road, but we have done it together and are stronger and better for it. I might be a linguist, but I'm not sure I know the right words to express my love and gratitude, and so maybe for the rest of our lives I will just have to try to tell you a little bit each day with words or without.

I must also give thanks to my children, Miles and Mabel. Although you *probably* set this dissertation back a couple of years or so, you have made my life so incredibly rich and bright. Everything looked different after you. Suddenly my identity as a student, as an academic, was shared by something else no less demanding or rewarding. You took away some of the fear of failure, if I did right by you who cared if I asked a dumb question or didn't perform perfectly while giving a presentation? You spread my worth out in a way I was unable to do on my own and made it bigger and I love you.

Now that my family has been paid their dues and no one will disown me, I can say that this dissertation and degree would not exist without the exceptional and brilliant Masaya Yoshida. Someday I will give you a 'World's Greatest Advisor' mug and fill it up with scotch, I promise. I have been privileged to share and collaborate with Masaya, and never had to doubt if he would go to bat for me if I was in need. He always knew when to push and when to be compassionate, when to give advice and when to let me stand on my own. I don't think I could have made it through the so very steep learning curve of parenting as a grad student without knowing I could always rely on him. He made the space for me to succeed and I thank him for it.

My committee members Alexis Wellwood and Brian Dillon as well, of course, deserve special recognition. Alexis came to Northwestern and shook things up a bit, and I was sad to see her go but happy for the experiences I was able to have while she was here. Alexis always asks the big questions, and I try to imagine her wider perspective when I get too lost in the minutiae of some data and my questions get too small. Brian has been tremendous help in shaping this dissertation, and I am thankful for the advice and guidance both he and his work have been able to offer.

None of this would have happened without sharing the journey with my cohort members Nayoun Kim and Tommy Denby. I have spent too many hours to count in the lab with Nayoun and her dedication and ethic is inspiring. Our projects were the better for her thoughtfulness, and she always brought me back some of her mom's kimchi when she visited Korea so she is ok in my book. Tommy is the silliest person I know, and good god does academia need some silliness once in a while. He was and remains a true friend, a brilliant linguist, a confidant, and one of the very best people I love. You're my family, and I can't wait to get back to CA and celebrate our good fortunes.

So many others at NU deserve my gratitude and praise. Thanks go out to Emily Cibelli, Angela Kanita, and Peter Baumann for their camaraderie. To Chun Chan for being such a patient man and a technical wizard. For all those in the SSS lab, thank you for your expertise and consideration of my work, and for the drinks and discussions we have shared over the years.

In particular thank you to David Potter and Mike Frazier. You were both so very kind when I arrived with a fresh face and a lot of naivety. Mike, I have always admired you for the depth of your ideas, but most of all for the immeasurable size of your heart. You give so much of yourself. I know you don't always think it's enough but for me it has been immense and profound. David, we shared so much during your time at NU but have shared yet more since you moved on. Thank you for candor, your adoration of my family, your unending support, and your generous and open spirit. I hope we give you as much as you give us.

I must offer my appreciation and gratitude to the department as a whole, especially Matt Goldrick and Irene Saak. As a woman with children in academia you can't always expect all the doors to be open for you. I know how hard everyone worked to ensure that I had the support, backing, and encouragement to succeed. Thank you for always making it easy to take care of my work, myself, and my family over all these years. Thanks and love go out to those in ELP as well, especially Julia Moore. Julia has been a remarkable mentor and friend, has always opened her door for me, and routinely astounds me with the depth of her compassion and dedication to her staff and students. Your impact on me and those around you is deep, and I thank you for it.

At last to those of you had no direct hand in this dissertation, but always stood behind me. Thanks to my brothers and sisters, and the sisters of my heart Ann Hargens and Melissa Moe. To my mother as well, an absolute treasure who is proud of me as long as I'm doing whatever the hell I want. Some of you were loud and vocal to bolster me up, and others said little

because you knew if I wanted it then I would make it happen. It appears both approaches were successful. Finally, I wish you could have seen me do it dad.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Language comprehension in real time involves not only the incremental combination of linguistic units, but necessarily relies on forming relationships between items that have previously been processed and those introduced into the current parse. These kinds of non-adjacent relationships are termed ‘long distance’ dependencies, and require linking elements to other elements that may have been encountered quite far away in terms of time and linear distance. For such kinds of interpretation to occur, comprehending language requires the storage of previously processed material in memory, and the subsequent search and retrieval of that material to successfully link it to a dependent element and achieve a successful parse (Lewis, Vasishth, & Van Dyke, 2006; Wagers & Phillips, 2014; Wanner & Maratsos, 1978). This dissertation addresses the kind of information that can influence the memory retrieval process, and the content of that retrieval. As these issues are intricately linked with the subjects of memory storage and ultimate linguistic interpretations, this dissertation also sheds light on the representations of stored material, as well as how interpretations might differ from the content of retrieval.

To explore these topics, this dissertation asks the following questions: (1) Can relations in previously processed material affect subsequent processing? (2) Does an increase in size or complexity of a retrieved element affect processing? And (3), is there evidence that the parser alters or amends retrieved material in real time? To these ends, this study employs the investigation of pronoun-antecedent dependencies[1] and ellipsis constructions to probe the

[1] Note that throughout this thesis I at times refer to the coreference relation between a pronoun and a noun as a "dependency relation", although I acknowledge that pronoun–antecedent relationships may not be

underlying processes of memory representation and retrieval. In these various domains, I show that previously processed and stored relations between two elements in particular structural positions can affect the formation of subsequent dependencies, that structurally more complex antecedents are costly to the retrieval process, and that there are online consequences of ellipsis-antecedent mismatch that may reflect structure building at the ellipsis site and post retrieval operations. Common threads underlie these observations, specifically the importance of parallelism in various domains and the existence of hierarchical structural material in both memory and interpretation.

1.1 Introduction

A key feature of natural language is the ability to construct non-local dependencies, and considerable attention has been given to the processes underlying dependency formation, in particular to how items are retrieved from memory for subsequent processing. During online sentence comprehension, in order to process and achieve the proper interpretation of a dependency, the parser must recognize the tail of a dependency, access previously processed material, and determine which available candidate for the head of the dependency optimally

---

dependencies in the strictest sense of the term (Hudson, 1980, 1993; Kimball, 1973). However, I also note that there are ample studies that employ this terminology to refer to the relation established between a pronoun and its antecedent (e.g. Grant, Sloggett, & Dillon, 2020; Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2007; Kwon, 2008; Piñango, Finn, Lacadie, & Constable, 2016; Yoshida, Kazanina, Pablos, & Sturt, 2014), and in line with this previous literature I use this term to refer to the fact that the interpretation of a pronoun is determined by that of its antecedent.

resolves it. A great deal of recent research has probed how the parser accomplishes this by examining various linguistic relations such as wh-gap dependencies (Fodor, 1989, 1993; Kim, Brehm, Sturt, & Yoshida, 2019; McElree, 2000; McElree, Foraker, & Dyer, 2003; Nicol, Fodor, & Swinney, 1994; Parker, 2017), subject-verb agreement (Dillon, Mishler, Sloggett, & Phillips, 2013; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Pearlmutter, Garnsey, & Bock, 1999; Van Dyke & McElree, 2011), reflexive-antecedent dependencies (Badecker & Straub, 2002; Cunnings & Felser, 2013; Dillon, 2011; Dillon et al., 2013; Parker & Phillips, 2017; Sturt, 2003), and ellipsis-antecedent dependencies (Kim, Brehm, & Yoshida, 2018; Martin & McElree, 2008; Shapiro & Hestvik, 1995; Yoshida, Dickey, & Sturt, 2013), among others.

A central issue to the examination of these dependencies is concerned with the search space available when the parser attempts to locate a controlling element, and how this space is navigated. At times it appears quite obvious which element needs to be retrieved, such as 'Miles' in (1a), however in other instances (1b and 1c) several candidates for retrieval may exist in memory, 'Sean' and 'Miles' may both be considered. It is the latter cases which prove most informative to the mechanisms involved in dependency resolution. In general, two main issues are at play here. The first concerns the *manner* in which the search space is examined, and the other concerns the *type* of information that is relevant during antecedent search. Proposals regarding the first issue have in general detailed two main types of approaches: serial search mechanisms and cue-based mechanisms. Approaches to the second issue have appealed to systems that favor feature matching based on lexical, morphological, and syntactic features, or have advocated for systems that give precedence to structural positions. These different alternatives and combinations all have their appeal and are explained in more detail below.

(1)

a. Miles indulged himself.

b. Sean said Miles indulged himself.

c. Sean said Miles indulged him.

Navigating the search space

Retrieval of a linguistic element from memory is initiated via the cues or instructions specified by the dependent element, but what precisely these instructions consist of has been extensively and enthusiastically discussed in recent research. Earlier models of the search process proposed serial search mechanisms (Ehrlich, 1980; Gordon & Hendrick, 1998; McElree & Dosher, 1989; McElree et al., 2003), in which information in a specified domain is iteratively considered according to some ordering principle, e.g. recency or prominence. One notable benefit of such models is the ability to accommodate structural specifications or relations that are pertinent to the dependent element (e.g. c-command), and it may be that traversal of this nature is selectively engaged (Dillon, 2014). However, these models also predict that search time should rise as the content and complexity of the memory representations increase, an effect which has not reliably surfaced in experimental studies.

More recently, however, the majority of research has found support for a cue-based mechanism in content-addressable memory (Lewis et al., 2006; Martin & McElree, 2008, 2011; McElree et al., 2003). A primary attraction of this group of models is that they are based upon general principles of cognitive processing and avoid the domain specificity entailed by models that use specialized linguistic mechanisms. Cue-based models assume that individual items are encoded in memory during initial processing with certain features, and that matching these

features to those of the dependent element is the basis of retrieval itself. Features may come from the lexical entry of an item, such as person, number, or gender, or may index the part that item plays in the local syntactic environment, for example grammatical role. Cues provided by the dependent element provide direct access to memory representations, and this access occurs in parallel for multiple candidates. Key predictions of such models are that multiple partially or fully matching memory representations will result in interference and/or retrieval errors, and that the time course of access is constant regardless of the content of the representation.

Much of the evidence to support cue-based models has emerged from investigations of whether the parser respects structural constraints when locating and evaluating the antecedent of a dependency. Under popular cue-based models, syntactic structure generally should have little or no effect on how the antecedent is located and retrieved: all possible antecedents are accessed and evaluated, regardless of structural position. This lack of structural sensitivity of the retrieval mechanism predicts that it should be possible to observe interference during retrieval from syntactically 'inaccessible' positions. This claim appears to be confirmed in numerous studies, including in filler-gap dependencies (Gordon, Hendrick, & Johnson, 2001; Van Dyke & McElree, 2006), negative polarity item dependencies (Orth, Yoshida, & Sloggett, 2020; Xiang, Dillon, & Phillips, 2009), and most notably in the examination of number attraction in subject-verb agreement, which refers to cases where a subject and verb do not agree, but a plural distractor causes an increase in perceived grammaticality (Dillon et al., 2013; Lago et al., 2015; Pearlmutter et al., 1999; Wagers, Lau, & Phillips, 2009) as shown in (2). In (2), the illusion of grammaticality can be explained if number is used as a cue to retrieve a nominal head, and the local noun is incorrectly retrieved even though it is not in a structurally licit position to license plural agreement morphology on the verb.

(2) *The key to the cabinets **were** rusty from many years of disuse.

Further support for cue-based models comes from investigations of antecedent length and complexity. One notable example comes from a series of studies performed by Martin and McElree (2008). Over the course of several experiments, they manipulated antecedent complexity in various ways, for example by alternating between simple and complex noun-phrases, and by contrasting antecedents that had simple verb-phrases against antecedents that contained wh-complements. In none of these experiments did antecedent complexity result in longer processing times as reflected by a speed accuracy trade-off (SAT) task. In fact, there was some indication in one experiment that complex antecedents resulted in *faster* processing as compared to simple antecedents. Similar results (although without evidence of a speed-up due to complexity) were found in Paape et al. (2017) using a self-paced reading paradigm.
Although these studies are often cited as providing evidence for the implementation of a cue-based retrieval mechanism in the construction of long-distance dependencies, other studies suggest that the search process may be structurally constrained, and that the parser may respect structural properties of the antecedent. Particularly compelling evidence for a grammatically sensitive retrieval mechanism comes from investigations of forwards and backwards anaphora, and the online application of the binding principles that constrain their distribution (Kazanina et al., 2007; Nicol & Swinney, 1989; Nicol, 1988; Sturt, 2003; Xiang et al., 2009). Another example of such research comes from Dillon et al. (2013), who directly compare the processing of reflexive dependencies to subject-verb agreement dependencies in English, as shown in (3). Dillon et al. observe retrieval interference in examples like (3b), which involves subject-verb

agreement, however fail to observe an interference effect in examples like (3a), which involve a reflexive dependency. This contrast in processing between dependency types is difficult to explain under a cue-based model without invoking additional mechanisms that provide the parser instructions as to which structural positions are relevant for the search process.

(3)

a. The new executive who oversaw the middle manager(s) apparently doubted himself/*themselves on most major decisions.
b. The new executive who oversaw the middle manager(s) apparently were dishonest about the company's profits.

To give a clear and concrete case of how search is carried out, examine the example in (1) reproduced below in (4). The reflexive pronoun 'himself' in (4b) initiates a search in memory for an appropriate antecedent. By hypothesis, the instructions for such a search specified by the dependent element 'himself' could take the form of features that are required to match those of the controlling element, or those instructions could specify which structural positions to examine as a host for an appropriate antecedent, and then evaluate elements in those positions as possible matches. These two possibilities are predicted to play out in different fashions under serial search and cue-based models. For example, an exclusively feature matching system employing cues such as +*masculine* and + *singular* would consider both 'Sean' and 'Miles' as potential antecedents for the reflexive 'himself, resulting in competition and interference. In contrast, a serial search mechanism respecting serial order would perform a backwards traversal, encounter the referent 'Miles', and would successfully match that noun phrase (NP) to the reflexive. In (4c), while the cue-based predictions are identical to those of (4b), serial search first considers

the local subject, and then moves on to the more distant subject, a process which is predicted to increase search time.

(4)

a. Miles indulged himself.

b. Sean said Miles indulged himself.

c. Sean said Miles indulged him.

Clearly the above alternatives are lacking a crucial piece of information regarding the resolution of pronouns, as it is well known that in addition to morphological features, the resolution of pronouns is governed by syntactic constraints involving c-command and the structural configuration of the dependency (Chomsky, 1981b). In 4b, while both NPs are in a position to c-command the reflexive, only the NP ‘Miles’ is within the local domain of the pronoun. How the parser implements these constraints in various retrieval frameworks is not always straightforward. Although the common consensus has been reached that the parser uses a cue-based retrieval system that evaluates candidates in parallel, the dynamic character of syntactic relations such as c-command are difficult to represent as features (for a way of circumventing this issue see Kush, Lidz, & Phillips, 2015). Serial search models have the benefit of being able to traverse structure, and so a model that prioritizes certain structural positions is easier to accommodate in these frameworks.

Memory representations

The discussion of *how* items are retrieved unavoidably requires a discussion of *what* is actually represented in memory after processing and encoding of a linguistic unit. Regardless of the

search theory, representations must contain sufficient detail to distinguish them from other similar elements and provide ample information to be matched to the specifications indicated by the dependent element. Much of what we know about what information is encoded and stored comes from experimental evidence observing the impact of a particular feature on retrieval, for example the observation of a gender mismatch effect when attempting to link a pronoun and an antecedent indicates that gender is a feature that can be associated with a memory representation. Given the quantity of evidence suggesting feature matching as a component of memory retrieval in language processing, I will assume that memory representations consist of feature bundles, although I note that this does not entail that retrieval and interpretation solely consist of feature matching.

The most explicit description of how linguistic information is stored in memory and subsequently retrieved comes from the influential model proposed by Lewis and Vasishth (2005). In this model, units (chunks) are defined as 'a representational element that enters into novel relations with other elements' (p. 381). An example from their paper is shown below, where a complete structure on the left is broken down into its component parts for storage in declarative memory.

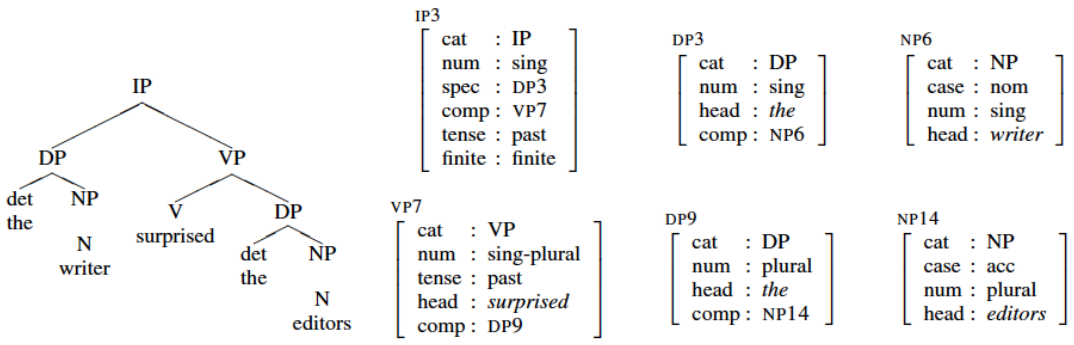


Figure 1: Example of chunks in the model proposed by Lewis and Vasishth (2005)

As one can observe from this illustration, these chunks can be imbued with a variety of features, including but not limited to lexical and morphological features such as number, gender, and phrasal category, and additionally can accommodate the immediate structural environment of the chunk. The authors assume an X-bar structure (Chomsky, 1986), encoded such that chunks represent maximal projections with features corresponding to X-bar positions (specifier, comp, head), and that much grammatical knowledge is represented procedurally in specific production rules residing in declarative memory.

One difficulty with this theory of representation is its (in)ability to represent relational information, notably c-command, but in essence any information defined over two constituents that are not referenced in the same chunk. An issue that will arise in Chapter 1 of this dissertation involves the apparent influence of parallelism of dependencies, which appears to be instance of what I might term 'indexical priming', such that the establishment of coreference in one clause influences the selection of an antecedent for coreference in a subsequent clause. How such a relation like this might be represented in a cue-based model is unclear.

Structural information

In addition to the nature of structural information in memory, a longstanding question concerns whether hierarchical syntactic structure is recovered from an antecedent in the processing of dependencies. Although some less widespread theories maintain that a pronoun is interpreted from the full representation of the antecedent (e.g. Elbourne, 2001; Postal, 1969), this debate has primarily concerned itself with the representation and processing of ellipsis. Two questions are open in this debate: (1) is there unpronounced structure in an ellipsis site, and (2) if so, is that

structure recovered or copied directly from the antecedent? Addressing this first question, there appears to be a fair amount of evidence that the parser indeed builds or incorporates structural material in the ellipsis site. One important source of this evidence come from connectivity effects, in which a fronted element is interpreted as occurring in a position lower than its surface form, and compatibility between the overt element of the elided clause and some part of the elided structure is required. A widely cited example is the requirement of case matching in overt case marking languages such as German, in which the case of the fronted wh-element must match the case of its non-elliptical counterpart (Merchant, 2001; Ross, 1969b), as shown in (5).

(5)

a. Case Matching (German)

   Er will jemandem schmeicheln, aber sie wissen nicht

   he wants someone.DAT flatter but they know not

   *wer/ *wen/ wem.

   who.NOM who.ACC who.DAT

   'He wants to flatter someone, but they don't know who' _

b. Er will jemandem schmeicheln, aber sie wissen nicht

   he wants someone.DAT flatter but they know not

   **wer/ *wen/ wem er schmeichelen will.*

   who.NOM who.ACC who.DAT he flatter wants

   'He wants to flatter someone, but they don't know who he wants to flatter'

Another domain in which convincing evidence emerges is in the examination of the interaction of island effects and ellipsis. Syntactic islands are structures which block filler-gap

dependencies (Ross, 1967) as in (6a), and island effects have been observed during the interpretation of ellipsis as in (6b), indicating that there is detailed structural information in the ellipsis site. Example (6a) is an example of the complex NP constraint, which bans extraction from a relative clause modifying a noun. (6b) is similarly unacceptable, despite the fact that the relative clause in question has been elided. The unacceptability can be explained if there is phonologically null structure at the ellipsis site from which the wh-phrase has been extracted.

(6)

a. *They want to hire someone who speaks a Balkan language, but I don't know which (Balkan language) they want to hire someone who speaks *t*

b. *They want to hire someone who speaks a Balkan language, but I don't know which (Balkan language) they [$_{VP}$ do ~~want to hire someone who speaks *t*~~]

Island effects have been also been observed in sprouting constructions, which are themselves indicative of structural material in ellipsis. Sprouting (Chung, Ladusaw, & McCloskey, 1995) refers to a type of ellipsis in which an explicit wh-correlate does not exist in the antecedent clause, for example in (7). The example of sluicing in (7a) has an explicit object 'something' in the antecedent as the correlate to 'what', however no such correlate exists in the antecedent of (7b).

(7)

a. John smoked something, but I don't know what.

b. John smoked, but I don't know what.

The processing of sprouting type ellipsis constructions has been associated with a penalty at the ellipsis site (Dickey & Bunger, 2011), which has been attributed to a lack of structural parallelism between the antecedent and the elided material. Their results additionally support the idea that structural material is copied or recreated in the ellipsis site, and that in the case of sprouting, reanalysis must occur to 'add in' the missing structural node in order for the wh-phrase to receive its thematic role and create an interpretable wh-gap chain.

Extensions of these findings were pursued by Yoshida and colleagues (Yoshida, Lee, & Dickey, 2013), who compared sluicing and sprouting types of ellipsis while manipulating whether the antecedent would create an island violating structure inside of an ellipsis site. Although many types of ellipsis do not ameliorate island effects, sluicing appears to be insensitive to islands (Chomsky, 1972; Chung et al., 1995; Lasnik, 2001; Merchant, 2001), while sprouting is not (Chung et al., 1995). In the sentences in (8) from Yoshida et al. (2013), the sentences in (8a-b) do not contain any island, however the sentences in (8c-d) contain an adjunct island in the first clause. Inspecting the region at and directly after the wh-phrase, the researchers predicted that an effect of sprouting would only emerge in non-island violating ellipsis configurations (8b), which was indeed observed in a self-paced reading study. The authors take this as evidence that the parser did not attempt to create a wh-gap dependency across an island in sentences like (8d), which heavily suggests that the parser considers detailed syntactic structure during the processing of ellipsis. Although not as relevant to the current discussion, it is also an intriguing result that none of these sentences in fact contained ellipsis, but ellipsis was a possible continuation at the regions of interest, indicating that the parser did not wait for bottom-up confirmation of ellipsis before positing an elliptical structure.

(8)

a/b. Nick's father discovered [that he smoked something/secretly in the garden] …

c/d. Nick's father was startled [because he smoked something/secretly in the garden]…

… but it wasn't clear **what** precisely he got out of smoking in hiding.

The research discussed above points strongly to the presence of fairly detailed syntactic structure inside the ellipsis site, however how that structural information arrives at the ellipsis site is less clear. The terms 'structure building' and 'copying' are at times used interchangeably, and discussions that link these processes to the storage and retrieval of items in memory are sometimes lacking or insufficiently specified. Returning to the representation of syntactic information in cue-based architectures, any given chunk can point to its syntactic neighbors in terms of X-bar relations, but no chunk 'contains' another in a recursive manner. However, hierarchical structure can be represented via the use of pointers to other phrases, which indexes the relationship between chunks in memory. It is crucial that this information is present, as we know from previous research, for example the syntactic priming literature (Branigan, Pickering, & McLean, 2005; Kim, 2006a; Kim, 2006b; Kim, Carbary, & Tanenhaus, 2014, among many others), that hierarchical information must be represented in some fashion. For example, in a speech production task, Xiang et al (2019) manipulated whether participants read and repeated conjoined sentences that included either a prepositional dative (NP PP) or a double object construction (NP NP) in the first clause, and a continuation in the second clause that either had parallel overt structure, VPE, or a simple intransitive verb. Participants were then asked to describe an image. The authors found a significant priming effect such that conditions in which

VPE occurred primed participants' productions to an equal extent as sentences with overt content, but no priming effect was witnessed in the intransitive verb conditions.

Accounts which assume structure at the ellipsis site have generally described its arrival there via some type of null copy (Chung et al., 1995; Fiengo & May, 1994; Frazier & Clifton, 2001; Merchant, 2001; Sag, 1976; Yoshida, Dickey, et al., 2013); cited here are notable accounts however this idea has been present since at least Ross (1969). These accounts may differ in terms of the derivational status of elliptical structure (pre spell-out, post spell-out, or both), however they assume that at *some* level there is silent syntactic structure. Copy accounts are often contrasted with cue-based or pointer accounts, however disentangling their predictions requires certain assumptions about what it means to copy or share a structure. For example, in Frazier and Clifton (2001), the authors proposes a 'cost-free copy' mechanism, in which antecedent structure is recycled into the ellipsis site, but assert that this operation does not impose a penalty on the parser. They find support for this proposal in an investigation of sluicing, where they do not observe an effect of antecedent length on self-paced reading times. This observation is consistent with their hypothesis as well as with complementary studies that find evidence of structure within the ellipsis site, however the lack of a structural complexity effect on ellipsis resolution has also been used to make the contrasting argument that merely a pointer to a memory representation is present at the ellipsis site (e.g. Martin & McElree, 2008).

What does it mean for structure to be copied? Describing the difference between sharing, reactivating, pointing to, building, reconstructing, or inferring syntactic structure is rarely explicit, and it seems that many proposals use some or all of these terms interchangeably. How could these operations (if different) proceed, and given what we assume about the representation of items and memory how they are accessed? It is difficult to argue against the fact that the

relevant memory unit(s) must be reactivated, and this can proceed straightforwardly following cue-based models that rely on retrieval cues at the dependent element. The main issue of this discussion regards what happens next, or if reactivation itself is sufficient for successful interpretation. It could be that the 'sharing' of these feature bundles is the end of retrieval, however if that is the case it is unclear how a compositional semantic interpretation is achieved in the absence of a syntactic interpretation. Alternatively, one might also consider an additional process that looks more typically like parsing, in which the reactivated material is used as input in a structure building process launched by the ellipsis site, or perhaps even by the prediction of an ellipsis site.

Relations and Structure

The implementation of the process discussed above by the parser and the retrieval mechanism has attractive qualities. Quite obviously it supplies the structural material that we see evidence of in ellipsis, but it also allows for the recreation of relations that are difficult to represent in feature-based memory architectures, for example c-command. C-command presents an encoding challenge for memory storage, especially as the distance between two elements grows. Consider two elements, X and Y, that are in a relationship such that X c-commands Y. At the point of encountering X, Y is not yet part of the local structural environment as it has not yet been encountered. To effectively encode c-command relations during incremental processing, as each new item is stored all previous items in memory must be updated with to reflect whether they do or do not c-command the new item. There are instances in which this could be relatively easy for

the parser, however it is clear that as the amount of material in the parse increases this is a computationally heavy and unlikely process.

The issue of representing c-command in memory demonstrates the overall difficulty of representing any item-to-item relations in feature-based models. Another example that may arise less frequently is the indexical relation created by the resolution of a dependency in previously processed material. Consider again X and Y, however in this scenario imagine that Y is a pronoun that selects X as its antecedent. Upon encountering Y, X is retrieved and co-indexed with the new item Y. This process seems computationally quite feasible, however as we will see in Chapter 1, there is evidence that the relationship between X and Y is represented in memory with reference to the relative structural positions of X and Y, and that the structural relationship is sufficiently independent to influence subsequent relationships that are not tied to any particular lexical items. Encoding this kind of structural relation, while not as computationally difficult as the case of c-command, again poses a challenge for feature-based models.

(Non) Isomorphism

The final issue of relevance for this study is the issue of isomorphism in the resolution of dependencies, in particular the matter of isomorphism of relational items and material such as anaphors and indices during online processing. Many theories and experimental studies have addressed the identity condition for ellipsis, with ample accounts supporting theories of syntactic identity (e.g. Chung et al., 1995; Fiengo & May, 1994; Hestvik, 1995; Lappin, 1992; Sag, 1976; Williams, 1977), semantic identity (Dalrymple, Shieber, & Pereira, 1991; Ginzburg & Sag, 2000; Hardt, 1993; Keenan, 1971; Merchant, 2001; Sag & Hankamer, 1984), or some hybrid of the two

(Chung, 2013; Kehler, 2002; Merchant, 2013). This is an active and dynamic debate, and can provide insight into the kind of material we believe is explicitly represented in memory.

Of particular interest in the domain of isomorphism is how interpretations are achieved when there is apparent non-isomorphism between an antecedent and the interpretation of an ellipsis site. Investigating such scenarios has occupied a certain subset of the field for some time, but there remains observations and conclusions to be mined from this vein. In light of the topics I have already discussed, I am especially interested in what the processing of non-isomorphic interpretations can tell us about the memory representations involved during retrieval, and if there is evidence that (perhaps fine-grained) processing penalties arise, or if there is evidence of real-time operations altering or repairing material.

Outline of the Dissertation

The structure of this thesis consists of three main parts. In the first chapter, I examine the processing of pronoun resolution in coordinate and subordinate structures, with the aim of investigating novel factors that may influence the selection of an antecedent. Across four experiments, I use coordinate and subordinate constructions in which both clauses have multiple referents but only one pronoun, and probe whether the relationship constructed in the first conjunct subsequently affects access of the antecedent for the pronoun in the second conjunct. I additionally examine effects of antecedent locality on antecedent preference and processing. In an offline judgment task, I find a robust influence of parallelism across conjuncts in coordinate structures, such that parallel antecedent-pronoun relationships are highly favored. In contrast, no effect of locality emerges in offline measures. This experiment is followed by an eye tracking while reading experiment, using a gender mismatch paradigm to assess whether a parallelism

effect arises during real time processing of the pronoun in the second conjunct. I find effects of both parallelism and locality on the processing of the second dependency, indicating that the relationship processed in the first conjunct is represented in memory in sufficient detail to impact subsequent retrieval processes. These results are mostly replicated (with the exclusion of a locality effect in subordinate constructions) when the clauses are in a subordinate type construction, indicating that this phenomenon is not specific to coordination but is more likely a general preference for parallelism. The findings from these studies point toward (1) a structure referencing memory representation of the pronoun-antecedent relationship and (2) the ability of this relationship to affect antecedent search.

Chapter 2 of this dissertation builds off of the indications of detailed structural information in the antecedent to investigate complexity effects in ellipsis constructions. In one offline and one online experiment I test whether it is possible to observe the effects of antecedent size on acceptability judgments as well as on eye tracking measures, and furthermore seek to replicate the observation of a sprouting-induced penalty. Both experiments elicit effects of antecedent size and sprouting, such that longer more complex antecedents result in degraded acceptability judgments and longer fixations, and sprouting also is both dispreferred and disrupts processing of the ellipsis site. The novel complexity effect I witness may provide some challenges for direct access models that predict no effect of antecedent size or complexity in ellipsis resolution.

Following up on the results of the first two experiments in Chapter 2, I take a deeper look into how structural constraints such as syntactic islands may interact with ellipsis and antecedent complexity effects. Making minimal changes to the stimuli used in the previous two experiments, I compare sluicing constructions with adjunct vs. argument extraction in the context

of factive islands, and furthermore manipulate the size of the potential antecedent. I again find a complexity effect, but observe that this effect appears to be modulated by whether the resulting interpretation of ellipsis based on the selected antecedent will violate a factive island constraint. These intriguing findings are discussed in light of cue-based retrieval and effects of parallelism on the scope an adjunct is able to take with factive verbs.

Having visited the search space and the content of retrieval, Chapter 3 moves to exploring phenomena that may occur as post retrieval operations. This chapter also focuses on ellipsis, however probes non-isomorphic indexical relations and morphology during the resolution of verb phrase ellipsis (VPE). The first two experiments of this chapter address the topic of strict and sloppy identity in ellipsis. I begin with an acceptability judgement which demonstrates a clear preference for the strict interpretation of VPE compared to the sloppy interpretation, indicating the preservation of the binding relation in the antecedent phrase. This issue is further pursued in an eye tracking while reading study which additionally manipulates the gender of the pronoun antecedent of the elided pronoun in sloppy interpretations to see if there is a cost associated with mismatching gender between the subject of the clause containing the VPE and the isomorphic form of the elided pronoun. The results of this study are in certain ways more complicated previous ones in this thesis and will be discussed in much greater detail in the chapter discussion, though in broad strokes I find much weaker evidence for facilitation resulting from the strict interpretation. I do, however, observe an interesting interaction that indicates that both morphological form and gender features are retrieved in these contexts.

The final experiments of this chapter also look at pronoun resolution and VPE, but in this case the area of interest is in instances of vehicle change. Using a complex combination of cataphora and VPE, I ask whether comprehension of sentences involving presumed vehicle

change are judged more harshly than those that do not, whether there is evidence of vehicle change occurring during online processing, and if that occurrence affects processing adversely. Although the results of the acceptability judgement portion of these experiments was unpredicted (most probably due to unrelated methodological issues), eye tracking results pointed to the computation of vehicle change during processing of the ellipsis, a process which was costly to the parser. These findings are discussed in light of assumptions about the types of features retrieved during ellipsis resolution, as well as how a successful and acceptable parse can be achieved.

Chapter 4 concludes.

Chapter 2: Parallelism Effects in Pronoun Resolution

Introduction

In this chapter, I pursue the topic of dependency resolution, in particular the resolution of the dependency between pronouns and their antecedents, to further probe what kinds of information are available to bias the parser when evaluating and selecting the antecedent. Through studies on complex sentence structures where two sentences are combined either by "and" or by "whereas", I ask whether the relation established in the first of the sentences affects the pronoun resolution process in the second sentence. More specifically, I ask the following two questions by examining the processing of sentences like (9).

(9) a. *Sean* said *Miles* loves <u>his</u> mother.

b. *Sean* said Jane loves <u>his</u> mother, and *Mabel* said *Ann* hates <u>her</u> father.

First, is the pronoun resolution process subject to locality effects? In other words, when the parser can find two candidate elements that may serve as the antecedent of the pronoun, does the parser always pick the closest (or the most local) element as the antecedent (Cunnings, Patterson, & Felser, 2014; Frazier, Ackerman, Baumann, Potter, & Yoshida, 2015; Kazanina et al., 2007; Van Gompel & Liversedge, 2003)? In (9a), for example, the pronoun *his* can find two possible candidates for the antecedent, namely *Sean* and *Miles*. In such a circumstance, does the parser pick *Miles* the closer NP as the antecedent of his or, *Sean* the far-away NP? If the parser prefers to have the antecedent of the pronoun in the local environment, then *Miles* should be selected as the antecedent in (9a).

Second, when there are two or more sentences connected in one sentence, is the parallelism between two sentences considered in terms of the selection of an antecedent? More concretely, when a pronoun 'finds' its antecedent in a particular structural position, is the selection of an antecedent for a subsequent pronoun impacted by this previous relationship? For example in (9b), in the first conjunct, for the pronoun *his* it is most natural to have *Sean*, a distant but gender matching Noun Phrase (NP) as the antecedent. Thus, in (9b) it is most likely that *his* forms a non-local referential dependency with *Sean*. The question is whether this non-local resolution of the referential dependency in the first conjunct affects the referential dependency formation in the second conjunct. If the parser prefers for the conjoined or subordinated clauses to hold a certain parallelism, and if the dependency structure is subject to this parallelism consideration, then in (9b), the parser picks the non-local NP *Mabel* as the antecedent of *her*. On the other hand, if the parser does not consider referential dependency relation in terms of the parallelism relation between the two connected clauses, and the parser prefers the local resolution of the referential dependencies, then the parser should take *Ann* as the antecedent of the pronoun.

To investigate these questions, I present a series of sentence completion and eye tracking experiments which modulate the position of potential antecedents to examine processing effects of locality and parallelism using sentences conjoined by "and", and extend this work to determine whether any potential effects can be attributed to general priming or coherence relation principles via the investigation of sentences connected by "whereas". To preview my

results, I find evidence for both locality and parallelism during pronoun resolution in both coordinate ("and") and subordinate ("whereas") structures[2].

Background

Pronoun resolution and locality effects

Pronouns have been extensively studied in the retrieval literature as a route to understanding the memory mechanisms that underlie language comprehension, and are sensitive to numerous factors which appear to guide the selection of an antecedent for successful resolution. Pronoun resolution is subject to various constraints, namely condition B of Binding Theory (Chomsky, 1981a), morphological and syntactic properties of the pronoun and its antecedent (Badecker & Straub, 2002), as well as a complex set of discourse constraints (Büring, 2005), including but likely not limited to (a) the prominence or salience of the potential antecedents (Clifton Jr & Ferreira, 1987; Gordon et al., 2001; Gordon, Hendrick, & Johnson, 2004; Grosz, Weinstein, & Joshi, 1995), (b) real world knowledge about the previously processed referents (Hobbs, 1979), and (c) discourse coherence relations (Chambers & Smyth, 1998; Elman, Kehler, & Rohde, 2006; Kehler, 2002; Kehler, Kertz, Rohde, & Elman, 2007; Kehler & Rohde, 2013; Kertz, Kehler, & Elman, 2006; Konieczny, 2005; Rohde, Kehler, & Elman, 2007; Streb, Hennighausen,

[2] When referring to the constructions I investigate in this study, I call sentences connected by "and" coordinate clauses, and sentences connected by "whereas" subordinate clauses. However, the distinction between coordination and subordination is both subtle and complicated. From a syntactic perspective, clauses connected by "and" and "whereas" show different distributions, but from a discourse-pragmatic perspective "whereas" exhibits the properties of coordination. I discuss this issue further in the background section.

& Rösler, 2004). Other preferences that have been proposed also refer to a bias to resolve pronouns to an antecedent in subject position (Crawley, Stevenson, & Kleinman, 1990) or to an antecedent that fulfils the same grammatical role as the pronoun (Chambers & Smyth, 1998; Smyth, 1994). To this diverse and sometimes contradictory array of observed patterns, we may suspect that locality in terms of dependency length also has a role to play.

Although human language allows for dependencies to be arbitrarily long (Chomsky, 1973, 1977), there appears to be evidence that in the absence of other mitigating factors the parser is motivated to resolve a dependency as soon as possible, resulting in a bias towards shorter dependencies (Aoshima, Phillips, & Weinberg, 2002, 2004; Gibson, 1998; Gibson & Warren, 2004; Phillips, 2006; Stowe, 1986; Traxler & Pickering, 1996). This locality bias has been most robustly demonstrated through the investigation of wh-dependencies, which has shown that in sentence comprehension speakers attempt to link a wh-fronted filler to the first potential gap site encountered in the downstream material (Stowe, 1986). Research employing acceptability judgement tasks (Hofmeister, Jaeger, Sag, Arnon, & Snider, 2007; Konieczny, 2000), production tasks (Konieczny, 2000), and ERP studies (Fiebach, Schlesewsky, & Friederici, 2002; Kluender & Kutas, 1993a) have also observed results that point to an effect of locality on dependency formation and resolution.

The locality bias is not limited to wh-dependencies, but also appears to influence cataphora resolution where a pronoun precedes the antecedent. In an eye tracking study, Van Gompel and Liversedge (2003) found evidence of an active search mechanism for antecedents of a pronoun during the processing of cataphora. In an experiment using biclausal sentences, a pronoun in the first clause either matched or mismatched the gender of the main subject of the second clause. An appropriate antecedent for the pronoun appeared in all sentences, however the

gender manipulation resulted in the appropriate antecedent either appearing in subject or object position of the main clause, as shown in (10).

(10)

a. gender match

When he was at the party, the boy cruelly teased the girl during the party games.

b. *gender mismatch*

When he was at the party, the girl cruelly teased the boy during the party games.

c. *control*

When I was at the party, the boy cruelly teased the girl during the party games.

A mismatch effect was found immediately following the main clause subject, at *cruelly*, such that first-pass reading times were slower when the gender of the main subject did not match the gender of the initial pronoun. The authors argue that this mismatch effect reflects the formation of a dependency with the first potential antecedent position prior to bottom-up information about the compatibility of the noun phrase occupying that position. Subsequent research (Kazanina et al., 2007) replicates this finding, and further shows that this search is constrained by grammatical factors, such that a gender mismatch effect is absent when the potential antecedent is in a structurally illicit position. These studies are also compatible with approaches such as Syntactic Prediction Locality Theory and Dependency Locality Theory (Gibson, 1998, 2000; Warren & Gibson, 2002), which have argued that the apparent eagerness of the parser to complete open dependencies is a result of the storage costs associated with keeping dependencies open during processing, and the cost of integrating two elements of a dependency.

Although the process of cataphoric pronoun resolution shares many similarities with filler-gap dependencies, inasmuch that a dependent element precedes a controlling element thus opening a search for an antecedent prior to bottom-up information about potential antecedents, the parser can only become aware of the presence of an anaphoric relation after encountering and recognizing the pronoun, which appears after the potential antecedents. Rather than evaluating candidates as they appear during comprehension, the parser must search through already-processed-material in memory for antecedent candidates.

There is, however, evidence that locality effects emerge in the realm of pronoun resolution in the more usual configuration as well, e.g. during a backwards search process. Some earlier research found that when pronouns were resolved to more distant antecedents, processing was delayed in some manner, reflected by comprehension latency (Clark & Sengul, 1979; Streb et al., 2004), eye tracking measures (Ehrlich & Rayner, 1983), and N400 effects (Streb et al., 2004). Interestingly, these results contradict with first mention accounts which posit that the first mentioned participant in a sentence holds a privileged position in regards to ambiguous pronoun resolution (Gernsbacher & Hargreaves, 1988; Gernsbacher, Hargreaves, & Beeman, 1989). It is worth noting that many of these earlier locality studies employed constructions where potential antecedents were significantly removed from the pronoun in question, often separated from the pronoun by one or more sentences. More recent research has found locality effects intra-sententially. For example, across two experiments Cunnings et al. (2014) manipulate the gender match between a quantified noun phrase, a proper name, and a pronoun, to investigate relative timing differences between variable binding and coreference assignment during pronoun resolution. Although their central question is whether a variable binder antecedent is initially preferred as opposed to a merely coreferential antecedent, they find that readers attempt to link a

pronoun to the linearly closest antecedent irrespective of its status. In (11a), the quantified noun phrase is linearly more distant from the pronoun (Experiment 1), whereas in (11b, Experiment 2) it is linearly closer. The quantified noun phrases used (e.g. 'every soldier') were assumed to be strongly associated with either male or female gender, and gender congruence between the pronoun and each potential antecedent was manipulated in a 2x2 design.

(11)

a. Every soldier who knew that *James/Helen* was watching was convinced that **he/she** should wave as the parade passed.
b. It looked to *James/Helen* that every soldier was completely convinced that **he/she** should wave as the parade passed.

The authors find that proximity of the potential antecedent to the pronoun affected processing of the pronoun, but that type of noun phrase had little to no effect. When the more local noun phrase mismatched in gender with the pronoun (e.g. *…Helen….he…*), longer reading times as reflected by eye tracking measures were observed at and after the pronoun, regardless of whether the local antecedent was a QP or a proper name, and regardless of whether the more distant noun phrase matched in gender.

This body of research seems to indicate that locality has a role to play during the pronoun resolution process. This is not to say that other proposed heuristics are incorrect, but rather that the pronoun resolution process is complex and some heuristics may be more valuable or 'weighty' to the parser. For example, the above studies demonstrate a locality bias, but do not rule out a competing bias for first mention.

Early proposals of the antecedent search process suggested that this search may have occurred in a backwards fashion (Ehrlich, 1980; Gordon & Hendrick, 1998; McElree & Dosher, 1989; McElree et al., 2003) linearly traversing and evaluating the previously processed material, which presumably could have demonstrated the same kind of eagerness for resolution that forwards search does, resulting in locality effects such that more recently processed antecedents would be favoured over more distant ones. However, as dependency resolution has been more widely investigated, the common consensus has been reached that the parser uses a cue-based retrieval system that evaluates candidates in parallel, which among other things is supported by findings of interference effects from more distally located antecedents (Franck, Colonna, & Rizzi, 2015; Franck, Vigliocco, & Nicol, 2002; Wagers et al., 2009). However, the effect of distance can alternatively be recognized as a matter of memory decay, and as such the distance between the head and tail of a dependency can still be an important factor in the dependency resolution process. Popular models of the parsing architecture, for example the ACT-R (Adaptive Control of Thought - Rational; (Anderson, 1996)) model adapted by Lewis and Vasishth (2005) also assume an element of decay, such that more recent material (e.g. local material) has a higher level of activation than non-local material, affecting both retrieval latency and comprehension accuracy.

Parallelism

Studies have shown that various types of parallelism of conjuncts in coordinate structures is helpful to the parser, such that a second conjunct is easier to parse if it is parallel to the first conjunct in some way (Altmann, Henstra, & Garnham, 1993; Frazier, Munn, & Clifton, 2000; Frazier, Taft, Roeper, Clifton, & Ehrlich, 1984). This phenomenon was notably observed by Frazier et al. (1984) who tested a series of conjoined sentences in which the conjuncts were

constructed such that their contents matched or did not match. For example, when animacy of the object DP was parallel across conjuncts (12a) the second conjunct was read faster than when animacy was not parallel (12b).

(12)

a. *John telephoned the doctor and his friend telephoned the lawyer*

b. *John telephoned the doctor and his friend telephoned the museum*.

Parallelism effects were also found in constructions manipulating active and passive voice, sentential vs. DP objects, non-shifted vs. shifted heavy NPs, and thematic role. Later research also showed a preference for parallelism by comparing constructions that were or were not structurally parallel (Frazier et al., 2000). The underlined phrase of the second conjunct was read faster in (13a) than in (13b), which is attributed to the parallel internal structure of the conjuncts of (13a).

(13)

a. *Hilda noticed a strange man and <u>a tall woman</u> when she entered the house.*

b. *Hilda noticed a man and <u>a tall woman</u> when she entered the house.*

Subsequent research has shown evidence for parallelism effects in a diverse array of constructions, including different kinds of elliptical constructions (Arregui, Clifton Jr, Frazier, & Moulton, 2006; Dickey & Bunger, 2011; Fox, 2000; Kehler, 2000; Mauner, Tanenhaus, & Carlson, 1995), and does not appear to be language specific (Apel, Knoeferle, & Crocker, 2007;

Knoeferle, 2007; Knoeferle & Crocker, 2009). However, little research has investigated the impact of parallelism on long distance dependencies.

Recently, Parker (2017) examined the processing of wh-filler-gap dependencies in the context of coordinate structures. In the coordination context, it has been observed that if a wh-element is extracted from a conjunct, it must be extracted from all the conjuncts as in (14), known as the across-the-board (ATB) extraction restriction (Gazdar, 1981; Ross, 1967; Williams, 1978).

(14) The beer which the friends were enthusiastically selling __ and drinking __ was brewed in California.

Parker compared conditions in which the position of the gap site was either parallel or non-parallel in the second conjunct by alternating the category of the filler (PP vs. NP) and parallelism in sentences with an NP filler, as shown in (15).

(15)

PP +parallel

a. *The harsh chemicals **with which** the technician sprayed the sensitive equipment ___ and prepared <u>the sterile beakers</u> ___ were manufactured in China*

NP +parallel

b. *The harsh chemicals **which** the technician sprayed the sensitive equipment with ___and prepared <u>the sterile beakers</u> with ___ were manufactured in China*

NP -parallel

c. *The harsh chemicals **which** the technician sprayed ___ and prepared <u>the sterile beakers with</u> ___ were manufactured in China*

In the parallel conditions there is a late gap in the prepositional phrase in both conjuncts, however in the non-parallel condition there is an early object gap in the first conjunct and a later prepositional gap in the second conjunct. If parallelism impacts the expectations of the gap position in the second conjunct, then the parser should be surprised at the presence of an overt direct object in the second conjunct of the non-parallel condition, resulting in a 'filled gap effect' (Stowe, 1986). In a self-paced reading experiment Parker (2017) found a significant disruption at the first two words of the filled gap position, such that 'the sterile' was read significantly slower in the non-parallel condition as compared to the parallel conditions, indicating that parallelism had a considerable effect on the construction and expectations of multiple gap dependencies. This work was of particular importance to the experiments detailed in this paper as it provides evidence that dependency formation and processing is among the greater set of elements over which parallelism may operate.

Although there have been well documented observations of robust parallelism effects at various levels of processing, the source of these effects remains unclear. One appealing approach to this problem is to attribute the source of parallelism effects to the source of more general structural priming. Due to its remarkable similarity to observations of the facilitative effects of structural priming (Arai, Van Gompel, & Scheepers, 2007; Branigan et al., 2005; Scheepers & Crocker, 2004; Traxler, 2008), it is reasonable and attractive to posit that parallelism effects are simply priming effects in a coordinate environment. Indeed, there seems to be some evidence that certain instances of parallelism can be attributed to general priming considerations. Using

corpus-based modelling, Dubey, Sturt, and Keller (2005) and Dubey, Keller, and Sturt (2008) found evidence for parallelism effects for noun phrases in both coordination and arbitrary syntactic environments, although they note that the effect was strongest in coordination. Building off this work, Sturt, Keller, and Dubey (2010) investigated the processing of noun phrases in coordinate and non-coordinate structures as in (16) using eye tracking, and found statistically equally sized parallelism effects in both coordinate and subordinate contexts.

(16)

(a) A demanding boss and a lazy worker did not do the job properly.

(b) A demanding boss said that a lazy worker did not do the job properly.

In contrast to these findings, other research has appeared to show that parallelism effects - unlike priming - are constrained by specific linguistic environments (e.g., coordination). Returning to earlier work by Frazier et al. (2000, see (13)), the authors did not witness facilitative effects of parallelism when the noun phrases served as subject and object as opposed to conjuncts (e.g. *a strange man noticed a tall woman*), which was further corroborated by similar experiments performed in German which failed to observe a parallelism effect outside the coordination context (Apel et al., 2007). These results are supported by an examination of parallelism of constituent order in German (Knoeferle, 2014), shown below in (17) , which across three experiments did not find significant parallelism effects when the conjuncts were conjoined by “but” or “while” as opposed to “and”.

(17)

a. **'and'/ 'but' / 'while' parallel**

*Der Polizist berichtete, dass den Trinker (*obj*)$_{NP1}$ der Rüpel (*subj*)$_{NP2}$ verfluchte* ***und / aber/ während*** *den Säufer (*obj*)$_{NP3}$ der Rowdy* (subj)$_{NP4}$ erschreckte, was niemanden weiter überraschte.

'The policeman reported that the drunkard (obj)$_{NP1}$ the lout (subj) $_{NP2}$ cursed **and / but / while** the tippler (obj)$_{NP3}$ the thug (subj)$_{NP4}$ frightened, which didn't surprise anybody.'

b. **'and' / 'but'/ 'while' non-parallel**

Der Polizist berichtete, dass der Trinker (subj)$_{NP1}$ den Rüpel (obj)$_{NP2}$ verfluchte **und /aber / *während*** den Säufer (obj)$_{NP3}$ der Rowdy (subj)$_{NP4}$ erschreckte, was niemanden weiter überraschte.

'The policeman reported that the drunkard (subj)$_{NP1}$ the lout (obj)$_{NP2}$ cursed **and / but /while** the tippler (obj)$_{NP3}$ the thug (subj)$_{NP4}$ frightened, which didn't surprise anybody.'

In Experiment 2, "and" and "but" were contrasted in parallel and non-parallel constituent order conditions, and in Experiment 3 "and" and "while" were compared. The results from eye tracking revealed a significant parallelism effect when clauses were conjoined by "and" in both Experiment 2 and Experiment 3, but not when the connective was "but" or "while". The author contends that the meaning associated with the connective, e.g., "and" for conjoining, "but" for contrasting, modulated the parallelism effect, which is unexpected under a general priming account.

In the domain of pronoun resolution, Kehler et al. (2007) investigated what factors influence pronoun resolution, and identified that pronoun resolution is subject to certain parallelism effects and that these effects on pronoun resolution follow from coherence relations. Past studies on pronoun interpretation have pointed out that NPs serving as certain grammatical functions and NPs located in certain positions are preferred as the antecedent of a pronoun. For example, it has been observed that the NP serving the same grammatical role as that of a pronoun is preferred to be the antecedent for the pronoun (Chambers & Smyth, 1998; Sheldon, 1974; Smyth, 1994). In (18a), the preferred antecedent for *him* is *Kerry*, which is an object. This preference is understood to reference the fact that the pronoun is serving as the grammatical object in the second conjunct, and so therefore the NP that is serving the same grammatical role (the grammatical object) is preferred to be the antecedent.

(18)

a. Bush narrowly defeated Kerry, and Romney trounced him.
b. Bush narrowly defeated Kerry, and Romney congratulated him.

Grammatical role parallelism thus shows that pronoun resolution in a complex sentence may be subject to a parallelism effect. In (18b), however, the most reasonable interpretation of *him* seems to be *Bush*, which does not serve the same grammatical role as the pronoun.

To accommodate these conflicting observations of antecedent preference, Kehler et al. argue that a Grammatical Role Parallelism bias (for 18a) and a Subject Preference (for 18b) follow from the more general discourse coherence relations such as resemblance relations or cause and effect relations. Most relevant to my research, a resemblance relation may hold

between two elements if the events described are similar in nature or if they contrast, and in such a relation parallelism between the elements (as outlined in (19)) is strongly preferred. That said, what precisely is meant by 'similar' or 'contrasting' events is not entirely clear - does it refer to related events that may happen (e.g. 'reading' and 'writing'), or would a difference in events such as 'lifting a box' vs. 'pondering a question' count as a contrast?

(19) **Parallel**

Infer $P(a_1, a_2, ...)$ from the assertion of S_1, and $P(b_1, b_2, ...)$ from the assertion of S_2, for a common P and similar *ai* and *bi*.

In (18a), *Bush* and *Romney* are similar elements both serving as subjects, and there is a resemblance relation between the two clauses. This means that the NPs that serve as the object *Kerry* and *him* are also similar elements. This parallel relation, but not grammatical role parallelism, is the source for the preference for the grammatical object *Kerry* to be the antecedent of the pronoun.

In the studies I have reviewed, an explicit distinction of the relevant senses of *coordination* and *subordination* is not always present, and their status in syntax and discourse is at times unclear. In general, coordination and subordination have been defined as syntactic concepts, which are subject to syntactic tests targeting their structural status. Structurally, we can observe that coordination involves two elements of seemingly equal status, while in subordination one element is a head and the other is a dependent (Huddleston & Pullum, 2005). Although the exact structural representation of coordination has been debated for some time, certain tests unequivocally distinguish it from subordination. For example, subordination

appears to allow for fronting of the subordinated clause, whereas coordination resists such permutation as shown in (20a-b). Additionally, coordination is known to allow across-the-board (ATB) extraction although subordination does not (20c-d). Other tests include the availability of constructions such as gapping in coordination and parasitic gaps in non-coordinate structures.

(20)

a. **Whereas** Bill ate cheese, John drank wine ___
b. * **And** Bill ate cheese, John drank wine ___
c. What did Bill buy __ **and** John drink ___?
d. * What did Bill buy ___ **whereas** John drink ___?

While most studies focused on parallelism effects in coordination have relied on purely syntactic notions of coordinate vs. non-coordinate structures, the mapping between these structural concepts and discourse concepts is not straightforward, and various asymmetries surface. Furthermore, the criteria and classification of coherence relations remains under debate (Jasinskaja & Karagjosova, 2015; Knott & Sanders, 1998). Some criteria have been proposed to classify coordinating and subordinating discourse relations (Asher & Vieu, 2005), although the relations appear to be 'overridable' in specific contexts. However, the relevant point is that certain types of connectives may create structurally subordinate environments, but also hold the adjacent discourse units in a resemblance relation. It is worth noting that the *effect* of something like a resemblance relation is in essence indistinguishable from a simple priming account, with

the caveat that priming effects should only be observed when the discourse encourages it. What types of elements or connectives encourage resemblance is a bit murkier.

As Kehler (2002) concentrates on a structurally coordinate context when discussing parallelism, it is not clear whether effects of the resemblance relation are seen only in the coordination context or if it can be extended to the subordination context, although given that some subordinators induce a contrast between the events they link, I surmise that this is the case. However, as parallelism effects in the processing literature have been seen in structurally defined coordination but not always in subordination, whether an effect will surface when structure and coherence are at odds is at issue. For example, in the first set of experiments in this paper, structure (coordination) and the coherence relation (resemblance) are in tandem in the sense they both potentially encourage parallelism, however in the second set of experiments structure (subordination) and the relation (resemblance) are pitted against each other in the sense that structurally subordinate environments are not predicted to encourage parallelism, but the resemblance relation is.

To summarize, the present study seeks to establish whether pronoun–antecedent dependency formation is subject to parallelism effects, and to determine whether any such effects are attributable to general priming or coherence effects or are specific to a coordination context. Additionally, this study employs constructions that also allow an investigation of locality effects in the process of antecedent retrieval. One possibility is that during backwards search for an antecedent of a pronoun, a previously processed pronoun–antecedent dependency does not impact selection of an antecedent. While a lack of evidence for parallelism operating over the processing of pronoun dependencies does not unequivocally rule out the possibility of pronoun–antecedent relations having a representation that is stored in short term memory, it would

minimally indicate that such a representation is impervious to parallelism. Alternatively, an observation of parallelism effects would confirm both that a pronoun–antecedent relation is represented in memory, and that this representation is subject to parallelism. Locality of an antecedent is expected to surface independently of parallelism, since there is substantial evidence of the impact of memory decay of previously processed material. Note that I do not have predictions concerning the relative strength of these factors should they surface, i.e. it may be that parallelism merely attenuates the effect of locality, or vice versa. Given the observation of a parallelism effect, this research will examine whether facilitation due to parallel conjuncts is a phenomenon specific to a structurally defined context (coordination), or it can be accounted for by a general priming or coherence driven account.

This Study

The structure of this research is as follows: Each experiment consists of two parts, a forced choice paradigm and an eye tracking study, which allows for an investigation of both offline and online comprehension. Experiment 1 uses conjoined constructions containing pronouns which unambiguously select an antecedent in the first conjunct, and determines whether the length of the pronoun–antecedent dependency in the first conjunct affects the selection or processing of the dependency in the second conjunct. To foreshadow the results, an observation of both locality and parallelism effects are reported. Experiment 2 addresses the source of parallelism effects by using identical stimuli as in Experiment 1, with the exception that in all conditions the connective introduces subordination in contrast to coordination by use of “whereas” as opposed to “and”, and provides evidence that parallelism effects are not limited to coordination.

Experiment 1

Experiment 1a

Experiment 1a serves to test whether participants are sensitive to the parallelism of anaphoric dependency resolution in an offline forced-choice task. Specifically, this task tested whether participants were more willing to construct a local or long-distance pronoun–antecedent dependency based on the type of dependency they processed in the first conjunct of a coordinate structure. I predict that parallelism will bias the construction of dependencies in the second conjunct, such that subjects will be significantly more likely to choose a pronoun referring to the matrix antecedent in the matrix dependency condition, and a pronoun referring to the local antecedent in the local dependency condition. A table detailing the predictions that different frameworks make is provided in Table 1, although given the types of connectives used in both this experiment and Experiment 2a the predictions of discourse coherence accounts are indistinguishable from simple priming accounts. . Given the offline nature of this task, a strong locality effect is not predicted.

1st Conjunct Dependency	**Factor**			
	Locality	*Priming*	*Parallelism*	*Discourse Coherence*
Baseline	local	NA	NA	NA
Matrix	local	matrix	matrix	matrix
Local	local	local	local	local

Table 1:Predictions made by different frameworks for Experiment 1a. Cell contents refer to the pronoun–antecedent relation made by the chosen pronoun.

Methods and Materials

Participants

Participants were 48 speakers recruited on Amazon Mechanical Turk. Participants were self-reported native English speakers, and had IP addresses within the United States. Subjects were compensated $2.00 for their participation. All subjects gave informed consent under approval of the Northwestern University Institutional Review Board, and were run under the protocol *Clausal Ellipsis: Its Structure and Online Processing (STU00082465)*.

Materials

A set of 36 target items and 36 fillers was created, and this set was pseudo-randomised per participant, resulting in 1728 critical observations. Target items were presented according to a standard Latin square design, and no more than 2 target or 2 filler sentences appeared in a row. Filler items consisted of a short sentence using the sluicing elliptical construction, with the choice of either 'why' or 'what' to complete the sentence (e.g. *Mary, for some reason, bought clothes, but I don't know* ___ (why/what)). Target items consisted of three different conditions: baseline, matrix dependency, or local dependency. An example of the three conditions is shown below in (21):

(21)

BASELINE

a. Mary said Tom hates ___ father.

MATRIX DEPENDENCY

b. Bill said Jane loves his mother, and Mary said Tom hates ___ father.

c. Bill said Jane loves her mother, and Mary said Tom hates ___ father.

Target items were constructed so that there were always two available antecedents of opposite gender. These possible antecedents were always in subject position, either as the subject of the matrix clause or the subject of the embedded clause. This design was used so that a participant's choice of *his* or *her* would unequivocally pick out either the matrix subject or the embedded subject as an antecedent. The baseline condition was always a simple non-coordinate structure, which corresponds to the second conjunct from the matrix and local dependency conditions. Stimuli was gender-balanced such that in half of the items the matrix subject was male and the embedded subject was female, and vice-versa.

Stimuli were also designed such that for all items (conditions b and c) the choice of a pronoun that created a parallel dependency in the second conjunct was a different lexical item than the pronoun in the first conjunct. For example, if there was a matrix dependency with a male antecedent in the first clause, then the dependency choice that would create a parallel relationship in the second conjunct would have a female antecedent. More concretely, the pronoun *his* appears in the first conjunct of (21b), and so the parallel pronoun choice in the second conjunct would be *her*. This design choice was implemented to control for the possibility of lexical priming. In this experiment, as in the following experiments, the discourse coherence relation is one of resemblance, which also predicts parallelism of conjuncts.

Procedure

Participants performed this experiment on Amazon Mechanical Turk. Participants were

instructed that they would see 72 English sentences, and that each sentence would contain a gap. Subjects were told that they should fill each gap by choosing one of the two options shown below the sentence. The instructions stated that subjects should choose the option that sounded most natural to them.

Prior to the experimental items, participants were given a practice trial that had an unambiguous answer, as well as a practice trial with two grammatical options. These trials were syntactically and semantically unrelated to the experimental trials. All practice and experimental items were shown to the participants on one page, and previously completed items were not masked from view.

Analysis and Results

Results were analysed in a single linear mixed effects logistic regression model using random subject and item intercepts[3], using the lme4 package in R. Results revealed that the experimental conditions highly impacted the rate at which different dependencies were constructed. Data was treatment coded, with condition a (baseline) set as the reference level. The dependent variable was a binary variable indexing whether subjects chose a local antecedent, coded as '1', or a matrix antecedent, coded as '0'. The choice of a local or matrix antecedent in the baseline condition was at chance[4], as shown by the lack of significance of the baseline condition in the

[3] glmer (choice ~ 1 + condition + (1 | subject) + (1 | item), data = data, family = "binomial")

[4] I note here that this result is a good indication that *overall* the embedded verbs in the baseline condition did not prefer a local or matrix antecedent, however I do observe some biases on an item-by-item basis, and embedded verb bias was not strictly controlled in my materials. For example, in the baseline condition of one item ('George said Lily mesmerizes ___ clients'), the local choice 'her' was selected 63% of the time. However, when this sentence was preceded by a conjunct with a matrix pronoun–antecedent relation, selection of the local choice dropped to 50%. Conversely, when

logistic regression model, however the matrix condition strongly predicted a matrix pronoun choice, and the local condition strongly predicted a local pronoun choice. Table 2 gives counts and percentages of local and matrix antecedent choices, and estimates, standard error, z-values and their associated p-values from the regression model are reported in Table 3.

	Matrix dependency pronoun	**Local dependency pronoun**
Baseline	278 **48%**	297 **52%**
1st Conjunct Matrix Dependency	416 **72%**	160 **28%**
1st Conjunct Local Dependency	153 **26%**	423 **74%**

Table 2:Raw counts and percentages for pronoun–antecedent dependency choice by condition.

condition	**estimate**	**standard error**	**z-value**	**p-value**
baseline	-.07	.16	-.44	.66
matrix	1.17	.14	8.66	<.001
local	-1.11	.14	-8.13	<.001

Table 3:Estimates, standard error, z-value, and p-value from a single logistic regression model for Experiment 1a.

the preceding conjunct had a local pronoun–antecedent relation, the percentage rose to 75%. Thus, I appear to see parallelism ‘overriding’ or ‘adding to’ any biases present in the materials.

Discussion

There are two main findings from this analysis. First, as predicted, the presence of a dependency in the first conjunct strongly influenced the kind of dependency constructed in the second conjunct. This finding supports the claim that parallelism operates over pronoun resolution, such that participants were strongly biased to create parallel dependencies. Second, an effect of locality either in the baseline or critical conditions was not found. Although it is possible that parallelism effects could have masked a locality effect in the critical conditions, the lack of a locality effect was unsurprising. As the source of locality effects is suspected to be rooted in the decay of recently processed items in memory, there is not a compelling reason to expect effects to surface when the task is offline in nature and allows for reflection and rereading of the material.

Experiment 1b

Experiment 1b employs the manipulations of Experiment 1a in an eye tracking study aimed to gauge whether parallelism or locality guide pronoun resolution during online sentence processing. While the parallelism effects observed in an offline task are notable, it may not be indicative of the kinds of factors that impact retrieval during online language comprehension. The inclusion of an online task allows us to investigate whether parallelism quickly, directly, and unconsciously impacts the retrieval of an antecedent in real-time processing when a comprehender is unable to reflect upon the material. I predict that parallelism effects in online processing will surface in this experiment as facilitation at or directly after the pronoun in the second conjunct when the two dependencies are parallel, as indexed by the observation of faster eye tracking measures (e.g. shorter duration of relevant measures). I additionally predict that

locality will have a more marked impact during online processing, such that pronouns that resolve to the more local antecedent will also demonstrate facilitation compared to those that resolve to the more distant one. A table of predictions made by different factors is presented in Table 4.

Condition	**Factor**			
	Locality	*Priming*	*Parallelism*	*Discourse Coherence*
Local/Parallel	faster	faster	faster	faster
Matrix/Parallel	slower	faster	faster	faster
Matrix/Non-Parallel	slower	slower	slower	slower
Local/Non-Parallel	faster	slower	slower	slower

Table 4:Predictions for eye tracking measures in Experiment 1b. Cell contents indicate the predicted effect of that factor on eye tracking measures at or directly after the pronoun in the second conjunct.

Methods and Materials

Participants

Forty participants from the Northwestern University undergraduate community were recruited for this experiment. All participants were native English speakers, and all had normal or corrected-to-normal vision. Subjects received course credit for their participation. All subjects for this and subsequent experiments gave informed consent under approval of the Northwestern University Institutional Review Board, and were run under the protocol The Retrieval and Interpretation of Antecedent Material (STU00206750). Two subjects were subsequently excluded prior to analysis due to low comprehension question accuracy (<65%).

Materials

Materials consisted of 24 sentences like (22), in which the length of a pronoun–antecedent dependency was manipulated in both the first and second conjuncts by changing the gender of the matrix and embedded subject noun-phrases, resulting in two Parallel conditions and two Non-parallel conditions. Names were chosen which were assumed to have a strong association with a particular gender, and names were checked post-hoc against the materials of other investigations of GMME, and were also subjected to an informal norming study, revealing that names were either strongly female or strongly male. The pronoun in the second conjunct unequivocally resolves to a noun-phrase either in matrix or embedded position, resulting in two Local conditions and two Matrix conditions. Each experimental sentence was followed with a comprehension question which targeted different areas of the sentence to ensure thorough comprehension of the entire item.

(22)

Local/Parallel

a. Michael thought Emma missed **her** dog and Sarah thought Max despised **his** cat with a great hatred

Matrix/Parallel

b. Emma thought Michael missed **her** dog and Max thought Sarah despised **his** cat with a great hatred

Matrix/Non-parallel

c. Michael thought Emma missed **her** dog and Max thought Sarah despised **his** cat with a great hatred

Local/Non-parallel

d. Emma thought Michael missed **her** dog and Sarah thought Max despised **his** cat with a great hatred

The materials used in the eye tracking experiment are a subset of the items in the forced choice experiment, and are marginally different than those used in Experiment 1a, the most important change being that the positions of the antecedents across conditions are manipulated instead of the genders of the pronouns. This change was implemented to (1) maintain lexical consistency at the critical region of analysis, and (2) to avoid any effect of lexical priming of the same pronoun across conjuncts. Furthermore, 3 embedded verbs were changed to avoid any bias due to discourse coherence. Finally, a prepositional phrase (e.g. ... *with a great hatred.)* was added to allow for the observation of any spillover effects. 104 unrelated fillers were included in the experimental materials. A full set of experimental items are included in the Appendix.

Procedure

Participants gaze was recorded using a tower-mounted SR Research Eyelink 1000 Plus at a sampling rate of 2000 Hz. Fixations shorter than 40 ms were incorporated into adjacent fixations. Each experimental session began with four practice trials, two of which had comprehension questions. Re-calibration was performed between trials if necessary, and subjects were given short breaks as needed. The experimental sentences were presented according to a standard Latin Square, and were randomised. All sentences appeared on one line, in Monaco 14-point font. Participants answered a yes/no comprehension question after every trial.

Analysis and Results

For the purposes of this experiment, I will concentrate on eye tracking measures at three regions: the critical region containing the pronoun in the second conjunct (*his cat),* the spillover region containing the preposition (*with*), and the second spillover region containing a determiner and an adjective (*a great),* as shown in (23). No significant effects were found in any subsequent region. The following analysis is based on four eye tracking measures: first fixation duration, first pass duration, regression path duration, and total time duration. First fixation duration refers to the length of the gaze for the first time a region is fixated upon. First pass duration includes the duration of all fixations to a region prior to gaze exiting to either the left or the right of the region. Regression path duration is calculated by summing all fixations within a region and all fixations to areas left of a region prior to exiting the region to the right for the first time. Total time is calculated as a sum of all time spent in a region, including first pass duration and any time spent re-reading a region.

(23) Michael thought Emma missed her dog and Sarah thought Max despised [**his cat**]$_1$ [**with**]$_2$ [**a great**]$_3$ hatred

Statistical analyses were carried out on log-transformed durations for each region and measure using linear mixed effects regression (Baayen et al., 2008) in R using the lme4 package. Intercepts were allowed to vary across subjects and items, and slopes of the fixed effects and their interaction were allowed to vary across subjects and items[5]. Factors were treatment coded, and analyses were conducted by comparing a converging maximally inclusive linear mixed effects regression model (LMER) to a reduced model, in line with Barr et al. (2013). Table 5 reports the means and standard error in milliseconds of reading times. $\chi 2$-values and their

[5] For example: lmer (logTime ~ (1 + dep_length * parallelism |subject) + (1 + dep_length * parallelism | item) + dep_length * parallelism, data = data)

corresponding p-values accompanied by the estimates and standard error calculated from the maximal model are reported in Table 6, with significant or near significant effects indicated in bold.

Dependency Length	Parallelism	Critical region	Spillover 1	Spillover 2
FIRST FIXATION				
local	parallel	274 (9)	233 (14)	264 (9)
matrix	parallel	305 (12)	214 (11)	244 (9)
matrix	non-parallel	282 (9)	234 (11)	259 (11)
local	non-parallel	274 (10)	237 (10)	263 (14)
FIRST PASS				
local	parallel	361 (13)	236 (14)	310 (14)
matrix	parallel	365 (15)	214 (11)	286 (14)
matrix	non-parallel	365 (14)	236 (11)	307 (17)
local	non-parallel	374 (17)	237 (10)	311 (18)
REGRESSION PATH				
local	parallel	442 (25)	302 (32)	942 (128)
matrix	parallel	510 (39)	296 (39)	997 (147)
matrix	non-parallel	539 (37)	440 (73)	860 (110)
local	non-parallel	492 (33)	431 (68)	791 (109)
TOTAL TIME				
local	parallel	574 (25)	290 (17)	417 (21)
matrix	parallel	654 (32)	277 (16)	459 (27)

matrix	non-parallel	703 (35)	320 (21)	465 (27)
local	non-parallel	646 (32)	303 (19)	432 (23)

Table 5:Means and standard error in raw values for each fixation measure and region for Experiment 1b.

Region	Effect	Estimate	Std error	χ^2 (df)	p-value
First Fixation					
Critical	**dependency length**	**.07**	**.03**	**5.11 (1)**	**.024**
	parallelism	.036	.03	1.36 (1)	>.1
	interaction	.034	.059	.34 (1)	>.1
Spillover 1	dependency length	-.05	.044	1.3(1)	>.1
	parallelism	-.036	.04	.81 (1)	>.1
	interaction	-.036	.09	.16 (1)	>.1
Spillover 2	dependency length	-.04	.04	.96 (1)	>.1
	parallelism	.003	.05	.005 (1)	>.1
	interaction	-.07	.09	.75 (1)	>.1
First Pass					
Critical	dependency length	.006	.037	.029 (1)	>.1
	parallelism	-.002	.038	.005 (1)	>.1
	interaction	-.018	.07	.05 (1)	>.1
Spillover 1	dependency length	-.052	.045	1.33 (1)	>.1

	parallelism	-.033	.042	.68 (1)	>.1
	interaction	-.06	.088	.454 (1)	>.1
Spillover 2	dependency length	-.05	.05	1 (1)	>.1
	parallelism	.002	.06	.001 (1)	>.1
	interaction	-.03	.08	.12 (1)	>.1
Regression Path					
Critical	dependency length	.072	.044	2.6 (1)	>.1
	parallelism	-.046	.051	.806 (1)	>.1
	interaction	.014	.109	.02 (1)	>.1
Spillover 1	dependency length	-.042	.076	.308 (1)	>.1
	parallelism	**-.21**	**.083**	**5.67 (1)**	**.017**
	interaction	-.023	.149	.025 (1)	>.1
Spillover 2	dependency length	.1	.11	.8 (1)	>.1
	parallelism	.06	.11	.32 (1)	>.1
	interaction	-.07	.2	.12 (1)	>.1
Total Time					
Critical	**dependency length**	**.108**	**.042**	**6.61 (1)**	**.011**
	parallelism	**-.08**	**.043**	**3.41 (1)**	**.065**
	interaction	.028	.082	.114 (1)	>.1
Spillover 1	dependency length	-.019	.05	.159 (1)	>.1
	parallelism	-.038	.066	.344 (1)	>.1

	interaction	-.025	.11	.052 (1)	<.1
Spillover 2	dependency length	.06	.05	1.55 (1)	
	parallelism	.03	.05	.34 (1)	>.1
	interaction	.04	.11	.12 (1)	>.1

Table 6:Estimates, standard error, χ^2 values, and p-values for each fixation measure and region in Experiment 1b.

At the critical region, a significant main effect of locality was observed in first fixation time and total fixation time, such that fixation times for conditions with a local (short) dependency in the second conjunct were faster than those for conditions with a matrix (long) dependency. Additionally, there was a non-significant trend in total reading times for a main effect of parallelism ($p = .065$), such that total reading times tended to be shorter in conditions where the dependencies were parallel across the two conjuncts. These results indicate that retrieval of a more local antecedent is easier for the parser, and suggests that parallelism of the pronoun–antecedent relationships also facilitates processing of the second pronoun.

A main effect of parallelism emerged in the first spillover region, such that conditions with parallel dependencies elicited shorter regression path durations than those with non-parallel dependencies. No other effects reached significance. Bar plots of the dependent measure for statistically significant region and measure combinations are shown in Figure 2.

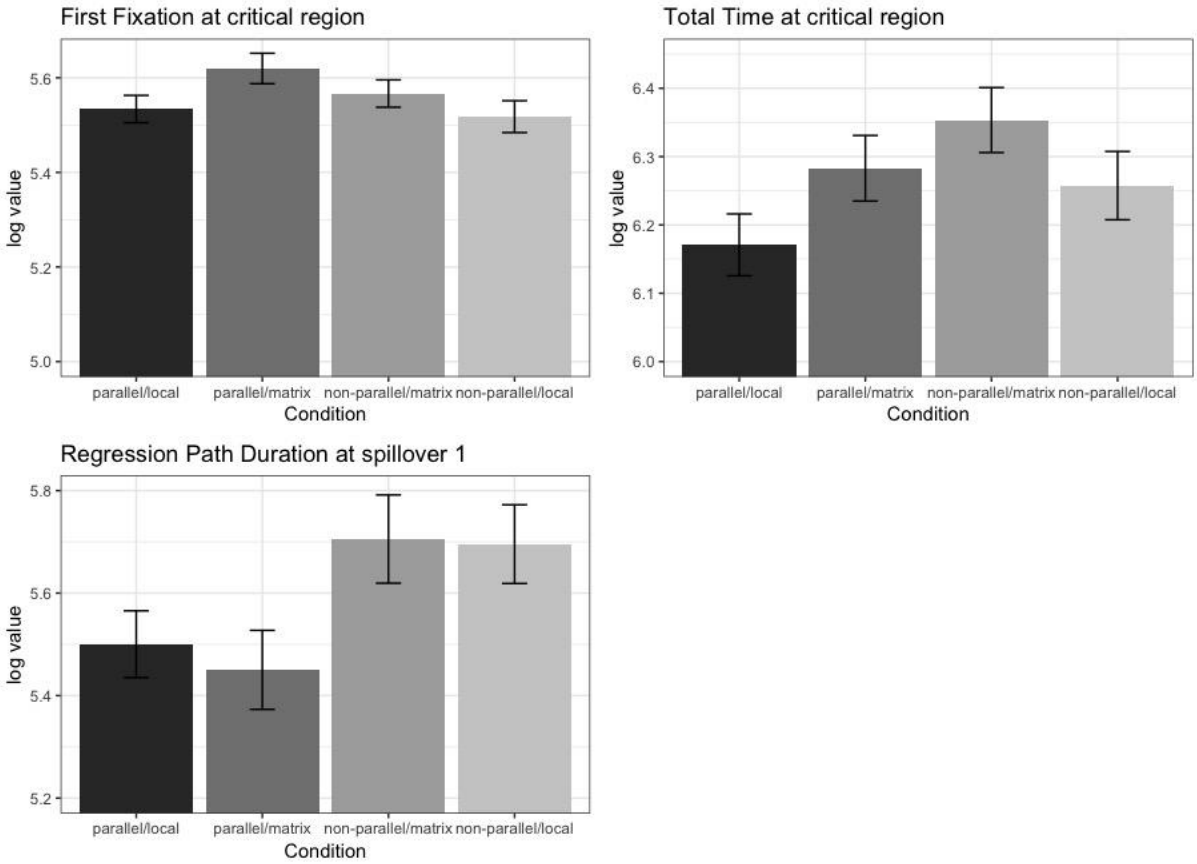


Figure 2: Bar plots for first fixation duration and total time at the critical region, and regression path duration at the first spillover region. Condition is on the x-axis and log fixation time is on the y-axis. Error bars represent standard error.

Discussion

The results from this eye tracking study provide strong indications of a parallelism effect in online processing. The implications of this and the previous study are discussed in more detail in the discussion section for Experiment 1 as a whole.

Experiment 1 Discussion

The results from Experiments 1a and 1b indicate that two factors appear to be influencing the ease of access of potential antecedents in the second conjunct of a coordinated structure: (1) how distant the antecedent is from the dependent element (locality), and (2) whether the antecedent resides in a parallel position to that of an antecedent of a previously processed dependency

(parallelism). While an effect of locality was not observable in the offline forced choice Experiment (1a), a clear effect of locality surfaced at the critical region in the online eye tracking experiment in both first fixation and total time. This effect is straightforwardly explained based on the relative activation levels of antecedents stored in memory due to how distantly in time they were processed, i.e. due to decay more recently processed items have a higher baseline activation than less recently processed items, which in turn may affect the retrieval latency of a given item.

The second main finding from these experiments is the apparent effect of parallelism on pronoun dependency formation in both offline and online measures. In the offline forced choice experiment, the length of the dependency processed in the first conjunct, e.g. whether the pronoun used picked out the matrix subject vs. embedded subject, dramatically affected the choice of pronoun in the second conjunct such that parallel dependencies were far more likely to occur than would be predicted by chance. Furthermore, a clear effect of parallelism also was observed during eye tracking at the first spillover region, as reflected by an increase in regression path duration for non-parallel conditions, as well as a tendency towards an increase in total time. These results strongly point to parallelism operating over pronoun dependency relations.

Experiment 2

Experiments 2a and 2b were created to test the source of the parallelism effects I witnessed in Experiment 1a and 1b. More specifically, Experiment 2a and 2b are designed to test whether parallelism effects can be observed in a subordinate context using “whereas” in a similar way as I see in the coordinate context using “and”.

Experiment 2a

Experiment 2a is a forced choice study aimed at determining whether the parallelism effects witnessed in Experiments 1a-b are specific to conjuncts coordinated by "and", or can be extended to other constructions such as a subordinate environment using "whereas". If parallelism effects are not strictly limited to syntactically coordinate environments, then a bias for the construction of parallel dependencies as indexed by pronoun choice (as in Experiment 1a) is predicted. If, however, the relevant domain is structural in nature, then no parallelism effect is predicted to surface. As there was no evidence for offline locality effects in Experiment 1a, and there is not an obvious reason why the introduction of "whereas" would impact this, I do not predict locality of an antecedent to influence pronoun choice in Experiment 2a. As in previous experiments, a table outlining different predictions is offered in Table 7.

1st Conjunct Dependency	**Factor**			
	Locality	*Priming*	*Parallelism*	*Discourse Coherence*
Baseline	local	NA	NA	NA
Matrix	local	matrix	no preference	matrix
Local	local	local	no preference	local

Table 7:Table of predictions for Experiment 2a. Cell contents refer to the pronoun–antecedent relation made by the chosen pronoun.

Materials and Methods

Experimental items were adopted from Experiment 1b[6], with two key adaptations. The first change made was to eliminate the possessive pronoun in the second conjunct to create a gap (the pronouns were shown after the sentence), and the second was to replace the conjunction "and" with "whereas", which served to create a subordinate rather than coordinate context. Following the tests outlined earlier in this paper, "whereas" appears to create a syntactically subordinate environment, although conversely it may also introduce expectations of parallel coherence relations.

Participants

Participants were 40 Northwestern undergraduates recruited from undergraduate classes in the Linguistics department. All participants were self-reported native speakers, and were at least 18 years of age. Subjects received course credit for their participation, and gave informed consent under the approval of the Northwestern University Institutional Review Board. Prior to analysis 2 participants were excluded due to extremely fast response time and lack of variation in their responses, e.g. choosing the button on the same side of the screen for large blocks of trials.

Materials

6 stimulus sets were created following a standard Latin square design, with items occurring in pseudo-random order. Each set consisted of 24 experimental items and 84 unrelated fillers, and no more than 2 experimental items were presented in a row. All items consisted of a sentence containing a gap, and two options to fill that gap. All options to fill the gap were pronouns, for

[6] The stimuli for Experiment 2a were adapted from Experiment 1b rather than 1a since some revision of the materials occurred between Experiments 1a and 1b, as noted in the materials section of Experiment 1b.

both target and filler items. Exactly like in Experiment 1a, target items were categorised into three distinct conditions: baseline, matrix dependency, or local dependency, with the critical distinction being that clauses were related with the subordinator "whereas" as opposed to the conjunction "and", as shown in (24).

(24) **BASELINE**

a. Mary said Tom hates ___ father (his/her)

MATRIX DEPENDENCY

b. Bill said Jane loves his mother whereas Mary said Tom hates ___ father (his/her)

LOCAL DEPENDENCY

c. Bill said Jane loves her mother whereas Mary said Tom hates ___ father (his/her)

Besides the substitution of "whereas" for "and", all other design features and considerations are identical to those in Experiment 1a.

Procedure

Participants accessed this experiment via the Northwestern University Department of Linguistics research participation website (SONA), where they were given a link to a survey hosted externally on Firebase (Amazon Web Services). Participants were not required to perform this experiment in the lab. Upon accessing the experiment, subjects were instructed that they would see 108 English sentences, and that each sentence would contain a gap. They were told to fill the gap by choosing one of two options given below the sentence, and to choose whichever option sounded the most 'natural' to them. Prior to the experimental items, subjects were shown two example items to demonstrate the task. Each item was presented to the subject on a unique page, and participants were unable to return to previously answered items.

Analysis and Results

As in Experiment 1a, results were analysed via a logistic regression model[7]. The counts and percentages of pronoun choice are summarised in Table 8.

	Matrix dependency pronoun	**Local dependency pronoun**
Baseline	133 **44%**	171 **56%**
1st Conjunct Matrix Dependency	183 **60%**	121 **40%**
1st Conjunct Local Dependency	92 **30%**	212 **70%**

Table 8:Raw counts and percentages for pronoun–antecedent dependency choice by condition for Experiment 2a.

A single linear mixed effects logistic regression model using random subject and item intercepts demonstrated that experimental conditions dramatically impacted the construction of either a local or matrix dependency in the second conjunct. A binary variable indexing whether subjects chose a local antecedent, coded as '1', or a matrix antecedent, coded as '0' served as the

[7] glmer (choice ~ 1 + condition + (1 | subject) + (1 | item), data = data, family = "binomial")

dependent variable. The choice of a local or matrix antecedent in the baseline condition was at chance, however the matrix condition strongly predicted a matrix pronoun choice, and the local condition strongly predicted a local pronoun choice. Estimates, standard error, z-values and their associated p-values are reported in Table 9.

condition	**estimate**	**standard error**	**z-value**	**p-value**
baseline	.37	.29	1.25	.21
matrix	-.86	.22	-3.94	<.001
local	.66	.21	3.16	.01

Table 9:Estimates, standard error, z-value, and p-value from a single logistic regression model for Experiment 2a.

Given the suggestive nature of the numerical difference between parallel choices in the local and matrix conditions, an additional mixed effects logistic regression model with subject and item intercepts[8] was performed using 'parallel choice' as the dependent measure. For the matrix and local conditions this measure indexed whether a parallel pronoun was chosen to complete the sentence, coded as '1', or a non-parallel one, coded as '0'. The baseline condition was excluded from this model, and the data was treatment coded with the matrix condition set as the reference level. The dependent measure of 'parallel choice' as opposed to 'local' was employed so as to more directly assess the relative effect of parallelism across the two conditions. Without including the baseline condition, any significant effect of having a local relation in the first conjunct could simply index that more local choices were made in the local

[8] glmer(parallel_choice ~ 1 + condition + (1 | subject) + (1 | item), data = data, family = "binomial")

condition as compared to the matrix condition rather than indexing whether a larger *parallelism* effect surfaces in one of the two conditions. Analysis revealed that there was a significant difference in parallel choice between the local and matrix conditions, such that the parallel choice was more often selected in the local condition as compared to the matrix condition. Estimates, standard error, z-values and their associated p-values are reported in Table 10. This result seems to indicate that a locality effect is present in the sentences containing two clauses connected with the subordinator 'whereas', but not in the baseline condition.

condition	**estimate**	**standard error**	**z-value**	**p-value**
matrix	.55	.22	2.52	.01
local	.5	.19	2.7	.007

Table 10:Estimates, standard error, z-value, and p-value from a single logistic regression model analyzing parallel choice for Experiment 2a.

Discussion

The finding from this analysis was remarkably clear: parallelism had a marked impact on pronoun choice even in a subordinate environment. Subjects were far more likely to choose a pronoun that created a parallel dependency than a non-parallel dependency. Furthermore, as in Experiment 1a, I did not observe an offline effect of locality as evidenced in the logistic regression model by the lack of significant bias in the baseline condition. Interestingly, however, a locality effect did appear to emerge in the critical sentences containing the connecter 'whereas'. I refrain from drawing strong conclusions based on this difference, however this may be a reflection of online processing occurring prior to the selection of a pronoun in the offline

task.

Experiment 2b

Experiment 2b expands on Experiment 2a via an eye tracking while reading study in which parallelism is varied in a subordinate environment to examine whether parallelism effects are unique to coordinate contexts using "and" in online processing. The inclusion of an online task is crucial to (1) show that parallelism is operating in real-time processing and (2) identify the context in which such effects are elicited (coordination with "and" vs. subordination with "whereas") during online comprehension. Given the results of Experiment 1a, I predict that a parallelism effect will be present such that parallel dependencies will cause facilitation witnessed at or directly after the pronoun in the second conjunct, resulting in faster eye tracking measures in the parallel conditions. Additionally, locality is also predicted to affect processing, such that pronouns that have a more local antecedent will also demonstrate facilitation. Predictions made by different frameworks are provided in Table 11.

Condition	**Factor**			
	Locality	*Priming*	*Parallelism*	*Discourse Coherence*
Local/Parallel	faster	faster	same	faster
Matrix/Parallel	slower	faster	same	faster
Matrix/Non-Parallel	slower	slower	same	slower
Local/Non-Parallel	faster	slower	same	slower

Table 11:Predictions for eye tracking measures for Experiment 2b. Cell contents indicate the predicted effect of that factor on eye tracking measures at or directly after the pronoun in the second conjunct.

Method and Materials

Participants

Forty participants from the Northwestern University undergraduate community were recruited for this experiment. All participants were native English speakers, and all had normal or corrected-to-normal vision. Subjects received course credit for their participation. One subject was subsequently excluded prior to analysis due to low comprehension question accuracy (< 75%).

Materials

Materials were composed of 24 sentences like (25), in which the length of a pronoun–antecedent dependency was manipulated in both the first and second sentences by changing the gender of the matrix and embedded subject noun-phrases, resulting in two Parallel conditions and two Non-parallel conditions. Names were identical to those used in Experiment 1b. As in Experiment 1b, the pronoun in the second sentence unambiguously resolves to a noun-phrase either in matrix or embedded position, resulting in two Local conditions and two Matrix positions. The only notable difference between these stimuli and those of Experiment 1b is the use of "whereas" in Experiment 2b to create a subordinate environment in contrast to the use of "and" in Experiment 1b. A full set of experimental items is included in the Appendix.

(25)

Local/Parallel

a. Michael thought Emma missed **her** dog whereas Sarah thought Max despised **his** cat with a great hatred

Matrix/Parallel

b. Emma thought Michael missed **her** dog whereas Max thought Sarah despised **his** cat with a great hatred

Matrix/Non-parallel

c. Michael thought Emma missed **her** dog whereas Max thought Sarah despised **his** cat with a great hatred

Local/Non-parallel

d. Emma thought Michael missed **her** dog whereas Sarah thought Max despised **his** cat with a great hatred

Procedure

Procedure was identical to that of Experiment 1b. Participants' gaze was recorded using a tower-mounted SR Research Eyelink 1000 Plus at a sampling rate of 2000 Hz. Fixations shorter than 40 ms were incorporated into adjacent fixations. Each experimental session was begun with four practice trials, two of which had comprehension questions. Re-calibration was performed between trials if necessary, and subjects were given short breaks as needed. The experimental sentences were presented according to a standard Latin Square, and were randomised. All sentences appeared on one line, in Monaco 14-point font. Participants answered a yes/no comprehension question after every trial.

Analysis and Results

For this experiment, I will concentrate on eye tracking measures at three regions: the critical region containing the pronoun in the second conjunct (*his cat)*, the spillover region containing the preposition (*with*), and the second spillover region containing a determiner and adjective (*a*

great). Regions are shown in (26). No significant effects were found in any subsequent region. As in Experiment 1b, first fixation duration, first pass duration, regression path duration, and total time duration were analysed.

(26) Michael thought Emma missed her dog whereas Sarah thought Max despised [**his cat]$_1$ [with]$_2$ [a great]$_3$** hatred

As in Experiment 1b, statistical analyses were carried out on log-transformed durations for each region and measure using linear mixed effects regression and treatment coded factors[9]. Table 12 reports the means and standard error in milliseconds of reading times. $\chi 2$-values and their corresponding p-values accompanied by the estimates and standard error calculated from the maximal model are reported in Table 13, with significant or near significant effects indicated in bold.

Dependency Length	Parallelism	Critical region	Spillover 1	Spillover 2
FIRST FIXATION				
local	parallel	257 (9)	226 (10)	250 (8)
matrix	parallel	277 (9)	238 (11)	254 (11)
matrix	non-parallel	271 (10)	243 (10)	276 (11)
local	non-parallel	276 (9)	244 (12)	268 (10)

[9] For example: lmer (logTime ~ (1 + dep_length * parallelism |subject) + (1 + dep_length * parallelism | item) + dep_length * parallelism, data = data)

FIRST PASS				
local	parallel	326 (13)	231 (10)	291 (12)
matrix	parallel	345 (15)	238 (11)	298 (18)
matrix	non-parallel	352 (14)	243 (10)	351 (17)
local	non-parallel	371 (18)	244 (12)	325 (15)
REGRESSION PATH				
local	parallel	520 (49)	495 (94)	724 (112)
matrix	parallel	487 (36)	298 (21)	899 (129)
matrix	non-parallel	609 (89)	384 (42)	1237 (153)
local	non-parallel	620 (69)	340 (35)	1155 (171)
TOTAL TIME				
local	parallel	646 (35)	346 (21)	468 (25)
matrix	parallel	651 (32)	312 (18)	512 (30)
matrix	non-parallel	631 (32)	314 (19)	524 (27)
local	non-parallel	687 (37)	319 (21)	480 (25)

Table 12:Means and standard error in raw values for each fixation measure and region in experiment 2b.

Region	Effect	Estimate	Std error	χ^2 (df)	p-value
First Fixation					
Critical	dependency length	.036	.03	1.38 (1)	>.1
	parallelism	-.021	.03	.507 (1)	>.1
	interaction	**.116**	**.06**	**3.69 (1)**	**.05**

Spillover 1	dependency length	-.004	.05	.005 (1)	>.1
	parallelism	-.044	.06	.64 (1)	>.1
	interaction	.887	.08	.763 (1)	>.1
Spillover 2	dependency length	.002	.04	.004 (1)	>.1
	parallelism	**-.08**	**.04**	**4.44 (1)**	**.04**
	interaction	-.05	.07	.44 (1)	>.1
First Pass					
Critical	dependency length	.02	.04	.4 (1)	>.1
	parallelism	**-.07**	**.04**	**4.32 (1)**	**.04**
	interaction	.1	.08	1.71 (1)	>.1
Spillover 1	dependency length	-.02	.05	.19 (1)	>.1
	parallelism	-.02	.05	.33 (1)	>.1
	interaction	.03	.11	.09 (1)	>.1
Spillover 2	dependency length	.006	.04	.02 (1)	>.1
	parallelism	**-.15**	**.05**	**11.37 (1)**	**.0007**
	interaction	**-.15**	**.08**	**3.02 (1)**	**.08**
Regression Path					
Critical	dependency length	.02	.06	.12 (1)	>.1
	parallelism	-.09	.06	2.1 (1)	>.1
	interaction	.02	.1	.05 (1)	>.1
Spillover 1	dependency	-.03	.09	1.03 (1)	>.1

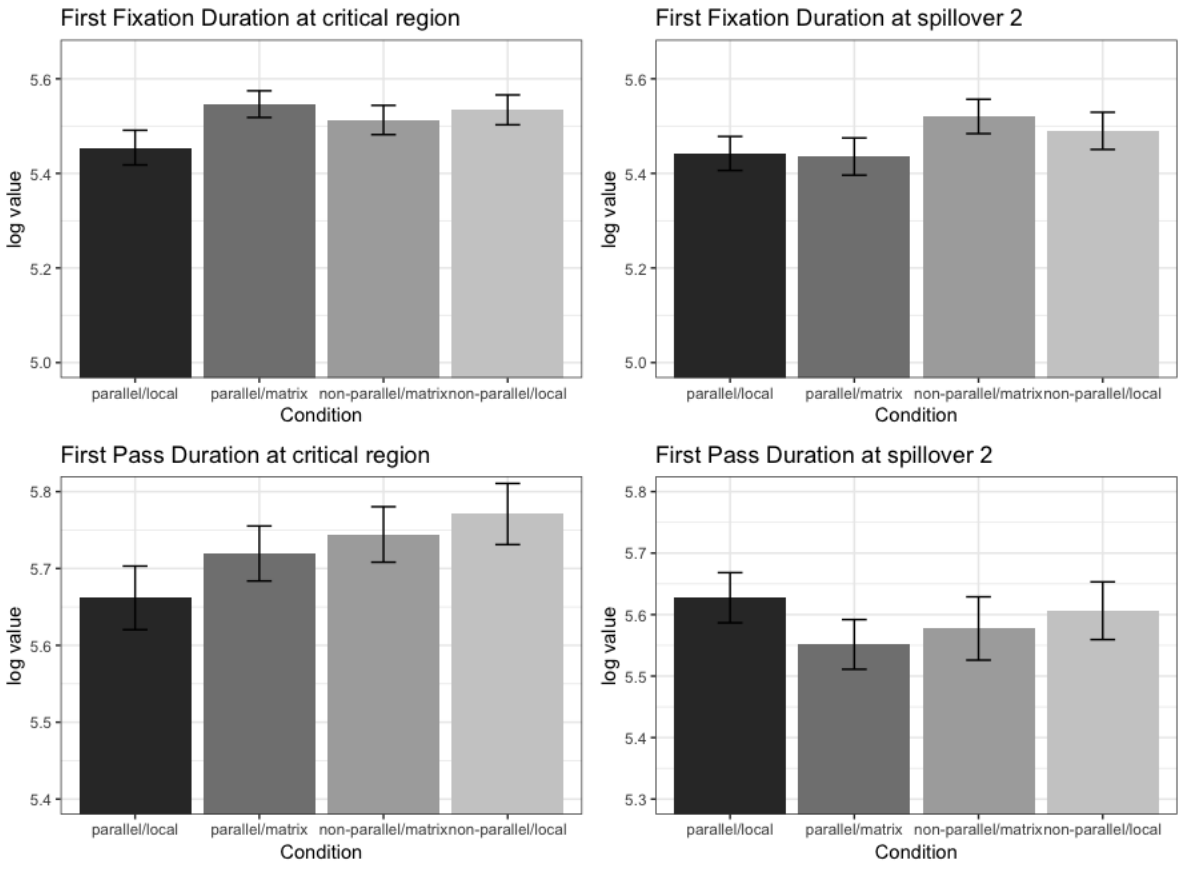
	length				
	parallelism	-.08	.08	.08 (1)	>.1
	interaction	-.20	.15	1.87 (1)	>.1
Spillover 2	dependency length	.12	.09	1.51 (1)	>.1
	parallelism	**-.24**	**.09**	**6.13 (1)**	**.01**
	interaction	-.05	.16	.09 (1)	>.1
Total Time					
Critical	dependency length	.007	.05	.02 (1)	>.1
	parallelism	-.03	.05	.35 (1)	>.1
	interaction	.04	.09	.17 (1)	>.1
Spillover 1	dependency length	-.07	.06	1.2 (1)	>.1
	parallelism	.04	.06	.49 (1)	>.1
	interaction	-.12	.12	1.09 (1)	>.1
Spillover 2	dependency length	.06	.05	1.38 (1)	>.1
	parallelism	-.08	.05	2.58 (1)	>.1
	interaction	-.1	.1	.86 (1)	>.1

Table 13:Estimates, standard error, χ^2 values, and p-values for each fixation measure and region in experiment 2b.

At the critical region, a significant interaction of parallelism and dependency length of the second conjunct was observed in first fixation time, such that fixation time for the condition with a local (short) dependency in the second conjunct was faster than the condition with a

matrix (long) dependency for the parallel conditions, but no such difference was observed in the non-parallel conditions. In addition, a main effect of parallelism emerged in first pass reading times at the critical region. While it is unclear why the length of the dependency should matter only when the dependencies in both conjuncts are parallel, the main effect of parallelism seems to indicate that the length of the dependency in the first conjunct is impacting the ease in which a subsequent dependency in the second conjunct is processed.

No significant effects were observed in the first spillover region. In the second spillover region, a main effect of parallelism surfaced in first fixation, first pass, and regression path duration, such that conditions with parallel dependencies demonstrated facilitated processing compared to conditions with non-parallel dependencies. Bar plots of significant measures are shown in Figure 3.



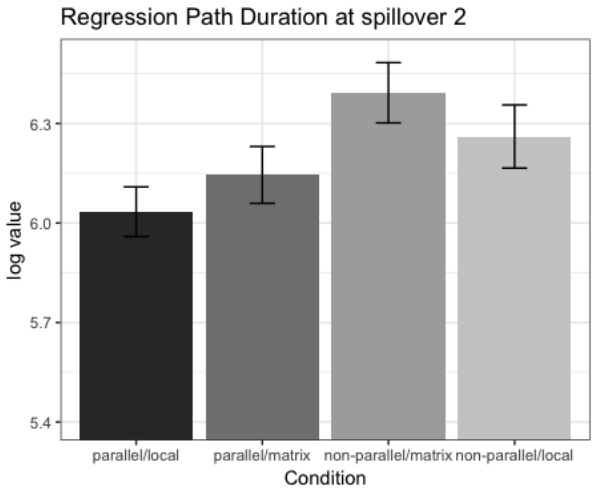


Figure 3:Bar plots of first fixation duration and first pass duration for the critical region, and first fixation duration, first pass duration, and regression path duration for the second spillover region. Error bars represent standard error.

Discussion

While the results from Experiments 2a-b differ from those of Experiments 1a-b in some ways, what is similar is that an effect of parallelism clearly emerges. In the offline experiment, an obvious effect of the first dependency is again observable such that subjects were far more likely to choose a pronoun which created a parallel dependency to that of the first dependency. In the online experiment, this effect first is seen at the critical region in first pass time, but more obviously is witnessed in the second spillover region in first fixation time, first pass duration, and regression path duration such that parallel conditions elicited faster eye tracking measures than non-parallel conditions. As in Experiment 1, these results strongly indicate that the position of the antecedent and the corresponding pronoun–antecedent relation in the first clause is impacting dependency formation in the second clause.

In opposition to Experient 1b, I do not see a main effect of locality in Experiment 2b using “whereas”. I am hesitant to reach any strong conclusions based on the observed interaction

such that locality appeared to facilitate processing only in the parallel conditions, however there is the potential that this could be attributed the strength of the expectation of parallelism introduced by "whereas".[10] It is possible that in terms of coherence relations that "whereas" induces *stronger* expectations than "and", and as such the violation of parallelism causes a more severe disruption in processing. The effects of this disruption may obscure the observation of locality effects in the non-parallel conditions, and subsequently facilitation due to locality of the antecedent is only identifiable in the parallel conditions.

Importantly, the replication of the parallelism effect in Experiment 2 is unexpected assuming that parallelism effects more broadly are attributable in some way to structurally coordinate constructions, such as those exemplified by "and", rather than being impacted by discourse level considerations or more general priming effects. The robustness of both the offline and online results supports the supposition that parallelism effects at various levels of representation are not limited to syntactic coordination.

A final note is warranted regarding some of the statistical assumptions made in this paper, in particular regarding the number of dependent measures and statistical models examined when analysing data collected from eye tracking experiments. Eye tracking is a valuable tool in the sentence processing arsenal for both the temporal resolution and the sheer quantity of data it provides. However, standard practices in eye tracking research have recently come under criticism (von der Malsburg & Angele, 2017) due to the potential for false positives to arise as a result of the quantity of statistical tests applied to the data. In this work I maintained .05 as my threshold for significance, however in subsequent work it may be more appropriate to apply a

[10] I thank an anonymous reviewer for this suggestion.

Bonferroni correction and increase the statistical power by expanding both my item set and subject pool.

General Discussion

Two forced choice sentence completion experiments and two eye tracking experiments explored whether pronoun dependencies are subject to parallelism effects, and if so whether such effects arise specifically in coordinate environments or can be attributed to general priming mechanisms or discourse coherence. I additionally investigated whether the distance of the relevant antecedent from the pronoun created a preference for more local antecedents. In Experiment 1a-b I manipulated the length of a pronoun dependency in two conjuncts conjoined by "and", and investigated whether the pronoun–antecedent relation in the second conjunct was preferred or facilitated when the length of the dependencies and position of the antecedent was paralleled across the conjuncts. In Experiment 2a-b I retained the manipulations of Experiment 1a-b, but changed the relationship between the clauses such that second dependency now occurred in a clause introduced by "whereas".

Locality of the antecedent does indeed appear to affect the processing of pronoun–antecedent dependencies, as evidenced by the eye tracking experiments, such that a linearly closer antecedent facilitates processing of the pronoun and subsequent regions. This result is expected under current models of retrieval that incorporate an element of decay[11] for items stored

[11] It has been suggested that decay may no longer be a useful explanatory factor in the retrieval literature (Berman, Jonides, & Lewis, 2009; Lewandowsky & Oberauer, 2009; McElree, 2006). I note, however, that decay is not an entirely unfounded factor (e.g.Van Dyke & Lewis, 2003), and that disentangling the effects of interference and decay is a difficult task.

in short term memory. This result, however, is also compatible with a feature overwriting account (Nairne, 1990; Oberauer & Kliegl, 2006) in which introduction of new items in memory may cause interference with older items that overlap in content, resulting in older representations becoming less distinct than newer ones. At present my results do not distinguish between these two accounts.

Crucially, the studies outlined and reported above provide evidence that parallelism (or priming) is operating over pronoun dependency relations, and demonstrate its effect in both offline and online measures. This effect is particularly clear in the forced choice experiments, where an extremely robust effect of parallelism was witnessed in both "and" coordination as well as subordination introduced by "whereas". The presence of a preference or facilitation of pronoun dependencies due to parallelism reveals that this relationship has a representation stored in memory, however it is not entirely obvious the precise source of the observed results, and how they can be accommodated in various frameworks.

One possibility is that I can attribute these parallelism effects as an instance of structural priming, which is an attractive account in that the structural location of the antecedent is easily specified, as well as the relative position of the pronoun. As Bock (Bock, 1986) noted and Pickering and Ferreira (Pickering & Ferreira, 2008) reiterate, 'structural priming provides evidence for the mental representation of syntax', and by observing an effect of parallelism in my studies I can assume that the pronoun–antecedent relationship indeed has a mental representation over which parallelism can operate. It is interesting, however, that my results appear to contrast with the findings of Knoeferle (2014), in which a simple priming account was rejected based on evidence that conjunction meaning ("and" vs. "but") as well as coordination ("and") vs. subordination ("while") impacted the parallelism effect on constituent order, such

that only “and” reliably elicited faster eye tracking measures. The author argues that a semantic account in which “and” encourages a parallel relation, whereas “but” and “while” encourage a contrastive relation, is the source of these differences. However, I am reluctant to view the comparison of my results as a true contrast – it may be that the difference in my pattern of results is a consequence of the differences between the representations being paralleled, and the context in which these representations are occurring. Less straightforward is the form which this representation might take: what precisely is being primed or paralleled?

My results may be compatible with models such as structured search or copy mechanisms. In the first framework, it is imaginable that the relation between the pronoun and antecedent could be represented as a kind of “search path” (Schneider & Phillips, 2001), which could then be primed when another pronoun is encountered in the second conjunct. In terms of a copy-based priming mechanism, I might theorise that the presence of coordination sets up a presumption of syntactic parallelism, which may involve copying the relevant structure from the first conjunct. If this structure also carries the indices of the pronoun–antecedent relationship, then in some sense the parallel relationship is already built into the structure before encountering any bottom-up information. In this type of copy-based priming scenario where the entire first conjunct including the pronoun-antecedent indices is activated, it is possible that the retrieval mechanism does not need to be employed at all. However, evidence of a locality effect in the parallel conditions of my experiments appear to refute this idea – if retrieval was not occurring in the second conjunct in these conditions then the distance of the antecedent from the pronoun (locality) should have no impact on the time course of processing. Furthermore, this type of account would need to explain why I continue to witness parallelism effects in the absence of

coordination since presumably structure copying would not be performed without some kind of trigger, e.g. the presence of a conjunction like “and”.

If these informal observations regarding the status of “whereas” are true, then the appearance of parallelism effects in certain syntactically subordinate environments may be due to this type of coherence, and may be contingent on the kind of relation introduced by the subordinating element. To probe this issue, one direction for future work could include a design that includes the factors I have already manipulated, namely syntactic environment and parallelism of dependencies, but also manipulates the kind of coherence relation introduced by the connective. If designed carefully to ensure that the alternate coherence relation predicts *non-parallel* dependencies in some conditions, this kind of design could shed light on the ultimate source of the effects I witness.

As previously stated, it is potentially more difficult to accommodate my results in currently assumed cue-based models of memory access and retrieval. The relationship between the pronoun and the antecedent is created by the act of retrieval and the ensuing coreference relation, however once this relationship has been established how does it influence subsequent pronoun relations? Antecedent selection is generally assumed to be feature based, however in the case of the pronoun dependencies in Experiments 1 and 2, it is insufficient to rely on syntactic position as the relevant cue to be primed or paralleled, as both potential antecedents are in subject position. Furthermore, it has been noted that it is difficult to represent relational information, such as c-command (for a review of this issue see (Kush et al., 2015), as features of individual items. Since the relationship between a pronoun and its antecedent is not yet apparent when encoding the antecedent, marking a noun as the antecedent of a pronoun would have to happen retroactively, and this would still be insufficient to “tell” the parser the relative position

of the previously retrieved antecedent when processing a new dependency. Also note that it is ambiguous in my experiments whether the structural position of the antecedent is the relevant information, or whether it is the relative positions of the pronoun and antecedent. The former may appear to be more easily represented as a combination of cues indexing syntactic position and embedding depth, and indeed in the experiments outlined in this paper a cue indexing subjecthood as well as a cue indexing whether that subject was in a matrix or embedded clause would be sufficient to single out the relevant antecedent. However, this kind of "embedding" cue would swiftly encounter problems once the structure surpassed two levels of embedding, inasmuch as information involving c-command would need to be encoded, thus returning us to the original problem of capturing relational information in a cue-based system.

Similarly, if the relative position of the antecedent and pronoun is the relevant information being reactivated, then its representational status in a feature-based system is also uncertain. Future studies manipulating both the position of the antecedent and the position of the pronoun across conjuncts would aid in clarifying what is being represented, but it is unlikely they will shed light on how to encode that representation in a feature-based system.

Apart from the nature of the dependency representation, we may also wish to ask how exactly parallelism of such a representation is affecting online processing in a cue-based framework. More concretely, is parallelism limiting the search space by "guiding" the retrieval mechanism to a parallel antecedent, or does parallelism of antecedent position relative to the pronoun facilitate pronoun resolution in terms of retrieval latency? Given that I do *not* primarily witness the parallelism effect in early eye tracking measures, I may speculate that an early (albeit) weak filter on the pool of potential antecedents is an unlikely scenario. Rather, I speculate that parallelism may be affecting the ease and speed with which an antecedent is

selected, resulting in slower eye tracking measures when the pronoun picks out a non-parallel antecedent. Alternatively, it is possible that in fact what I am witnessing is a reanalysis cost in the non-parallel conditions. This may be an attractive account given that I see an effect of parallelism in regression path duration in both eye tracking experiments, a measure which has been observed to reflect misanalysis during processing. In the non-parallel conditions of Experiments 1b-2b this measure may index recovery from an incorrect linking of the pronoun to the parallel antecedent. However, the distinction between these two accounts based on the types of measures in which an effect is observed is not straightforward, and I refrain making a strong claim to this effect.

An important point to emphasize is that the effect is significant in the absence of lexical priming - the parallel pronoun in all instances was the pronoun of the opposite gender of the previously processed pronoun, and all other lexical material in the embedded clauses was unique across conjuncts. This design choice allows us to more confidently assume that what is witnessed in these experiments is due to parallelism of pronoun antecedent coreference relations. However, there remains the possibility that while the lexical material is different, some sort of pattern of alternating gender values on the nouns and pronouns is the relevant information. For example if the pattern is [+masc] [-masc] [+masc] (**John** said **Mary** hated **his** dog) in the first conjunct, once the initial noun of the second conjunct is recognized as [-masc], there may be an advantage to continue the alternation with [+masc] and [-masc]. Although I cannot currently disambiguate between these possibilities, work is currently being conducted to examine contexts where no coreference is established to determine whether what matters is indeed the pronoun antecedent relationship.

Independently of the particular theory assumed or the kind of information parallelism is operating over, the results reported for Experiment 2a-b provide support for the idea that parallelism effects are not unique to “and”, but can likely be attributed to more general principles. If parallelism effects were truly due to coordinate structures and the expectation for parallelism that such a construction creates, then I would only expect to see the parallelism effect in Experiment 1, but not in the subordinate environment employed in Experiment 2. A less absolute position on the source of parallelism effects may be that while parallelism effects are possible and/or present in various environments, the effect is larger in “and” coordination. There does appear to be some evidence for this idea, for example Dubey, Sturt, and Keller (Dubey, Sturt, & Keller, 2005) and Dubey, Keller, and Sturt (Dubey, Keller, & Sturt, 2008) find an effect of parallelism in both coordinate and arbitrary environments, but also note that this effect is strongest in the case of coordination. However, they failed to replicate this finding using eye tracking (Sturt, Keller, and Dubey, 2010). Examining mown experiments, it is not clear if I see a difference in effect sizes between Experiments 1 and 2. Although on the surface it appears that Experiment 1a demonstrates a more dramatic effect of parallelism than Experiment 2a, there are other differences between these experiments that may give rise to such variation, for example the number of participants and the mode of presentation in the two studies. At this point I reserve speculating on whether there are notable contrasts between the parallelism effects that arise in coordination and subordination until such time that I can directly compare them within a single experiment.

As noted, the configuration of the stimuli used in Experiments 1 and 2 may allow for another interpretation. Since the sentences I used have only one level of embedding, a cue that indexed whether a subject was in the matrix clause or not would be sufficient to isolate the

parallel antecedent. However, this proposal would face difficulty once another embedded clause was introduced, and a binary cue of this nature would be unable to distinguish one embedded subject from another. While my stimuli do not permit us to observe whether parallelism effects would surface when comparing one embedded antecedent to another, I have a strong intuition that this would be the case. However, the sentences for such an investigation become quite unnatural and ludicrous. An alternative to adding yet more levels of embedding to the sentences used in Experiments 1 and 2 is to broaden the investigation to see whether I might witness evidence for the presence of structural information in memory in other configurations and environments. One such environment is ellipsis.

Chapter 3: Antecedent Structure Matters

Introduction

Recent research in human sentence processing has devoted substantial attention to the processing of ellipsis, in particular to the mechanism underlying the resolution of unpronounced material. Previous research has investigated various types of ellipsis, with much earlier work focused on verb-phrase ellipsis (Frazier & Clifton Jr, 2005; Kim, Kobele, Runner, & Hale, 2011; Martin & McElree, 2008; Roberts, Matsuo, & Duffield, 2013; Tanenhaus & Carlson, 1990) , although more recent studies have expanded to include constructions such as sluicing (Clifton, 1998; Dickey & Bunger, 2011; Frazier & Clifton, 2001; Harris, 2015; Yoshida, Dickey, et al., 2013), and gapping (Carlson, Dickey, & Kennedy, 2005; Frazier, 2015; Kaan, Overfelt, Tromp, & Wijnen, 2013; Kaan, Wijnen, & Swaab, 2004; Kim, Carlson, Dickey, & Yoshida, 2020), among others. During the processing of any type of ellipsis, the parser must engage the following processes. First, the parser must recognise an ellipsis site. In other words, the parser must recognise that some material in the sentence is missing. Second, once the ellipsis site has been recognised, the parser must search for and identify the antecedent. The interpretation of the ellipsis site is normally dependent on its antecedent. Therefore, to determine the meaning of the ellipsis site, the parser needs to find the antecedent of the ellipsis. Finally, once the antecedent is identified, the parser needs to integrate the materials in the antecedent into the ellipsis site (Yoshida, 2018).

(27) is an instance of the sluicing ellipsis construction (Ross, 1969a). In sluicing, the clausal material that follows the wh-phrase is missing. When the parser processes the sluicing construction, the parser must recognise that there is unpronounced material following "why", identify and access the antecedent, retrieve from memory the content of the antecedent "Mary slept", and integrate the retrieved information of the antecedent into the elided clause following "why".

(27) Mary slept for some reason, but I don't know why ~~Mary slept~~.

Background

Ellipsis resolution and retrieval models

While ellipsis is pervasive in language, the mechanisms that allow for successful and rapid processing of ellipsis are not well understood. Although the recognition of an ellipsis site as well as its antecedent is not a trivial matter (see Yoshida, 2018; Yoshida, Lee, et al., 2013), increasing attention has been paid to the nature of the process underlying the recovery of previously processed material, namely the antecedent of the ellipsis site, and the status of the material represented at the ellipsis site. Although there are many accounts of antecedent retrieval, recent research has lent support to models that assume a cue-based pointer to content addressable memory (Lewis & Vasishth, 2005; Martin & McElree, 2008). To reiterate, in these models, cues from the ellipsis site are checked against features of possible antecedents, such that a maximally matching antecedent will be retrieved and integrated into the current parse. These models are in contrast to serial search models, which assume that an ellipsis site triggers a linear iterative search of recently processed material.

Antecedent complexity and structural sensitivity

To restate some background information from Chapter 1, a lack of antecedent complexity effects on the time course of ellipsis processing has been used as support for models in which all elements are accessed directly and in parallel. Stimuli from a key example from Martin and McElree (2008) is recreated below in (28), where they were unable to witness any effect of complexity on a SAT task.

(28)

c. *Simple antecedent*

The history professor [understood Roman mythology], but the principal was displeased to learn that [the over worked students/*the overly worn books] attending summer session did not.

d. *Complex antecedent*

The history professor [understood Rome's swift and brutal destruction of Carthage], but the principal was displeased to learn that [the over worked students/*the overly worn books] attending summer session did not.

To explain the lack of complexity effects, Martin and McElree propose an unconstrained "Direct Access" cue-based model. In this model, as in other cue-based models, representations are directly accessed via the match between the ellipsis cues and the antecedent features. Furthermore, the output of initial processing is a memory representation that does not reflect the initial ease or difficulty of processing. That is, all memory representations are represented

equally, and that "representations with varying degrees of distinctiveness can be recovered in equal time" (Martin & McElree, 2011, p.2). In other words, retrieval processes during the processing of the ellipsis site should not be affected by the size or complexity of antecedent, and therefore processing speed at an ellipsis site should not differ.

Martin and McElree's results provide corroboration for an earlier study by Frazier and Clifton (2001). In this study the effect of antecedent length was examined in sluicing contexts, as shown in (30), and it was found that longer antecedents did not affect self-paced reading times. In contrast to Martin and McElree, Frazier and Clifton assume that structure is copied at the ellipsis site (Copy Alpha), but assert that this copying is "cost-free". They speculate that when the syntactic scope is clearly marked, for example by an interrogative phrase or "did", that a pointer to the antecedent directs the parser to the material which is then copied wholesale into the ellipsis site.

(29)

a.	Sarah left her boyfriend last May.	*Short antecedent*
b.	Sarah got up the courage to leave her boyfriend last May.	*Long antecedent*
c.	Tina did too.	*Ellipsis*

However, some earlier work did find a notable effect of antecedent complexity on reading times (Murphy, 1985), whose example stimuli are shown in (30), and it is worth noting that there were non-significant but suggestive trends in the study performed by Frazier and Clifton (2001).

(30)

a. Jimmy swept the floor. Later, his uncle did too.
b. Jimmy swept the tile floor behind the chairs free of hair and cigarettes. Later, his uncle did too.

Martin and McElree argue that the early result from Murphy can possibly be attributed to the quality of the memory representation of the antecedent, and furthermore assert that the structure sharing proposed by Frazier and Clifton is in essence entirely compatible with a pointer that directs the parser to the relevant representation to be shared, and that the subsequent copy operation Frazier and Clifton propose is unnecessary. In their examination of antecedent complexity, Martin and McElree used only the SAT task (although eye tracking was employed in one study to examine distance effects), which certainly by its nature provides insights on speed and accuracy, however it is not clear how loyally such a task would translate to other sentence processing measures, and whether a more naturalistic task may cast a different lens on real-time processing. Additionally, there may be some concerns about whether participants truly needed to process the ellipsis in order to determine the sensicality of the match between the subject of the embedded clause and the antecedent verb in (29). For example, to judge sensicality at the point of "the over worked students/*the overly worn books" in (29), the relation between "the books" and the main verb "understood" is sufficient to make an accurate judgement, and thus it may be that the rest of the antecedent material is neglected.

Although these studies are often cited as providing evidence for the implementation of a cue-based retrieval mechanism in the construction of long-distance dependencies, other studies suggest that the search process may be structurally constrained, and that the parser may respect

structural properties of the antecedent. Particularly compelling evidence for a grammatically sensitive retrieval mechanism comes from investigations of forwards and backwards anaphora, and the online application of the binding principles that constrain their distribution (Kazanina et al., 2007; Nicol & Swinney, 1989; Nicol, 1988; Sturt, 2003; Xiang et al., 2009). An example of such research comes from Dillon et al. (2013), who directly compare the processing of reflexive dependencies to subject-verb agreement dependencies in English, as shown in (31). Dillon et al. observe retrieval interference in examples like (31b), which involves subject-verb agreement, however failed to observe an interference effect in examples like (31a), which involve a reflexive dependency. This contrast in processing between dependency types is difficult to explain under a cue-based model without invoking additional mechanisms, which are potentially theoretically unmotivated.

(31)

a. The new executive who oversaw the middle manager(s) apparently doubted himself/*themselves on most major decisions.
b. The new executive who oversaw the middle manager(s) apparently were dishonest about the company's profits.

Evidence reported in the ellipsis processing literature also indicates that the consensus is not so clear regarding the retrieval mechanism. Specifically, results indicating that retrieval is unaffected by antecedent complexity are challenged by contradictory findings using extremely similar constructions. Carlson et al. (2005) examine the interpretation of gapping constructions, a type of ellipsis in which verbal material is absent in the second conjunct of a coordinate

structure. The authors argue that gapping constructions are not a uniform phenomenon, but rather that in some cases there are two available interpretations that employ structurally distinct representations of the second conjunct: one which involves ellipsis and one which involves across-the-board (ATB) movement, as shown in (32).These two representations vary in complexity, such that the subject interpretation involves recovering structurally more complex material than does the object interpretation. What Carlson et al. find is that when both analyses are available, comprehenders are biased towards the simpler object interpretation. Furthermore, in cases where both interpretations employ the same structural representation, the object bias is reduced. These results are attributed to differences in complexity, such that comprehenders avoid the interpretation that has greater structural complexity in the elided material.

(32)

Subject interpretation

a. [[IP Josh visited the office during the vacation] and [IP Sarah [IP during the week [IP t_{Sarah} visited the office t_{during}]]].

Object interpretation

b. Josh [vP visited [VP [VP t_{visit} the office during the vacation] and [VP t_{visit} Sarah during the week]].

In this research, we aim to determine whether complexity effects arise during antecedent retrieval during the processing of sluicing and sprouting constructions. Sluicing refers to clausal ellipsis which occurs following an interrogative phrase, as shown in (33a), where the interpretation of the elided clause is contingent on previously processed material in the sentence.

Sprouting is superficially extremely similar to sprouting, with the exception that there is no overt correlate to the wh-phrase in the antecedent material, as shown in (33b).

(33)

a. John ate **something**, but I don't know **what** ~~John ate~~ ___

b. John ate, but I don't know **what** ~~John ate~~ ____

Comprehenders appear to experience more processing difficulty for sentences involving sprouting (33b) than sluicing (33a) (Dickey & Bunger, 2011; Yoshida, Lee, et al., 2013). Specifically, there is an observed disruption in processing upon encountering the wh-phrase, which may be attributed to difficulty in inferring the unexpressed element in the antecedent. This effect overall can be explained in terms of parallelism, such that there is a penalty imposed due to a lack of parallel identity between the antecedent and elided clauses (Dickey & Bunger, 2011; Kim et al., 2011).

Sluicing is a potentially fruitful area to examine effects of antecedent complexity since there appear to be certain observations that certain interpretations are difficult to elicit despite their apparent grammaticality. In particular, it has been observed that when the wh-remnant in a sluicing construction is the correlate of an adjunct, then it is difficult to arrive at an interpretation which locates the adjunct in the embedded clause of a biclausal structure, as shown in (34).

(34) ? John said that Mary went to school **for some reason**, but I don't know **why** ~~John said that Mary went to school~~ ____

Islands

Another possible factor that may affect the retrieval process that has been under studied up to this point is the presence of an island (Ross, 1967), which also necessarily references structural information. Islands are structurally defined domains[12] that constrain the formation of long-distance dependencies produced by movement operations. For example, in (35), the so-called Complex Noun Phrase (NP) Constraint appears to prohibit extraction out of a definite determiner phrase (DP), which is demonstrated by the relative unacceptability of (35b) compared to (35a) where movement originates from within an embedded tensed clause.

(35)

[12] I note that not all approaches to island phenomena attribute their effects to structurally defined grammatical constraints. For example, some approaches include reference to structure, but ascribe island effects to processing demands on working memory (e.g. Christensen, Kizach, & Nyvad, 2013; Christensen & Nyvad, 2014; Deane, 1991; Hofmeister & Sag, 2010; Kluender & Kutas, 1993b). Others eschew an appeal to structure altogether and reference semantic or pragmatic considerations as the source of islands (for example Abrusán, 2011; Ambridge & Goldberg, 2008; Deane, 1991; Erteschik-Shir, 1973; Kroch, 1989; Kuno & Li, 1976; Szabolcsi & Zwarts, 1993). While it is true that many of the contexts that demonstrate bans on extraction are structurally complex, and I appreciate the appeal of non-domain specific theories as well as the intriguing insights offered by semantic and pragmatic accounts, I find criticism of these accounts quite credible (Phillips, 2013; Sprouse, Wagers, & Phillips, 2012; Yoshida et al., 2014). I believe there is substantial reason to refer to island constraints in structurally based terms, but refer the reader to the aforementioned material for alternative accounts.

c. Who$_i$ did Mary believe that John liked <who$_i$>?

d. *Who$_i$ did Mary believe the claim that Bill liked <who$_i$>?

Numerous types of islands have been identified since they were first documented, involving the apparent prohibition of extraction out of clausal adjuncts (Adjunct Island (Cattell, 1976; Geis, 1970)), coordinate structures (Coordinate Structure Constraint (Ross, 1967)), and subject position (Subject Condition (Chomsky, 1972)), to name a few among many (see also Kush, Lohndal, & Sprouse, 2018; Kush, Omaki, & Hornstein, 2013; Sprouse & Hornstein, 2013; Sprouse et al., 2012). Furthermore, island constraints have been categorized into two distinct classes, strong and weak islands, which refers to how absolute the constraint in question is (Abrusán, 2007; Cinque, 1990; Szabolcsi & Den Dikken, 1999; Szabolcsi & Lohndal, 2017).The term "strong islands" has been used to classify constraints which prohibit any extraction out of the relevant structure, whereas the term "weak islands" has been used for constraints which only prohibit extraction of certain kinds of phrases, e.g. they show "weak" effects. For example, in (36) the extraction of an argument out of a wh-island (weak island) appears to be permitted, however extraction of an adjunct results in unacceptability.

(36)

e. Which drink$_i$ did Ann ask how to mix t$_i$?

f. *How$_i$ did Ann ask which drink to mix t$_i$?

Although islands often result in strong reports of unacceptability, their effects sometimes seem to disappear or be reduced in the context of ellipsis (Chomsky, 1972; Chung et al., 1995;

Lasnik, 2001; Merchant, 2001; Potter, 2017; Potter & Görgülü, 2019; Yoshida, Potter, & Hunter, 2019). This effect, sometimes dubbed “island repair” or “island amelioration”, is demonstrated by the contrast of acceptability of (37a) compared to (37b). While both examples have the same meaning inasmuch as (37a) is the pronounced version of (37b), both also violate the Complex NP constraint. However, the example in which the offending structure has been elided (37b) seems to elicit greater perceived acceptability than the example in which the material and structure of the island is overtly pronounced.

(37)

a. *I believe the claim that she fired someone, but they don't know who I believe the claim that she fired.

b. (?) I believe the claim that she fired someone, but they don't know who.

One type of weak island that demonstrates this kind of argument-adjunct asymmetry is the so-called factive island, which prohibits extraction of an adjunct from within the clausal complement of a factive verb (Abrusán, 2007, 2011, 2014; De Cuba, 2006; Melvold, 1991; Oshima, 2006; Rooryck, 1992; Szabolcsi & Zwarts, 1993). Factive verbs are verbs which presuppose the truth of their complement, for example “hate” or “realize” (Adams, 1985; Beaver, 2010; Hooper, 1975; Karttunen, 1971; Kiparsky & Kiparsky, 1970). Example (38) demonstrates that extraction of an argument (38a) from the clausal complement of a factive verb is relatively acceptable (although perhaps not perfect), however extraction of an adjunct (38b) results in degraded acceptability.

(38)

a. What$_i$ do you regret that he solved t$_i$?

b. * How do you regret that he solved the problem t$_i$?

Of interest to the current research, one might wonder how island constraints may interact with ellipsis resolution, more particularly if the parser may respect island constraints by "guiding" the retrieval process to avoid selecting an antecedent which would create an island violation. For example, by using a factive verb combined with a sluicing construction as in (39), it may be possible to probe the sensitivity of the retrieval process to structural concerns such as islands. If the parser avoids retrieving an antecedent from which wh-extraction would violate an island, then it is indicative that structure is represented (a) in the memory representation of the antecedent or (b) in the ellipsis site post-retrieval, or both.

(39) *Bill hated that Sarah cooked the fish, but I don't know how$_i$ ~~Bill hated that Sarah cooked the fish t$_i$~~

Following from the summaries detailed above, it would appear that ellipsis and structurally defined phenomena such as islands may offer an avenue of investigation into determining both what kind of information is stored during processing, as well as what kind of information is retrieved. In this paper, we pursue whether complexity of the antecedent affects processing during ellipsis resolution, which would provide additional support for the presence of structural information in the memory representations of previously processed elements. In

addition, we employ a type of weak island to gain insight on whether the parser and the retrieval process will be impacted by the potential to create an island violation during ellipsis, which will also shed light on whether structural material is encoded in memory representations.

Experiments

The following experiments are designed to test whether the length of an antecedent affects the retrieval process during ellipsis, and whether island constraints are respected during the search and selection of an antecedent of ellipsis. Experiments 3a and 3b use sluicing and sprouting constructions to probe length effects via an acceptability judgment and eye tracking task. Experiments 4a and 4b use identical tasks to probe length effects and verb factivity in sluicing constructions.

Experiment 3a

Experiment 3a tested the sensitivity of comprehenders to differences in structural complexity of elided material in sluicing and sprouting, as well as to probe the claim that wh-adjunct extraction out of an embedded clause results in degraded acceptability. This experiment employs an acceptability rating task, with a 2x2 design with Ellipsis Type (sluicing vs. spouting) and Antecedent Length (long vs. short) as factors. A sample set of the stimuli is summarised in (40). Since this task is performed offline, it is not a direct test of online retrieval mechanisms. However, it is possible that processing difficulty may result in degraded judgments.

If offline and online processes in interpreting ellipsis are linked, direct access and cost-free copy accounts predict that we should not see a difference in acceptability judgments due to Antecedent Length. These accounts do not make explicit predictions regarding Ellipsis Type,

although they are compatible with the prediction that sprouting is more costly than sluicing and may therefore elicit poorer judgments. In sum, these models make a specific prediction that we will not observe an Antecedent Length effect in either sluicing or sprouting contexts.

Alternatively, if the retrieval process is sensitive to Antecedent Length, and this sensitivity is reflected in offline processing, then sentences with longer antecedents are predicted to be judged as degraded as compared to those with shorter antecedents.

Finally, if wh-adjunct extraction out of an embedded clause is dispreferred, then we may predict that the long sprouting condition will be judged as more acceptable than the long sluicing condition. In the long sluicing condition (40a) there is an overt antecedent for the wh-filler, and inclusion of the wh-correlate within the embedded clause directly after the embedded subject strongly encourages an embedded interpretation. In sprouting, however, the base position of this adjunct is unspecified, allowing for comprehenders to infer that the adjunct originated from the matrix clause. If comprehenders exploit this under-specification, then sprouting could ameliorate a dispreference for an embedded wh-adjunct gap.

Methods and Materials

Participants

64 self-reported native English speakers were recruited on Amazon Mechanical Turk. (Sprouse, 2011). Subjects were paid $2.00 for their participation. The Northwestern University Institutional Review Board approved this and the following experiment, and all participants indicated informed consent.

Materials

The stimuli for this experiment consisted of 24 experimental items, and 72 fillers. The independent factors of Ellipsis Type and Antecedent Length were crossed, such that each item had 4 conditions, as shown in (40).

(40)

Sluice/Long

a. Bill thinks that Mary, for some reason, quit her job, but I don't know why specifically, although I hope to find out soon.

Sluice/Short

b. Mary, for some reason, quit her job, but I don't know why specifically, although I hope to find out soon.

Sprout/Long

c. Bill thinks that Mary, after the meeting, quit her job, but I don't know why specifically, although I hope to find out soon.

Sprout/Short

d. Mary, after the meeting, quit her job, but I don't know why specifically, although I hope to find out soon.

64 unique lists were pseudo-randomized and counterbalanced, such that each participant saw only one condition from each item, and experimental items were separated by a minimum of one filler. Filler items included gapping and noun-phrase ellipsis constructions, however these fillers did not involve wh-filler/gap dependencies. A third filler type involved multiple wh-

phrases. No filler employed bi-clausal structures. The experimental items can be found in the Appendix.

Procedure

Upon viewing the experiment, participants were instructed that they would be completing a questionnaire in which they would be asked to judge how natural a sentence sounded. Prior to beginning the questionnaire, they were shown 2 example sentences and ratings, demonstrating a generally acceptable and generally unacceptable English sentence. Example items were identical across participants. Participants were instructed to rate each sentence on a scale of 1 (unnatural) to 7 (natural) based on their intuitions.

Analysis and Results

Two items were removed from analysis due to overall issues of interpretability (see Appendix). Raw judgments of the remaining experimental items were converted to z-scores within participants, and all subsequent analyses were based on these scores. The z-score transformation is useful in that in converts a participant's raw scores to values that represent the relationship between a subject's mean score and a particular rating's standard deviation from that mean. This corrects for individual differences in usage of the rating scale and enables us to compare ratings that originate from different normal distributions.

Linear mixed effects models were used to analyze the z-scored ratings[13]. The model was fit with a maximal random effects structure (Barr et al., 2013), and item random effects were eliminated until the model converged. The model was run using the lme4 package in R (Bates et al., 2015), and p-values for the fixed and random effects were evaluated using model comparison. Additionally, independent t-tests were performed to compare the effects of Antecedent Length within each Ellipsis Type.

Figure 4 presents the average z-score ratings for each condition of the experimental items. We found an overall effect of Ellipsis Type (β = -.19, S.E. = .05, χ^2 = 14.5 (1), p = .00001), such that sluicing conditions (mean z-score = .12) were judged to be more natural than sprouting conditions (mean z-score = -.07). We also found a significant effect of Antecedent Length (β = .12, S.E. = .05, χ^2 = 6.11 (1), p = .01), such that sentences with shorter antecedents were judged as more natural than longer antecedents within both the sluicing (.19 vs .04) and sprouting (-.02 vs. -.12) conditions.

Table 14 presents the estimated coefficients and standard error with for the Linear Mixed Effects model, and the p-values obtained using model comparison. Main effects were found for both Ellipsis Type and Antecedent Length, such that sluicing was judged as significantly more natural than sprouting, and sentences with shorter antecedents were judged as significantly more natural than longer antecedents.

[13] For example: model<-lmer(z_score ~ length*type + (1 + length*type | subject) + (1 + length*type | item), data = data). All subsequent models for experiments in this paper follow this structure for their respective dependent measures.

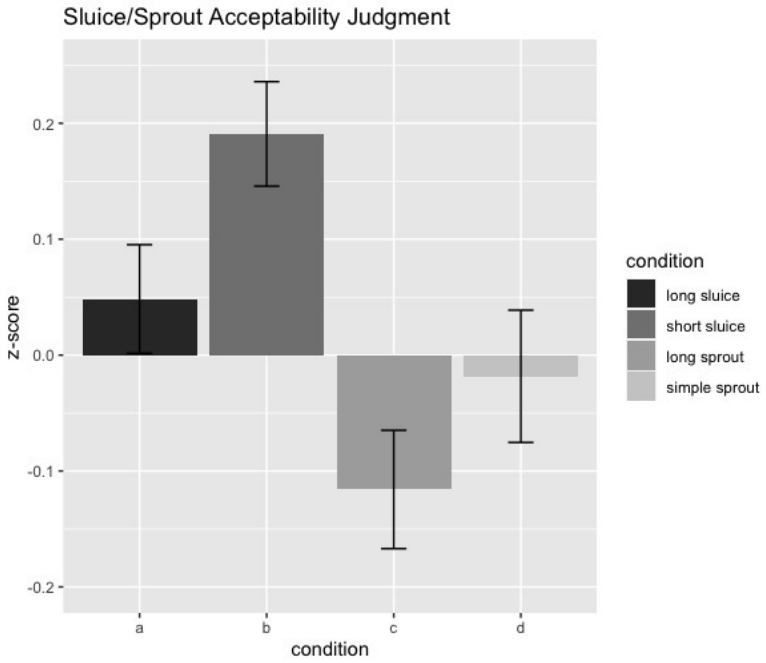


Figure 4:Bar plot of z-scores for Experiments 3a. Condition is on the x-axis, mean z-score is on the y-axis. Error bars represent standard error.

effect	**estimate**	**standard error**	**χ^2 (df)**	**p-value**
antecedent length	**.12**	**.05**	**6.11 (1)**	**.01**
ellipsis type	**-.19**	**.05**	**14.5 (1)**	**.0001**
interaction	-.04	.10	.2 (1)	>.1

Table 14: Estimates, standard error, χ^2, and associated p-values for Experiment 3a. Significant effects noted in bold.

Discussion

The results indicate that inferring an unexpressed adjunct in the elided material during the comprehension of sprouting constructions results in degraded judgments. This result is predicted

by accounts of sprouting which attribute processing difficulty to non-parallelism between the antecedent and the elided material (Dickey & Bunger, 2011).

Interestingly, the results demonstrate an effect such that longer antecedents contribute to the degradation of acceptability of both sluicing and sprouting constructions. This degradation could potentially challenge the predictions of both direct access and cost-free copy models. In a direct access model (Martin and McElree, 2011), the elided material is accessed through a pointer mechanism to a memory representation. Since in this model the memory representation does not subsequently copy structural information into the ellipsis site nor recreate it, length or complexity are not predicted to impact processing. However, an offline measure such as an acceptability rating may not reflect the processes at play during real-time comprehension, and so we are hesitant to make strong claims based on this effect. Furthermore, it is quite likely that longer sentences are *in general* more difficult for comprehenders, regardless of the need for retrieval, although the results regarding sluicing vs. sprouting are indicative that difficulty during initial processing impacts subsequent offline judgements. To more closely inspect the retrieval process occurring when reading these sentences, a finer grained and temporally sensitive tool is required.

Experiment 3b

Experiment 3b probes the results from Experiments 3a by employing an eye-tracking while reading task. If offline results are directly linked to online processes, we predict that we will observe differences in processing due to Ellipsis Type such that sprouting is more difficult than sluicing. Furthermore, given the observance of an effect of Antecedent Length in offline measures, we predict that this effect will surface more strongly in online measures such that the

retrieval of longer antecedents is more costly than the retrieval of shorter antecedents, which will be reflected by longer duration measures at or after the wh-element.

Methods and Materials

Participants

40 undergraduate participants from the Northwestern University Department of Linguistics Subject Pool volunteered to complete this task. Each participant received 1 unit of course credit required by an undergraduate level course in Linguistics. Participants were of normal or corrected-to-normal vision and hearing and were all self-reported native speakers of English. All participants provided informed consent.

Materials

The experimental items used in Experiments 1b were identical to those of Experiments 1a (see (28)) with the addition of a comprehension question following each item. The 24 experimental items were presented with 100 fillers. Lists were pseudo-randomised according to a standard Latin square design.

Procedure

Prior to the experiment, participants had several practice trials with feedback to accustom themselves to the procedure. Eye movements were recorded using a tower mounted Eyelink 1000 eye tracker. Recalibrations were performed as necessary, and participants were frequently offered breaks. Items were pseudorandomised and balanced such that each participant saw only one condition from each item, and at least one filler item occurred between critical trials.

Analysis and Results

Gaze duration was recorded and manually corrected for vertical drift. Fixations less than 80 ms were merged into adjacent fixations, and fixations of more than 2000 ms were excluded from analysis. My analysis of fixations is based on the following 4 measures: first fixation duration, first pass time, regression path duration, and total fixation time. First fixation duration is defined as the duration of the first fixation within a region. First pass time is defined as the sum of all fixations within a region before the first instance of the gaze exiting the region, either to the left or to the right. Regression path duration consists of the sum of all time spent in the region and to the left of the region prior to the first instance of the gaze exiting to the right of the region. Finally, total fixation time is calculated by summing all fixations to a region, including first pass time and re-reading time.

In this study we will focus on three regions of interest: the region containing the wh-remnant *why* (as shown in (41)), the first spillover region containing an adverb such as *specifically,* and the second spillover region containing a conjunction like *although.* All items were displayed on two lines due to length limitations of the presentation software, with the regions of interest always appearing on the second line.

(41) Bill thinks that Mary, for some reason, quit her job, but I don't know **why**$_1$ **specifically**$_2$, **although**$_3$ I hope to find out soon.

Since region 1 (*why)* was always of the same character length, raw fixation values were log-transformed to meet normality assumptions without undergoing residualization to account

for word length effects. Since region 2 varied in character length, raw values were log-transformed, and then used as the dependent variable in a linear model with word length as a predictor. The residuals from the linear model were then subsequently used as the dependent variable for statistical analysis of the spillover region.

In accordance with the discussion in Barr et al (2013), all analyses were conducted by constructing converging maximally inclusive models, and then comparing those models to a model reduced by a single fixed effect. Mean fixation times for each region, condition, and eye tracking measure are provided in Table 15. (β) and standard error (S.E.) were determined from the maximal model. ANOVA was used to compare the maximal model and the reduced model, and to calculate the χ^2 and significance (α = .05) as reported in Table 16.

Ellipsis Type	Antecedent Length	Critical region	Spillover 1	Spillover 2
FIRST FIXATION				
sluice	long	196 (6)	183 (4)	198 (4)
sluice	short	181 (5)	202 (4)	189 (3)
sprout	long	199 (7)	194 (5)	202 (5)
sprout	short	208 (7)	184 (4)	191 (5)
FIRST PASS				
sluice	long	205 (8)	202 (6)	631 (33)
sluice	short	185 (6)	210 (7)	603 (21)
sprout	long	214 (9)	207 (6)	601 (26)
sprout	short	213 (8)	207 (7)	578 (22)
REGRESSION PATH				

sluice	long	318 (39)	318 (23)	1622 (85)
sluice	short	218 (9)	300 (18)	1341 (84)
sprout	long	271 (22)	373 (43)	1899 (137)
sprout	short	283 (23)	327 (28)	1337 (84)
TOTAL TIME				
sluice	long	245 (14)	256 (12)	746 (39)
sluice	short	206 (8)	273 (13)	701 (30)
sprout	long	275 (25)	266 (12)	746 (55)
sprout	short	235 (11)	273 (14)	677 (30)

Table 15:Mean fixation values for each condition, region, and eye tracking measure for Experiments 3b.

Region	Effect	Estimate	Std error	χ^2 (df)	p-value
First Fixation					
Critical	antecedent length	-.02	.03	.48 (1)	>.1
	ellipsis type	**.06**	**.03**	**4.54 (1)**	**.03**
	interaction	**.12**	**.06**	4.1 (1)	**.04**
Spillover 1	antecedent length	-.03	.02	1.53 (1)	>.1
	ellipsis type	.02	.02	1.11 (1)	>.1
	interaction	-.06	.05	1.77 (1)	>.1
Spillover 2	antecedent length	-.05	.02	.56 (1)	>.1
	ellipsis type	-00007	.02	.29 (1)	>.1

	interaction	-.02	.04	.85 (1)	>.1
First Pass					
Critical	antecedent length	-.04	.02	1.2 (1)	>.1
	ellipsis type	.06	.04	2.57 (1)	>.1
	interaction	.11	.07	2.47 (1)	>.1
Spillover 1	antecedent length	0	.03	0 (1)	>.1
	ellipsis type	.006	.03	.05 (1)	>.1
	interaction	-.05	.06	.91 (1)	>.1
Spillover 2	antecedent length	.002	.04	.001 (1)	>.1
	ellipsis type	-.04	.04	.87 (1)	.>.1
	interaction	-.07	.08	.81 (1)	>.1
Regression Path					
Critical	antecedent length	-.08	.05	1.1 (1)	>.1
	ellipsis type	.08	.05	2.06 (1)	>.1
	interaction	**.24**	**.1**	**5.67 (1)**	**.02**
Spillover 1	antecedent length	-.03	.04	.32 (1)	>.1
	ellipsis type	.03	.05	.43 (1)	>.1
	interaction	-.006	.09	.005 (1)	>.1
Spillover 2	**antecedent length**	**-.28**	**.06**	**17.72 (1)**	**.00003**
	ellipsis type	.04	.06	.46 (1)	>.1
	interaction	-.12	.1	1.22 (1)	>.1

Total Time					
Critical	**antecedent length**	**-.1**	**.05**	**4.17 (1)**	**.04**
	ellipsis type	.08	.06	2.13 (1)	>.1
	interaction	.04	.1	.17 (1)	>.1
Spillover 1	antecedent length	.02	.04	.2 (1)	>.1
	ellipsis type	.02	.07	.27 (1)	>.1
	interaction	-.06	.07	.67 (1)	>.1
Spillover 2	antecedent length	-.03	.04	.56 (1)	>.1
	ellipsis type	-.02	.04	.87 (1)	>.1
	interaction	-.07	.08	.8 (1)	>.1

Table 16:Estimates, standard error, χ^2 values, and p-values for Experiments 3b.

Intercepts were permitted to vary across subjects and items. Slopes of Antecedent Length, Ellipsis Type, and the interaction between these factors, were also allowed to vary across subject and item. If the maximal model failed to converge, the random effects structure was simplified according to Baayen (2008). Factors were centered prior to analysis for this and all subsequent experiments in this study. To determine significance, the maximal converging model was compared using ANOVA to a reduced model with one term of interest removed.

In the critical region, e.g. *why*, we found a significant main effect of Ellipsis Type in first fixation duration ((β = ..06, S.E. = .03, χ^2 = 4.54 (1), p = .03), such that sprouting contexts elicited longer first fixations at the wh-region than sluicing contexts. This result corroborates previous

research which indicates that sprouting incurs a processing cost due to the added burden of inferring an unexpressed correlate of a wh-phrase. In addition, a significant effect of Antecedent Length was observed in total fixation duration ((β = -.1, S.E. = .05, χ^2 = 4.17 (1), p = .04), such that conditions with longer (bi-clausal) antecedents resulted in increased total fixation duration as compared to conditions with shorter (mono-clausal) antecedents.

Interactions were also observed at the critical region. A significant interaction was seen in first fixation duration (β = .12, S.E. = .06 χ^2 = 4.1 (1), p = .04), and a significant interaction was seen in regression path duration (β = .24, S.E. = .1, χ^2 = 5.67 (1), p = .02). Pairwise comparisons revealed that the simple sluice condition elicited significantly shorter first fixation durations and regression path durations than all other conditions. A likely explanation for this interaction is that the penalty imposed by sprouting is obscuring a length effect in the sprouting conditions, so an effect of Antecedent Length is only seen in the sluicing contexts.

No significant effects were observed in the first spillover region, however a robust effect of Antecedent Length was surfaced in regression path duration at the second spillover region (β = -.28, S.E. = .06, χ^2 = 17.72 (1), p = .00003). Figure 5 below provides bar plots of log transformed fixation measures of statistically significant measures.

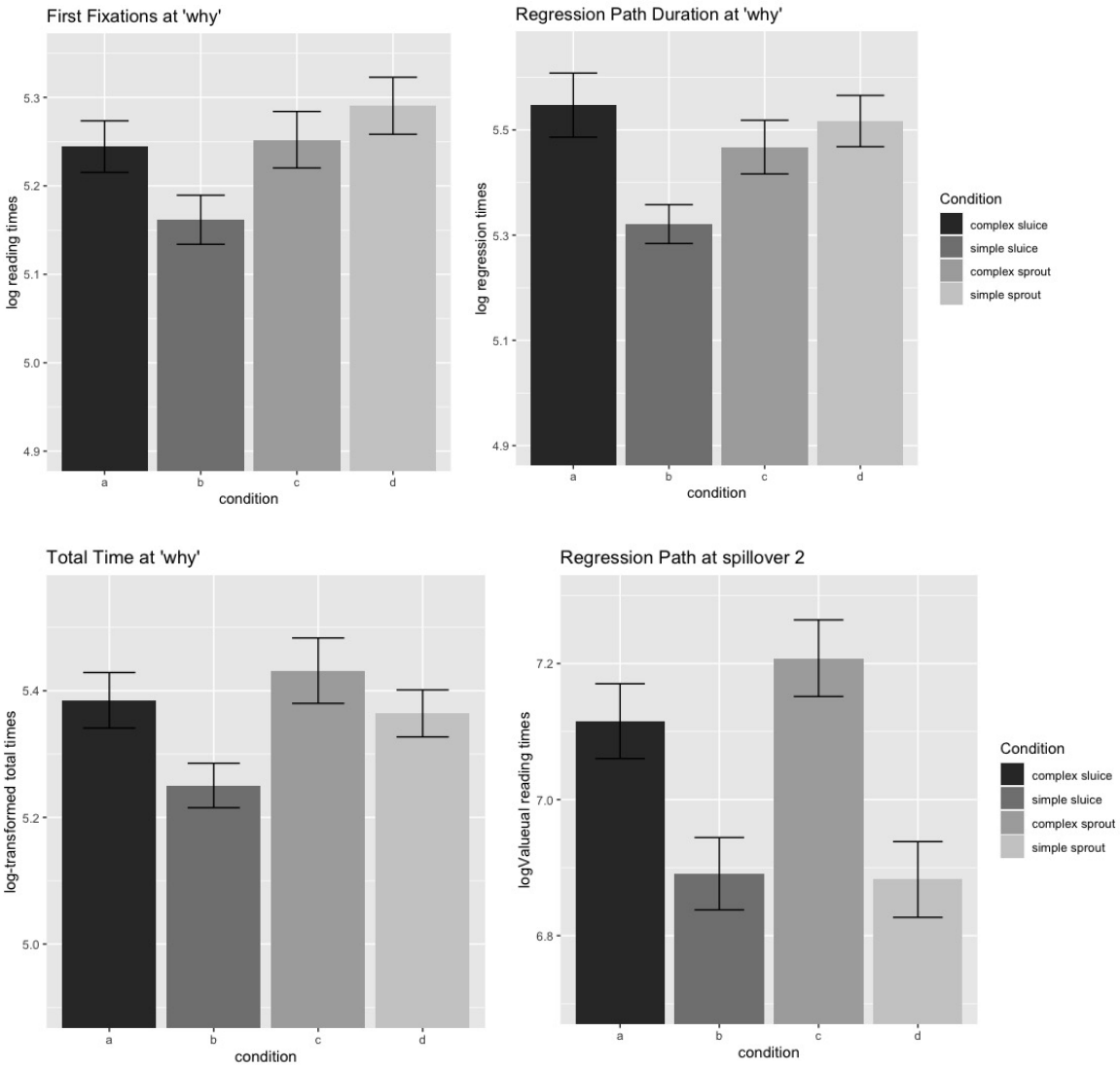


Figure 5:Bar plots for significant measures for Experiments 1b. Error bars represent standard error.

Discussion

The results of the eye tracking experiment reveal that both Ellipsis Type and Antecedent Length affect the online processing of the ellipsis site. Regarding Ellipsis Type, this result corroborates the effect witnessed in the acceptability judgment experiment as well as previous studies which seem to observe a penalty for sprouting constructions. The result supports the idea that non-isomorphism between the antecedent and the interpretation of the ellipsis site causes some

difficulty for the comprehender – in the case of sprouting that difficulty is likely related to the cost associated with having to build additional structure to accommodate a previously absent element (in this instance an adjunct) (Dickey & Bunger, 2011). While this result was unsurprising, it provides a valuable replication of previous work and may contribute to the debate over the presence or absence of structural material during the interpretation of ellipsis. It also affirms the claim the "for some reason" is indeed serve as the correlate for the wh-phrase "why".

The more novel finding from Experiments 3b is a reliable effect of Antecedent Length, such that longer antecedents appear to result in difficulty in processing material at or after the ellipsis site, as indexed by longer durations in reading times. At the critical region, "why", this effect emerges as a main effect of total time fixating on the region as well as solely in conditions which use the sluicing construction for regression path duration. This pattern of results may suggest that there is broadly some difficulty with longer antecedents (as indexed by total time), but that time spent fixating on previous material serves to both access the additional content of the longer antecedents as well as amend the recovered antecedent to allow for a source of the wh-element. This idea may also be supported by the strong effect of Antecedent Length at the second spillover region which does not seem to be impacted by Ellipsis Type. It may be that the recognition of sprouting and the ensuing changes to the structural content of the antecedent is completed before the parser moves on to subsequent material, whereas the difficulty associated with the retrieval and integration of other material lingers. However, the lack of an antecedent length effect at the first spillover region and its reappearance at the second spillover region may link the regression path duration result at "although" to integration difficulties for longer antecedents.

The interpretation of these results indicate that antecedent length has a role to play during the retrieval process, which runs contrary to the predictions of certain direct access models currently assumed in the sentence processing field (Lewis et al., 2006; Martin & McElree, 2008, 2011; Paape et al., 2017). When considered along with the effect of sprouting, a view of retrieval that includes sensitivity and respect for structural information becomes more sharply focused. However, the view afforded by these experiments would benefit from additional support from other experimental manipulations. To this end, minimal changes to the stimuli from Experiments 3a-b allow for an exploration of how retrieval may interact with island constraints, which is presented in Experiments 4a-b.

Experiment 4a

Experiment 4a builds off the materials used in Experiments 3a-b to manipulate whether retrieval of a biclausal antecedent at the ellipsis site creates a factive island violation. As in Experiments 1a, the length of the potential antecedent is alternated to further probe if the retrieval of longer antecedents causes more processing difficulty as opposed to shorter antecedents. As detailed in the background section, factive verbs presuppose the truth of their embedded complement, and extraction of an adjunct phrase out of the embedded complement is relatively dispreferred. This can be a fruitful area for investigation as if we observe different patterns of processing due to antecedent complexity between the island and non-island conditions, then we may be able to make certain inferences about how the retrieval process is respecting structural considerations such as islands during antecedent retrieval and ellipsis resolution. As in Experiments 3a, this experiment employs an acceptability rating task, with a 2x2 design with Wh-Type (adjunct vs. argument) and Antecedent Length (long vs. short) as factors. A sample set of the stimuli is

summarised in (42). While this type of offline task may or may not directly reflect processing difficulty, we employ it here to probe whether any lingering effects of online processing affect the final interpretation, and to see if any overall biases are observed due to Wh-Type.

Given the nature of this offline task, we may or may not see any lingering effects of online processes, although the results of Experiments 1a suggest that such processes may impact acceptability judgments even when all final interpretations are grammatical. If effects are observed, however, the clearest prediction is that length of the antecedent will affect processing, at least in the cases where there is *not* potential for an island violation (e.g. argument extraction). However, in the case of adjunct extraction, contrasting predictions can be made. One possibility is that in the case of adjunct extraction when there is a factive matrix verb (e.g. 42a), the parser retrieves the maximally long antecedent, and the ensuing interpretation violates a factive island. This presumably would result in degraded judgments. In contrast, a final interpretation that does not cause an island violation can be achieved if only the embedded clause is retrieved. If this is the ultimate interpretation of cases like (42a), then we do not predict a difference between (42a-b).

Methods and Materials

Participants

41 self-reported native English speakers were recruited on Amazon Mechanical Turk. Subjects received $2.00 in compensation, and all provided informed consent prior to completing the experiment.

Materials

The stimuli for this experiment consisted of 24 experimental items, and 72 fillers. The independent factors of WH-Type and Antecedent Length were crossed, such that each item had 4 conditions, as shown in (42).

(42)

Adjunct/Long

a. Bill hated that Mary, for some reason, bought clothes, and somebody should know why specifically, although they are keeping it a secret.

Argument/Long

b. Bill hated that Mary, in the morning, bought something, and somebody should know what specifically, although they are keeping it a secret.

Adjunct/Short

c. Mary, for some reason, bought clothes, and somebody should know why specifically, although they are keeping it a secret.

Argument/Short

d. Mary, in the morning, bought clothes, and somebody should know what specifically, although they are keeping it a secret.

4 lists were pseudo-randomized and counterbalanced, such that each participant saw only one condition from each item, and experimental items were separated by a minimum of one filler. Filler items included sentences manipulating verb phrase ellipsis, quantifier raising, and

reflexive pronoun resolution, however these fillers did not involve wh-filler/gap dependencies. The experimental items can be found in the Appendix.

Procedure

As in previous acceptability judgment experiments, upon viewing the experiment, participants were instructed that they would be completing a questionnaire in which they would be asked to judge how natural a sentence sounded. Subjects were presented with two example sentences prior to beginning the experiment, which demonstrated a generally acceptable and generally unacceptable English sentence. All participants viewed identical example sentences. Participants were instructed to rate each sentence on a scale of 1 (unnatural) to 7 (natural) based on their intuitions.

Analysis and Results

Following the analysis used in my previous experiments, raw judgments of the experimental items were converted to z-scores within participants, and all subsequent analyses were based on these scores.

Linear mixed effects models were used to analyze the z-scored ratings. The model was fit with a maximal random effects structure (Barr et al., 2013), and item random effects were eliminated until the model converged. The model was run using the lme4 package in R (Bates et al., 2015), and p-values for the fixed and random effects were evaluated using model comparison.

The average z-score ratings for each condition of the experimental items are shown in Figure 6, and Table 17 presents the estimated coefficients and standard error with from the

regression model with associated p-values obtained through model comparison. A main effect of Antecedent Length is observed (β = .62, S.E. = .08, χ^2 = 34.29 (1), p <.00001), such that sentences with shorter antecedents were judged as more natural than those with longer antecedents (mean z-scores .32 vs. -.31, respectively). A marginally significant main effect of WH-Type was also observed (β = -.15, S.E. = .08, χ^2 = 3.56 (1), p = .06), such that conditions involving adjunct extraction were judged to be more natural than those with argument extraction (mean z-scores .08 vs. -.08, respectively).

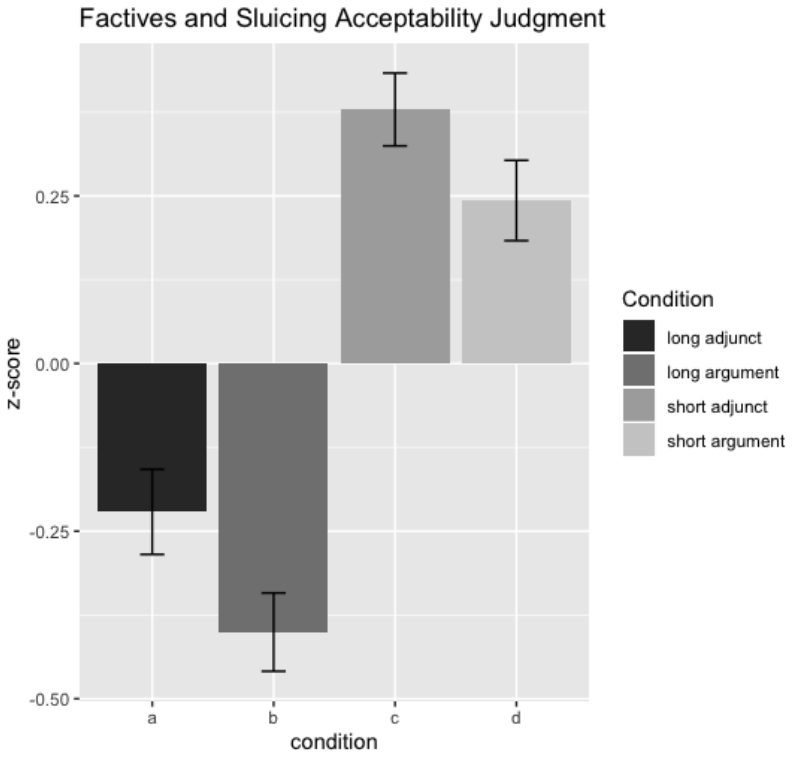


Figure 6:Bar plot of z-scores for Experiment 4a. Condition is on the x-axis, mean z-score is on the y-axis.

effect	**estimate**	**standard error**	**χ^2 (df)**	**p-value**

antecedent length	**.62**	**.08**	**34.29 (1)**	**< .000001**
wh- type	**-.15**	**.08**	**3.56 (1)**	**.06**
interaction	.02	.13	.04 (1)	>.1

Table 17:Estimates, standard error, χ^2, and associated p-values for Experiments 4a. Significant and near significant effects noted in bold.

Discussion

These results strongly indicate that the length of the retrieved material is significantly impacting the acceptability of the presented sentences. Given the results of Experiments 3a, this was not a wholly unexpected result, however the strength of the effect was somewhat surprising. Since the relevant difference between Experiments 3a and 4a is the presence of factive matrix verbs in the long antecedent conditions, one possibility is that factivity of the main verb encourages retrieval of the entire biclausal antecedent, resulting in a more notable length effect in Experiment 4a. This kind of explanation is in line with informal observations that it is "easier" to achieve a reading of the sentences in the long antecedent conditions in which the ellipsis site is interpreted as containing (1) the previously processed biclausal structure and (2) a gap in the adjunctival phrase of the embedded clause. However, if there is a preference for maximal retrieval in the long antecedent conditions, we might expect to see a penalty associated with the resulting factive island violation in the long adjunct condition (condition a), which does not appear to be observed.

Additionally, the main effect of WH-Type such that *adjuncts* were judged as more natural than arguments was an unexpected result. If an effect was to be witnessed, the prediction was that the long adjunct conditions should be dispreferred as they present the potential for an island

violation. However, not only is there a lack of an interaction, but the main effect runs in the contrary direction. It is possible that an explanation lies in the availability of other interpretations of the elided material, and their lingering effects in offline measures. In the adjunct conditions, there exist other potential attachment sites for the adjunct phrase that are not available for the argument. For example, in the long adjunct condition (reproduced below in (43)), although the wh-correlate "for some reason" is located in the embedded clause, it is possible to "sprout" the adjunct from the matrix clause. This interpretation does not result in a factive island violation. In contrast, in the long argument condition, the wh-element can only originate from the embedded clause. We surmise that given other documented effects of lingering parses (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013; Van Gompel, Pickering, Pearson, & Jacob, 2006), there is a possibility that even when an ungrammatical interpretation is forced during the resolution of ellipsis with "why", acceptability could arise due to lingering effects of previous parses.

(43)

Adjunct/Long

Bill hated that Mary, for some reason, bought clothes, and somebody should know why specifically, although they are keeping it a secret.

Argument/Long

Bill hated that Mary, in the morning, bought something, and somebody should know what specifically, although they are keeping it a secret.

The overall results of Experiment 4a leaves us with some conclusions as well as questions. The striking length effect certainly points to structural complexity having an observable effect on the interpretation of ellipsis in my materials, however the lack of an effect of the potential factive island and the observed preference for adjunct extraction require more investigation. An online task is performed in Experiment 4b to probe the temporal dynamics of the processing of the sentences from Experiment 4a.

Experiment 4b

An eye tracking while reading task in employed in Experiment 4b to determine whether the results witnessed in Experiment 4a are directly tied to the processing of the ellipsis site, and to see whether we can observe evidence that factivity of the matrix verb in the longer conditions affects the retrieval process in some way. Following the observations of Experiments 3b, we minimally predict that a length effect will surface at or after the wh-element in the conditions in which an argument is extracted, however given the results of Experiment 4a it may be that a main effect regardless of wh-type will surface. However, if factivity is relevant to the parser during the ellipsis resolution process, then two conflicting sets of predictions can be tested. One possibility is that the entire biclausal antecedent will be retrieved, resulting in a factive island violation, and this violation may subsequently result in reanalysis of the material such that the ultimate interpretation consists of retrieval of only the embedded clause. Such a sequence of events would most likely result in longer processing times at or after the wh-element that may be indistinguishable from a simple length effect, but may surface in an additive fashion such that the long adjunct conditions demonstrate longer reading times than the long argument condition. The

alternative is that the parser is able to access information about factivity of the matrix verb as well as the structural source of the wh-element during the evaluation of potential antecedents, recognise the potential for a violation, and pre-emptively avoid retrieval of the biclausal antecedent. In a model where antecedent candidates are evaluated in parallel this should result in an interaction such that the long adjunct condition patterns with the short conditions, and only the long argument condition elicits longer reading times.

Methods and Materials

Participants

41 undergraduate participants from the Northwestern University Department of Linguistics Subject Pool volunteered to complete this task. Each participant received 1 unit of course credit required by an undergraduate level course in Linguistics. Participants were of normal or corrected-to-normal vision and hearing, and were all self-reported native speakers of English. All participants provided informed consent.

Materials

The experimental materials used in Experiment 2b were identical to those of Experiment 4a (see (42)) with the addition of a comprehension question following each item. The 24 experimental items were presented with 100. Lists were pseudo-randomised according to a standard Latin square design.

Procedure

Procedure was identical to Experiments 3b.

Analysis and Results

Treatment of the raw eye tracking data was identical to that of Experiments 1b. In this study we will focus on three regions of interest: the region containing the wh-remnant *why* or *what* and the spillover regions containing an adverb such as *specifically* and a word such as *although* to continue the sentence, shown in (44). All items were displayed on two lines due to length limitations of the presentation software, with the regions of interest always appearing on the second line.

(44) Bill hated that Mary, for some reason, quit her job, but I don't know **why$_1$ specifically$_2$**, **although$_3$** I hope to find out soon.

As in previous eye tracking experiments, raw fixation values were log-transformed to meet normality assumptions. Since region 1 differed in length due to the manipulation of Wh-Type (*why* vs. *what)*, log-transformed values were then used as the dependent variable in a linear regression model with word length as a predictor, and the resulting residuals were used as the dependent measure in subsequent mixed effect regression models. Although the spillover regions did not differ in length across conditions, the residualized reading times were also used in these models to maintain consistency.

Analyses and model comparisons were conducted in an identical way as in previous experiments and in accordance with the discussion in Barr et al (2013). Mean raw fixation values for each eye tracking measure and region are reported in Table 18. Intercepts (β) and standard error (S.E.) were determined from the maximal model, and ANOVA was used to compare the

maximal model and the reduced model, and to calculate the χ^2 and significance (α = .05) as reported in Table 19 .

Antecedent	Wh-type	Critical region	Spillover 1	Spillover 2
FIRST FIXATION				
long factive	adjunct	191 (9)	200 (9)	221 (12)
long factive	argument	202 (14)	189 (7)	207 (6)
short non-factive	adjunct	203 (9)	194 (6)	218 (7)
short non-factive	argument	206 (10)	196 (7)	217 (8)
FIRST PASS				
long factive	adjunct	198 (10)	225 (12)	258 (17)
long factive	argument	217 (16)	216 (10)	229 (8)
short non-factive	adjunct	222 (11)	219 (9)	242 (11)
short non-factive	argument	220 (12)	216 (9)	246 (12)
REGRESSION PATH				
long factive	adjunct	405 (46)	404 (69)	519 (85)
long factive	argument	656 (211)	958 (183)	651 (127)
short non-factive	adjunct	427 (61)	449 (57)	492 (90)
short non-factive	argument	396 (49)	350 (35)	593 (108)
TOTAL TIME				
long factive	adjunct	318 (20)	337 (19)	394 (19)
long factive	argument	390 (23)	370 (19)	393 (23)
short non-factive	adjunct	318 (16)	333 (17)	379 (23)
short non-factive	argument	365 (22)	362 (19)	429 (23)

Table 18: Mean fixation values for each condition, region, and eye tracking measure for Experiment 4b.

Region	Effect	Estimate	Std error	χ^2 (df)	p-value
First Fixation					
Critical	antecedent length	.07	.04	2.02 (1)	>.1
	wh-type	-.003	.05	.006 (1)	>.1
	interaction	.03	.09	.09 (1)	>.1
Spillover 1	antecedent length	.03	.04	.07 (1)	>.1
	wh-type	-.01	.03	1.33 (1)	>.1
	interaction	-.007	.07	.79 (1)	>.1
Spillover 2	antecedent length	.03	.04	.83 (1)	>.1
	wh-type	-.01	.03	.08 (1)	>.1
	interaction	-.007	.07	.01 (1)	>.1
First Pass					
Critical	antecedent length	.08	.05	2.17 (1)	>.1
	wh-type	0	.05	0 (1)	>.1
	interaction	.06	.1	.29 (1)	>.1
Spillover 1	antecedent length	-.003	.04	.002 (1)	>.1
	wh-type	.03	.04	.87 (1)	>.1
	interaction	-.04	.1	.21 (1)	>.1
Spillover 2	antecedent length	.02	.04	.36 (1)	>.1

	wh-type	-.004	.04	.02 (1)	.>.1
	interaction	-.04	.09	.2 (1)	>.1
Regression Path					
Critical	antecedent length	-.04	.09	.2 (1)	>.1
	wh-type	-.07	.1	.49 (1)	>.1
	interaction	.19	.17	1.2 (1)	.>.1
Spillover 1	**antecedent length**	**-.17**	**.1**	**3.14 (1)**	**.07**
	wh-type	-.16	.1	2.5 (1)	>.1
	interaction	**.48**	**.19**	**5.39**	**.02**
Spillover 2	antecedent length	-.02	.11	.02 (1)	>.1
	wh-type	-.08	.1	.67 (1)	>.1
	interaction	-.09	.21	.16 (1)	>.1
Total Time					
Critical	antecedent length	.005	.05	.01 (1)	>.1
	wh-type	**-.12**	**.05**	**5.87 (1)**	**.02**
	interaction	.12	.1	1.37 (1)	>.1
Spillover 1	antecedent length	.02	.05	.12 (1)	>.1
	wh-type	**-.08**	**.04**	**3.34 (1)**	**.06**
	interaction	.05	.09	.28 (1)	>.1
Spillover 2	antecedent length	.05	.05	1.28 (1)	>.1
	wh-type	-.03	.04	.61 (1)	>.1

	interaction	-.1	.1	.9 (1)	>.1

Table 19:Estimates, standard error, χ^2 values, and p-values for Experiment 4b.

A significant main effect of Wh-Type was found in total fixation time at the critical region (β = -.12, S.E. = .05, χ^2 = 5.87 (1), p = .02), such that sentences with wh-extraction out of an adjunct resulted in shorter total times than conditions with arguments.

At the first spillover region, a significant interaction was observed in regression path duration (β = .48, S.E. = .19, χ^2 = 5.39 (1), p = .02) such that length of the antecedent affected reading times in the argument conditions, however did not have an effect in the adjunct conditions. A main effect of Antecedent Length approached significance in regression path durations (β = -.17, S.E. = .1, χ^2 = 3.14 (1), p = .07) such that longer antecedents resulted in longer regression path durations, and there additionally was a marginal effect of Wh-Type in total fixation time (β = -.08, S.E. = .04, χ^2 = 3.34 (1), p = .06) at the first spillover region. No significant effects were observed at any subsequent region. Bar plots of significant measures are provided below in Figure 7:

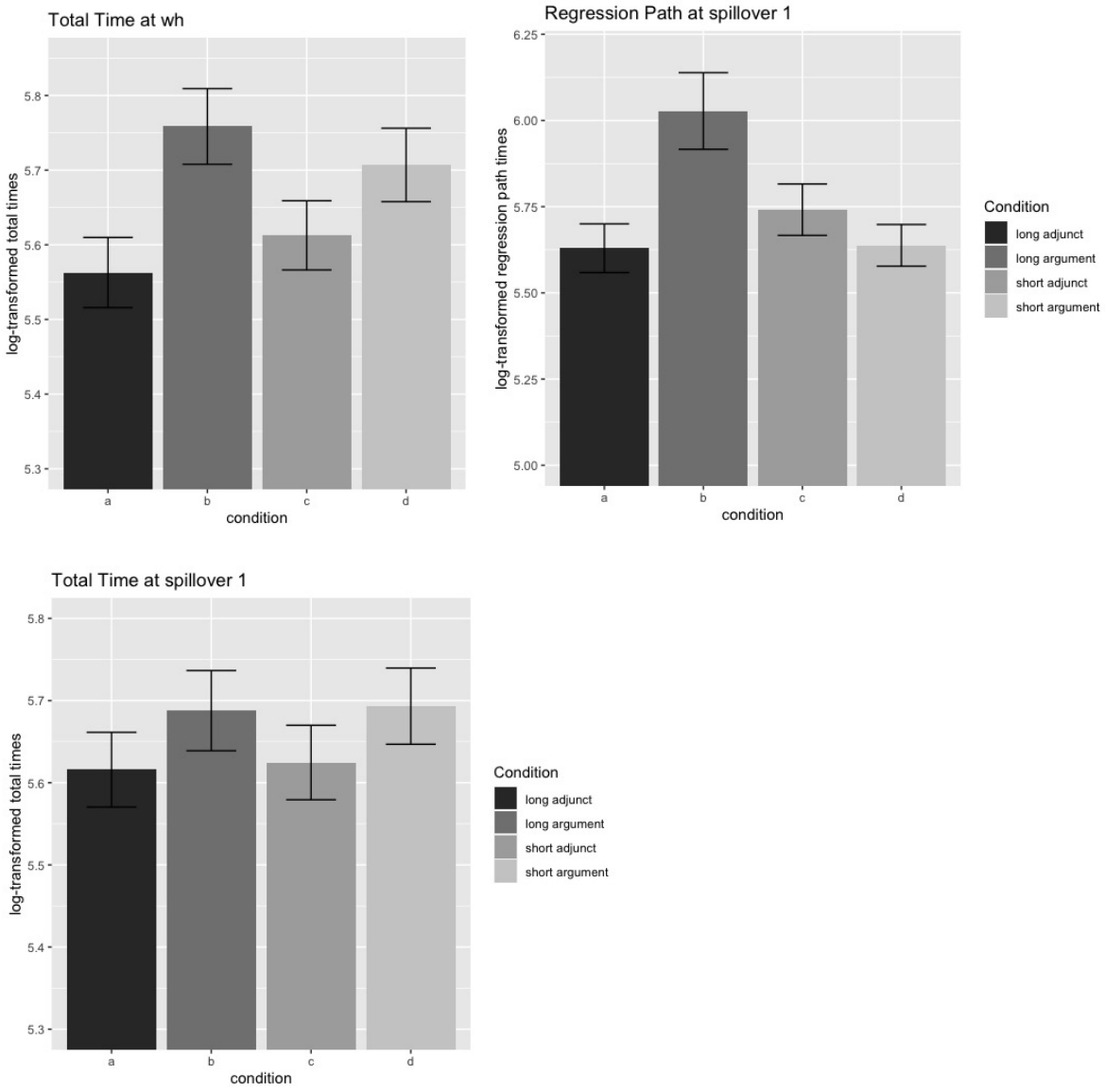


Figure 7:Bar plots of significant measure in Experiment 4b. Error bars represent standard error.

Discussion

The results of Experiment 4b are intriguing, although perhaps somewhat less clear than those from previous experiments. The finding that adjunct extraction appeared to facilitate processing compared to argument extraction was less surprising considering the results of the associated

acceptability judgment experiment, however still unexpected from a theoretical standpoint. If anything, one might expect that the possibility of an island violation in the long adjunct conditions might drive the *opposite* effect, such that the condition(s) with argument extraction would show an advantage over those with adjuncts. In Experiment 4a, we suggested that a lingering (positive) effect of a grammatical parses may persist even when an ungrammatical effect is ultimately forced, however it is not apparent how this kind of explanation would facilitate incremental processing. Furthermore, it does not explain the fact that a main effect of WH-Type is witnessed rather than an interaction in which only the long conditions are affected.

The more intriguing result from Experiment 4b is the observance of the interaction in regression path duration. The data demonstrate that subjects had more difficulty processing the long argument condition as compared to the short argument condition, but that the adjunct conditions showed similar profiles. In addition, the long adjunct condition appeared indistinguishable from both of the short conditions in terms of regression path duration. This result is informative in a number of ways. First, it appears that length of the antecedent is affecting processing during the resolution of ellipsis in the argument conditions, which confirms the findings and supports the conclusions of Experiments 3b. Second, and most interestingly, the lack of an effect in the adjunct conditions and the speed of the long adjunct condition supports the prediction that only the embedded clause is retrieved during resolution in the long adjunct condition. As noted, if reanalysis were occurring rather than initial retrieval of only the embedded clause, then a processing penalty reflected by longer fixation times would be expected to occur, which does not appear to be the case.

This result has implications for how we think of retrieval, and what kind of information is relevant during the retrieval process. For the parser to be able to avoid an island violation (rather

than *fix* an island violation), information about the semantic qualities of the matrix verb as well as the structural location of the extracted wh-element must be available during the retrieval process. More concretely, the parser must recognize that an element internal to the antecedent is factive, that there will be a gap in an adjunct clause of this antecedent, and furthermore that this combination is ungrammatical. Current cue-based theories may be able to accommodate the factivity of an internal element, but it is difficult to imagine what precisely this cue would consist of. The simplest answer is something along the lines of [+/- factive], but what then exactly does that cue do or require? Does it require something like the truth value of its complement, and if so is that now also represented featurally? When would a cue like [+/- factive] aid in retrieval or memory access? Furthermore the structural location of an wh-element and its relation to the factive verb may be difficult to imagine considering that the probe cues provided by 'why' will target a TP that itself does not provide information about the features associated with the VP 'internal' to it. It is possible that this difficult to imagine interaction in memory between matrix verb information and structure could be circumvented altogether if we assume that the position of the adjunct in the antecedent sets the ultimate position of "why" in the ellipsis site, an idea employing the notion of scope parallelism. This proposal is explored further in the General Discussion.

General Discussion

In this research, two acceptability judgment experiments and two eye tracking experiments sought to illuminate the retrieval process by investigating the effects of antecedent structure and factive islands during the resolution of ellipsis. In Experiments 3a-b, length of a potential antecedent and ellipsis type were manipulated, resulting in the findings that longer antecedents

appeared to cause difficulty for comprehenders, in contrast to predictions made by variants of direct-access models that assume little or no impact of structural information on the processing of ellipsis. Additionally, sprouting type ellipsis constructions appeared to disrupt processing of the wh-element, confirming previous research that found a penalty associated with having to infer the source of a wh-element when a correlate was not found in the antecedent (Dickey & Bunger, 2011). In Experiments 4a-b, the stimuli from Experiments 3a-b were altered to ensure that the matrix verb in the long antecedent conditions was factive, which allowed for an investigation of how island constraints may interact with the retrieval process. The potential for an island violation appeared to impact online processing such that the parser "avoided" the retrieval of a violation inducing antecedent. Finally, there was some evidence that argument extraction was dispreferred compared to adjunct extraction, although the source of this difference was somewhat unclear.

The most notable observation from the experiments of my work is that complexity of an antecedent appears to have a significant impact on the parser and the retrieval process during the processing of the ellipsis site. The relationship between antecedent structure and comprehension is such that sentences with longer antecedents are judged more harshly by comprehenders, and also cause more relative difficulty during the resolution of ellipsis. This observation directly contradicts the predictions of retrieval models that employ a direct access cue-based pointer system to resolve dependencies. One of the hallmarks of a direct access system such as that proposed by Martin and McElree (2008) is that retrieval times are unaffected by the length or complexity of an antecedent, as pre-existing memory structures are "linked" to the dependent element rather than copied or recreated at the point of interpretation. However, myresults indicate that an increase in structural complexity results in an increase in readings times, which is

incompatible with an account in which the ellipsis site contains merely a pointer to a stored value in memory.

Furthermore, the results in this paper suggest that even if we conclude that structural information is present in the memory representation of the antecedent, that information is not being "shared" between two elements in the sense suggested by Frazier and Clifton (2001).While the predictions of the pointer account and the cost-free copy account are essentially equivalent (as pointed out by Martin and McElree), the results presented in this work minimally indicate that if copying is taking place, that operation becomes more costly as the material to be copied is increased. However, given observations from Hall and Yoshida (2020) in which a structural relation between two elements in the memory is referenced during the resolution of pronouns, the copying account at least provides a means of reconstructing structure within the ellipsis site. However, to reiterate, my results directly contradict an account where the structural material is shared directly between the ellipsis site and a pre-existing memory representation, and at the very least indicate that if structure is copied then copying becomes more costly as the amount of material is increased.

Other evidence certainly exists that structural information is present in the ellipsis site, for example in connectivity effects (Merchant, 2001; Ross, 1967; Yoshida et al., 2019) or island sensitivity (Johnson, 2001b; Ross, 1967), however the observations from this study are the first recent results that show a correlation between antecedent complexity and the speed of ellipsis resolution. I note that recent investigations of this topic have found contrary results to the ones presented here, namely work performed by Paape and colleagues (Paape, Hemforth, & Vasishth, 2018; Paape et al., 2017), who do not observe an effect of structural complexity of the antecedent on ellipsis resolution. There does not appear to be an obvious reason why I have opposing

results, however it may be related to the temporal sensitivity of the experimental paradigm (self-paced reading vs. eye tracking). As opposed to the consideration of structural *relations*, it is not difficult to represent structural identity of items in a cue-based system, however the means by which a detailed hierarchical structure arrives at the ellipsis site is murkier.

What then are the alternatives? Some accounts, including those that presume that the relevant information for matching elided material and an antecedent is semantic identity, would require that syntactic information be reconstructed from the semantic relation between the antecedent and ellipsis in order to account for the evidence of structure in the ellipsis site. In such scenarios the structure may need to be built up inference by inference, although it is also possible to imagine that the identified antecedent is re-parsed. However, these kinds of accounts are questionable for several reasons, as noted among others by Frazier (2018). One objection is that not only would it be difficult to infer the syntactic information of the antecedent in the first place, but if the ultimate goal of sentence comprehension is to achieve the interpretation of a sentence (Frazier, 2018), and structure is a means to semantic interpretation (Beck & Tiemann, 2018; Kratzer & Heim, 1998; Larson, 1995; Partee, 2014), then an effort to determine the internal structure of an antecedent if a semantic interpretation has already been achieved seems to be at odds with the principle of economy. Although these accounts may not appear parsimonious, they would easily be able to account for the length effects witnessed in this study - as more material needs to be inferred or re-parsed duration of processing increases.

A semantic account, however, would run into difficulty when attempting to account for the results of Experiment 2b, where we see a *lack* of a length effect in the condition involving long potential antecedents and adjunct extraction. If the syntactic structure is inferred from the semantic identity of the antecedent, presumably the structural location of the extracted element

would not be clear until after the structure is inferred. In this scenario, we might expect to see added difficulty in the long adjunct condition as opposed to the lack of a length effect that is in fact witnessed.

The results from the factive island experiments are curious, and novel as far as I know. I have framed these results as evidence that the parser is "avoiding" retrieving the longer antecedent so that a violation does not occur, but what precisely does this avoidance entail? It is the combination of factivity and structural location[14] of the wh-element that has been extracted which induces the violation, which taken together seems a difficult thing to represent as a set of cues in a feature matching system. However, factivity itself could easily be encoded as a binary feature, and this combined with a post retrieval process that determines the source and path of wh-movement could account for the pattern of results demonstrated in Experiment 2b. It is possible that in the computation of the source and path of the wh-phrase, factivity blocks upwards movement of the wh-phrase from the embedded to the matrix CP.

The type of account outlined above, however, suffers greatly from running contrary to intuitions and evidence that the parser operates in an incremental and eager fashion (Aoshima et al., 2004; Staub & Clifton Jr, 2006; Stowe, 1986). Additionally, it attributes the processing costs we see to post-retrieval processes. A more plausible scenario may be that upon encountering the wh-element and retrieving the elided material, the parser attempts to complete the dependency opened by "why" and close it as soon as possible. The structurally closest option to complete the

[14] I would also point out that regardless of the source of island effects, the evidence from my experiments indicates that antecedent complexity has an effect on processing of the ellipsis site.

dependency (in the long adjunct conditions) is to sprout the adjunct from the factive matrix verb, shown below in Figure 8[15].

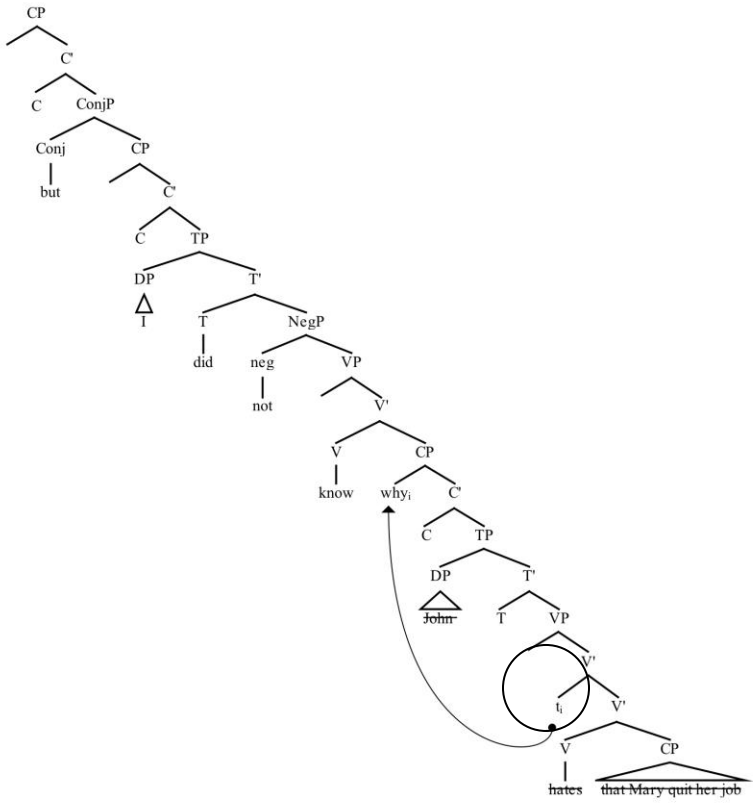


Figure 8:Tree demonstrating movement of wh-phrase from adjunct of the matrix VP to matrix spec CP. The circle indicates the structural position that is not present in the antecedent, but has sprouted in the elided structure.

This explanation neatly explains the difference between the long adjunct and long argument conditions in Experiment 4b: the dependency in the adjunct condition is completed quickly and processing is subsequently faster, whereas the dependency in the argument condition

[15] Trees made using Treeform (Derrick & Archambault, 2010)

has no sprouting option and must wait to be completed until the object position of the embedded clause. However, this explanation ultimately fails as well when comparing Experiments 3 and 4. In Experiment 3, we see an effect of length between long and short adjunct conditions, even though the same type of early dependency closure is possible in the non-factive long adjunct conditions. Furthermore, Experiment 3 reaffirms claims that sprouting is costly, and no such sprouting penalty appears for the long adjunct conditions of Experiment 4.

It is worth addressing the possibility that the observations I have made regarding antecedent length and the retrieval process are in fact a simple result of ambiguity the parser faces when determining the interpretation of ellipsis. In the long antecedent conditions of both Experiments 3 and 4, retrieval of just the embedded clause is perfectly grammatical, and interpreting the wh-element as initially originating in the matrix clause is also possible, at least in the adjunct extraction cases. There are reasons to doubt such an account, however. First, there does not appear to be convincing previous evidence that ambiguity affects processing adversely, and in fact some research has notably reported that ambiguity produces the opposite effect, such that it *facilitates* processing (Traxler, Pickering, & Clifton Jr, 1998; Van Gompel, Pickering, Pearson, & Liversedge, 2005; Van Gompel, Pickering, & Traxler, 2001). Furthermore, ambiguity as a driving force behind the observed results in particular fails to account for the results of Experiment 2b. It is the long argument condition as opposed to the long adjunct condition which elicits longer fixation measures, however it is the long adjunct condition which presents more ambiguity.

In consideration of the problematic accounts above, it again seems that the most parsimonious hypothesis is that only the embedded clause is retrieved factive island condition of Experiment 4. A possible route to understanding how this might be achieved could involve a

complex interaction between the scope the adjunct is able to take in the relevant sentences, the complexity of the antecedent specified by such scope, and the structural complexity of the antecedent that is ultimately retrieved. It may be that an adjunct phrase like “for some reason” wants to raise in a way similar to quantifier raising, but that this raising is blocked by a factive verb. For example, in a standard X-bar theory tree diagram in Figure 9 , the adjunct “for some reason” is able to raise to a position higher than the matrix verb “believes”. When the matrix verb is factive, however, this covert movement is blocked. Evidence for this kind of scope interaction is more apparent when interpreting sentences with quantifier raising, for example consider the contrast between 45a and 45b: a collective reading (such that there is a single reason that Mary quit her job that everyone hates/believes) in which the adjunct “for some reason” scopes over the quantifier seems easier to achieve in 45a, which has a non-factive verb, compared to 45b, which has a factive verb.

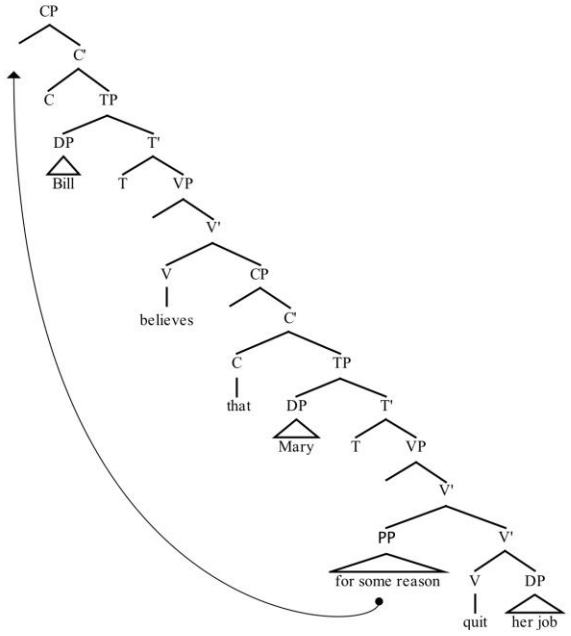


Figure 9:Tree diagram of the sentence " Bill believes that Mary, for some reason, quit her job". Line indicates possible movement to specifier position of the matrix CP. This movement is hypothesized to be blocked when the matrix verb is factive.

(45)

a. Everyone believes that Mary, for some reason, quit her job.

b. Everyone hates that Mary, for some reason, quit her job.

Thus, before recognition of the ellipsis site is achieved, the height of the adjunct in the antecedent has already been determined. Referring to scope parallelism (Chung et al., 1995; Johnson, 2001a; Merchant, 2001; Romero, 1998; Sag, 1976; Williams, 1977), it has been observed that in the interaction of ambiguous scope and ellipsis, the determination of scope in

the ellipsis interpretation must match the scope determined in the antecedent. While this has been most explored in quantifier scope, it has also been shown that "the scope a wh-phrase has, relative to its clause, must match the scope its correlate has to the clause" (Johnson, 2001a p. 4). Thus, if the adjunct in the antecedent is blocked from moving to the matrix clause by the factive verb, then the correlate in the antecedent has been determined to scope only over the embedded clause, and the subsequent interpretation of ellipsis must also place the wh-element in this position, as shown in Figure 10.

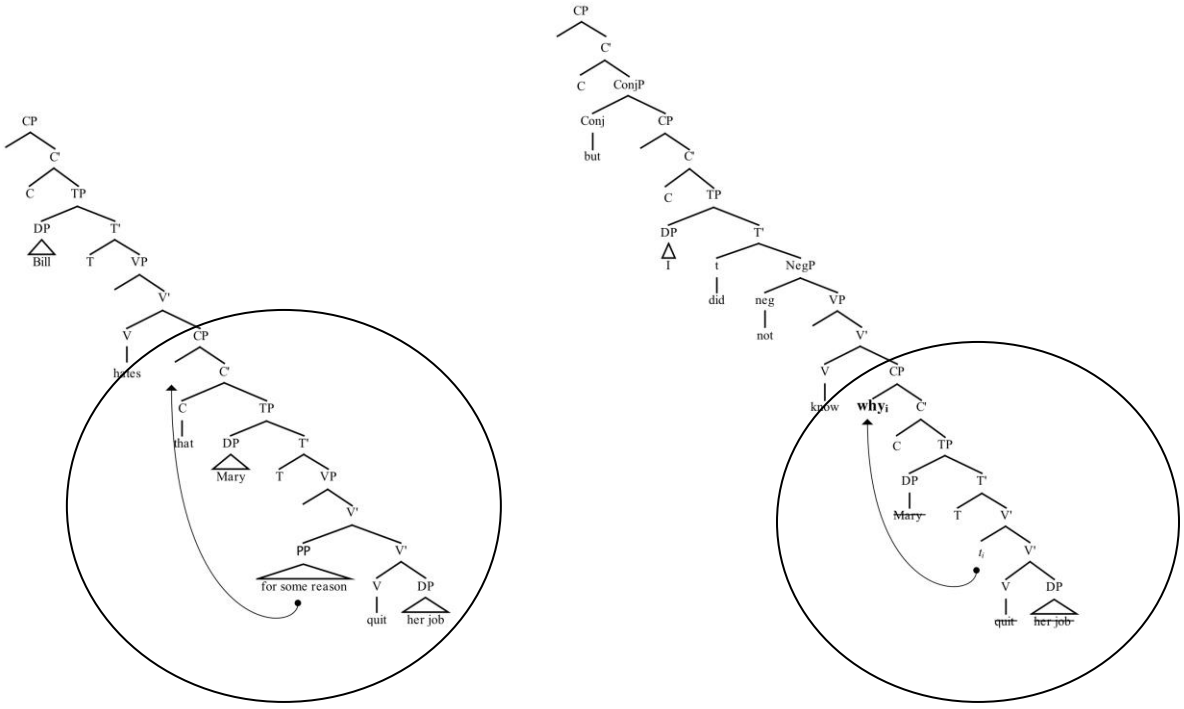


Figure 10:Tree on left demonstrates raising of adjunct to specifier position of the embedded CP in antecedent. Tree on right demonstrates movement of wh to parallel position from elided clause.

More concretely, in the non-factive cases, parallelism of scope guides the search process towards the longer, more complex antecedents, and in the factive cases scope parallelism guides the search to only the embedded clause. It is this determination, the scope of the adjunct, that

dictates the complexity of the antecedent to be retrieved. Finally, the complexity of the retrieved element affects processing, such that larger more complex antecedents are more costly to retrieve.

The observations and conclusions made in this chapter inform us as to the nature of the memory representation as well as the content of what is retrieved. It appears we must make room for structural information in both of these domains, and acknowledge that structural information is not just key to incremental processing and interpretation, but that is consulted during a range of other processes. One other domain in which structure has been argued to play a key role is in the identity constraints on ellipsis and in the interpretation of non-isomorphic content. I now turn to this topic in Chapter 4.

Chapter 4: Non-isomorphism in Ellipsis

Introduction

There exists an extensive literature concerned with the nature of ellipsis, in particular the constraints placed on the relationship between the ellipsis and the antecedent. It is apparent from a multitude of experiments and real-world examples that there are often differences between the antecedent of ellipsis and what is interpreted at the ellipsis site. These differences are sometimes tolerated, and at other times non-isomorphism is partially or wholly unacceptable. A large body of research has sought to determine the criteria used for matching an instance of ellipsis with its antecedent, which in broad strokes can be portrayed as a debate between theories which assume that the identity constraint operates over syntactic structures, and theories which assume that semantic representations are the relevant matching domain. The debate remains ongoing as the data do not neatly line up with either side, which has led to several hybrid models as well. While this chapter does not presume to settle or even tackle this central issue head on, I hope to demonstrate that some examples of non-isomorphism in ellipsis resolution are acceptable due to post-retrieval repair processes that change the retrieved material, and that the performance of these changes has an observable impact on real-time processing.

Background

Non-isomorphism in ellipsis

As stated, there are numerous examples of mismatches between an antecedent and an interpretation of ellipsis that are quite acceptable. For example, the sentence in (46) demonstrates

that the interpretation of ellipsis is not always identical to the pronounced form of the antecedent: in this case there is a voice mismatch between the elided material and the antecedent.

(46) These dishes were to have been washed, but nobody did ~~wash these dishes~~.

In (46), the antecedent employs passive voice, however the interpretation of the ellipsis uses active voice. The respective syntactic structures in (1) mismatch in voice as diagrammed in the trees in (11), however they are taken to be semantically equivalent. Sentences such as these demonstrate that there does not have to a perfect copy of the antecedent surface form interpreted at the ellipsis site.

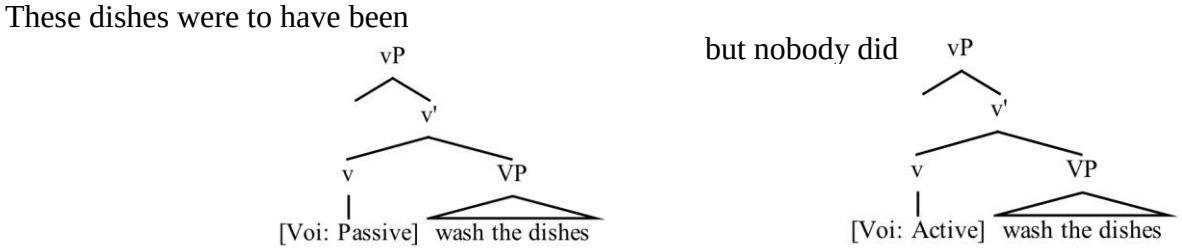


Figure 11:Trees for antecedent (left) and elided material (right) for example (1).

A quite common type of mismatch concerns the mismatch produced between the verb phrase ellipsis (VPE) site and its antecedent which contains a reflexive, as shown in (47):

(47)

a. John loved himself, and Bill did too.

b. John$_1$ loved himself$_1$, and Bill$_2$ ~~loved him$_1$~~ too.

c. John$_1$ loved himself$_1$, and Bill$_2$ ~~loved himself$_2$~~ too.

d. *John$_1$ loved himself$_1$ and Bill$_2$ loved himself$_1$ too.

In (47a), interpretation of the ellipsis site results in the equivalent of either (47b) or (47c). In the first instance, the so-called 'strict' identity (Ross, 1967), the pronoun retrieved during ellipsis resolution has an identical binding relationship as occurs in the antecedent, i.e. the pronoun is linked to 'John'. In (47c), the pronoun is interpreted as coreferring with 'Bill', the so-called 'sloppy' identity. One can see, however, that some type of mismatch occurs in either interpretation. To achieve (47b), the form of the pronoun cannot be isomorphic, as this would violate Binding Condition A (Chomsky, 1981b), which states that a reflexive anaphor must have an antecedent in its local domain, resulting in the unacceptability of (47d). Conversely, in (47c), the morphology of the pronoun is isomorphic, however the binding relation is not as now the antecedent for the pronoun is 'Bill' as opposed to 'John'.

A similar case in which a pronoun appears to have a different form within the ellipsis site is the phenomenon of "vehicle change" (Fiengo & May, 1994). Vehicle change refers to sentences in which certain Principle C violations fail to arise in elided structures, demonstrated in (48) and (49).

(48)

e. John loves Mary$_i$, and she$_i$ thinks Bill does too.

f. *John loves Mary$_i$, and she$_i$ thinks Bill loves Mary$_i$ too

g. *John loves Mary$_i$, and she$_i$ does too.

In example (48) above, (48a) presumably has the same interpretation as (48b), and if we assume that the ellipsis site in (48a) is associated with an exact copy of the VP antecedent in the previous clause, then we must explain the difference in acceptability between these two sentences. More specifically, if the ellipsis site in (48a) has the form of the last VP in (48b), then a Binding Condition C violation is predicted since the pronoun 'she' c-commands 'Mary'. However, given the high acceptability of the coreference relations interpreted in (48a) we must conclude that an isomorphic copy of the antecedent structure cannot be present. One suggestion to explain the absence of a Condition C violation is to posit that that the name in the antecedent 'becomes' a pronoun under ellipsis, thus evading ungrammaticality.

Furthermore, we must also explain why ellipsis appears to ameliorate a Condition C violation in (48a), but not in (48c). In the theory proposed by Fiengo and May (and re-examined in Merchant, 2001), the pronominal feature of nominals is allowed to vary under reconstruction, such that the interpretation of the ellipsis yields the pronominal correlate. Therefore, the interpretations of examples (48a-c) are actually those shown in (49a-c).

(49)

h. John loves Mary$_1$, and she$_1$ thinks Bill does too ~~($_{VP2}$ love her$_1$)~~.

i. *John loves Mary$_1$, and she$_1$ thinks Bill loves Mary$_1$ too

j. *John loves Mary$_1$, and she$_1$ does too ~~($_{VP2}$ love her$_1$)~~.

This proposal successfully accounts for the pattern of acceptability observed in (48). In (48b), the explicit R-expression in the embedded clause is c-commanded by a co-indexed

pronoun, and thus violates Condition C. In (48a), however, the R-expression has had its [-pronoun] feature changed (49a), and subsequently an R-expression is no longer c-commanded by a co-indexed pronoun, thus Condition C is not violated. Crucial support for this account comes from (48c/49c), where we can explain the unacceptability by referencing Binding Condition B, which states that the antecedent of a pronoun such as 'her' cannot be both local and in a position to c-command the pronoun. Proposing the presence of an element with a [+pronoun] feature in (48/49c) successfully predicts unacceptability due to Condition B effects since the pronoun 'she' c-commands 'her' in its local domain. The difference in acceptability of (48a) and (48c) is essential to Fiengo and May's claim, as it provides persuasive evidence that there is indeed a pronoun within the ellipsis site as opposed to an R-expression.

Many observations of antecedent mismatch have been observed, although most accounts of how these mismatches occur and are tolerated arise from the inspection of voice mismatches, as shown previously in (46), and reproduced below in (50).

(50) These dishes were to be washed, but obviously nobody did ~~wash these dishes~~.

Some argue that this kind of voice mismatch supports the view that an identity constraint operates over semantic representations (Dalrymple et al., 1991; Hardt, 1993; Kehler, 1993; Merchant, 2001) – the syntactic structures of the antecedent and the ellipsis of (50) are non-parallel, but the semantic representations are taken to be equivalent. However, even within the realm of voice mismatches a contradictory pair also exists, for instance this sentence in (51) from Kehler (2000):

(51) *This problem was looked into by John, and Bill did ~~look into this problem~~, too.

Sentences such as (51) have given credence to accounts that assume syntactic identity (Chung et al., 1995; Fiengo & May, 1994; Hestvik, 1995; Lappin, 1992; Sag, 1976; Williams, 1977). In general, however, accounts which adhere to syntactic identity under-generate acceptable sentences, however those which endorse a semantic account over-generate. To be able to account for this type of contradiction in attested examples of ellipsis, other explanations have made use of hybrid accounts, or have attributed the range of difficulty in interpreting ellipsis-antecedent mismatches to more general processing costs. One instance is Kehler (2000, 2002), who proposes that both syntactic and semantic identity are relevant to antecedent recovery, but that discourse relations mediate which of these types of information is employed during the recovery process. For example, when two clauses are in a resemblance type discourse relation, then syntactic identity is required, and any subsequent syntactic mismatch between the antecedent and the ellipsis clause results in degradation. In contrast, when a cause-effect type discourse relation holds between the two clauses, ellipsis is resolved via semantic representations and propositional identity is the relevant factor. This type of account has been supported by work performed by Kim and colleagues (Kim et al., 2011), who find a reduction in the syntactic mismatch effect under cause-effect relations, however the effect is not entirely eliminated. However, other experimental examinations of Kehler's proposal have failed to show that discourse consideration in fact mediate the acceptability of mismatches (Frazier & Clifton, 2006).

The variation in acceptability of voice mismatches has also been elegantly addressed purely in structural terms by Merchant (Merchant, 2008), who asserts that the acceptability of

mismatching voice is explained depending on the height of the structure being elided. He uses the comparison of VPE and pseudogapping (as in (52)) to demonstrate that voice mismatches are tolerable in VPE because a node lower than [v(oice)P] is elided, and thus voice is not subject to syntactic identity constraints because it is not contained within the ellipsis. Voice mismatch, however, is not tolerated in pseudogapping because the target of ellipsis is vP, and therefore voice is subject to syntactic identity. Trees illustrating this idea from Merchant (2008) are shown in Figure 12, where [E] on a head indicates that the complement is elided. The relevant point more generally is that tolerable voice mismatches *only* arise in the case that something lower than the head of vP is elided (for example VP). He asserts that this uneven distribution of voice mismatches in ellipsis constitutes a problem for theories of semantic identity, as without referring to the syntactic content of the elided material it is difficult to explain why voice mismatch should matter to one construction, but not to another.

(52)

k. **VPE**: The problem was to have been looked into, but nobody did ~~look into this problem~~

l. **PG**: * Roses were brought by some, and others did lilies ~~bring~~

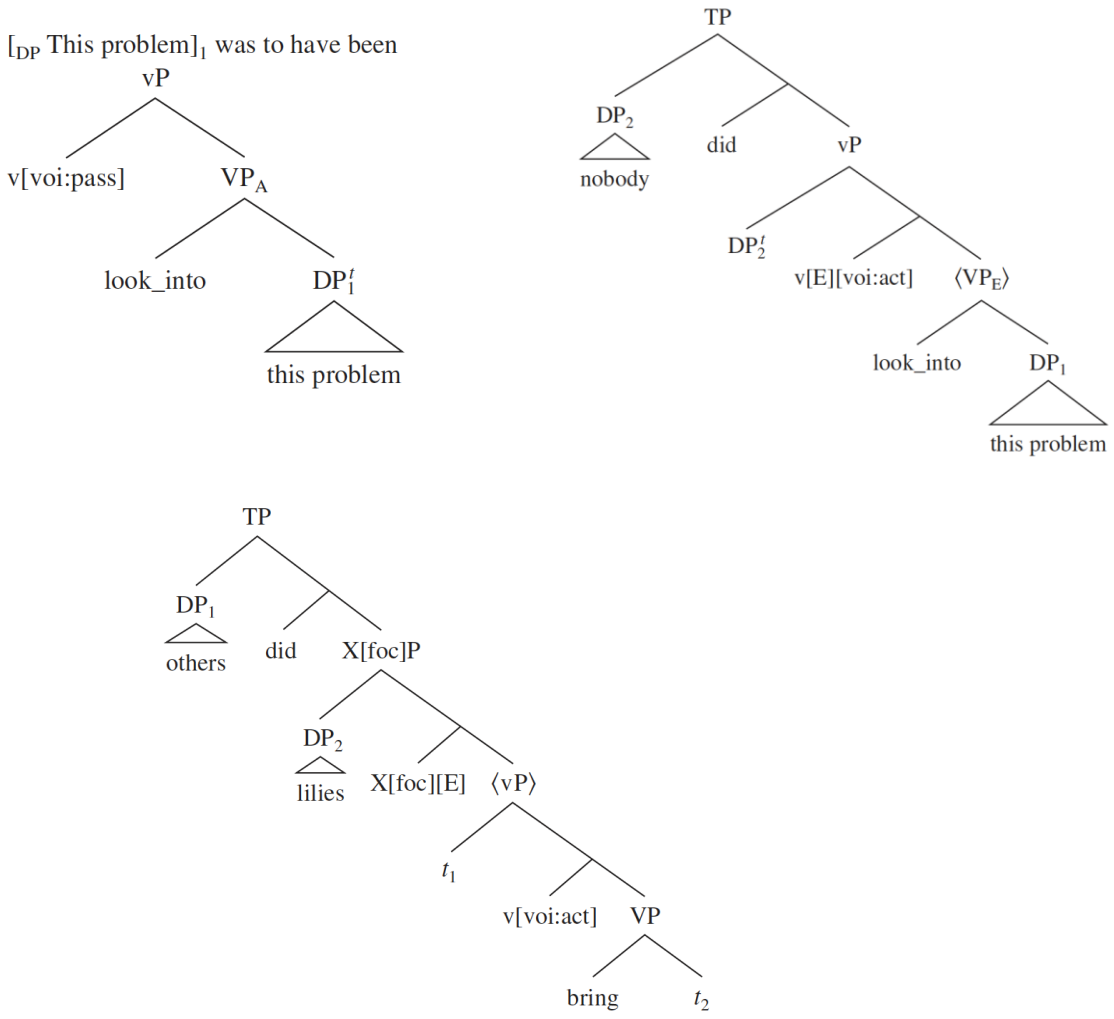


Figure 12: Trees demonstrating mismatching voice heads on tree structure for sentence in (22a), and different target of ellipsis on pseudogapping structure in (22b) .

Other attempts to reconcile the acceptability of syntactically defined antecedent-ellipsis mismatches have ascribed difficulty to processing considerations or post-retrieval repair operations. Experimental evidence suggests that rather than being a binary distinction, there is a cline of acceptability for syntactically mismatching antecedent-ellipsis pairs (Arregui et al., 2006; Kim et al., 2011), such that an increase in mismatch results in a decrease of acceptability. An oft cited example of this kind of approach is Kim et al. (2011), who through a series of

magnitude estimation experiments find the following pattern of acceptability, shown in (53) and (54) (where > indicates higher acceptability).

(53) **Mismatching voice**: Passive-Active > Active-Passive

m. The desert was praised by the customer after the critic did already.

n. The customer praised the desert after the appetizer was already.

(54) **Gerundive antecedents:** Verbal Gerund > Nominal Gerund

o. Singing the arias tomorrow night will be difficult, but Maria will.

p. Tomorrow night's singing of the arias will be difficult, but Maria will

The authors maintain that syntactic identity is the relevant domain, and that mismatches are in fact grammatical. They assert that mismatches are grammatical in that a canonical VP underlies all the mismatches in (53) and (54), but that degradation of acceptability is correlated with the number of derivational steps the parser is required to search through to locate the matching VP. The search heuristics they propose prioritize certain types of structures, e.g. no passivization, and also encourage maximal ellipsis (Takahashi & Fox, 2005). Violations of these heuristics require more search effort to recover a syntactically matching antecedent, and therefore are less acceptable.

Another approach that relies on extra-grammatical processing constraints or parsing heuristics to explain the relative acceptability of certain mismatches is given by Arregui et al. (2006) in what they call the 'VP Recycling' hypothesis. This approach differs from Kim et al. in that it ultimately regards all mismatches as ungrammatical, but also maintains syntactic identity (albeit over surface forms). In this proposal repair processes (like those used in garden path

recovery) create a syntactically matching antecedent by 'recycling' the materials from the antecedent. In essence, this kind of repair can be viewed as a form of parallelism, such that a non-parallel or non-matching interpretation is altered so that it is parallel to the antecedent. This form of recycling is a performance repair strategy that carries with it varying degrees of difficulty dependent on the number of repairs to be made, and as such predicts the cline that is observed in both their and Kim et al.'s experiments. For example, the preference for Passive-Active mismatch is predicted by this account since it is easier to recover an active form from the structurally more complex passive form than it is to restructure a passive from an active.

Building off of the work of these processing and repair based accounts, a recent proposal from Frazier and Duff (2019) refines the abilities and limitations of repair in a syntactic matching account. The authors maintain that when a fully matching antecedent cannot be found it will be repaired, but this repair is only carried out when there is sufficient evidence for the repair and relatively few operations are needed. An interesting contribution from this account is a proposal concerning the amount of syntactic material that can be held in working memory during sentence processing, a topic that is under-addressed. Based off a series of experiments implementing acceptability ratings, forced choice interpretation, and self-paced reading, the authors propose the Activated Syntactic Memory (ASM) hypothesis – that "memory holds the syntactic representation of the current sentence and the last potentially independent clause". While they admit that this is an idealization, and that clause size and complexity most certainly affect these limits, some of their evidence seems to support the hypothesis. For example, they observe that the connective (and subsequent syntactic structure) appears to impact the accessibility of antecedents: in coordination with two independent clauses accessibility of the

first clause is reduced compared to sentences that are connected with a subordinator like ‘after’, supporting their assertion that the last independent clause is held in memory.

A more recent account of acceptable mismatches and the cline of acceptability comes from Parker (2018), who asserts that the observed cline can be successfully captured by cue-based retrieval models without the application of parser-specific heuristics. Under this account, range of acceptability is a result of processing considerations linked to the cue-based retrieval of the antecedent, such that when the cues of the ellipsis site mismatch those of the antecedent there is a processing disruption. The size of the disruption (and subsequent degree of acceptability) is monotonically related to the degree of cue mismatch. Parker argues that this kind of explanation is parsimonious in that no violation of parsing heuristics or special repair rules are required, as are proposed in other processing based accounts (Arregui et al., 2006; Kim et al., 2011), but that it is the expectation made by the retrieval cues at the ellipsis site which is violated. This account, when combined with the computational claims of the ACT-R cognitive architecture, successfully predicts the observed cline in VPE, and furthermore provides empirical coverage not covered by other accounts, for example the preference for Active-Active matches vs. Passive-Passive matches.

Returning to the examples of strict and sloppy identity and vehicle change outlined above, one might ask how they can be accommodated in these frameworks, or whether their accommodation is necessary at all. It has been argued (and seems quite clear) that not all linguistic elements require an identical antecedent to be judged perfectly natural. Take for instance the sentence in (55), which is an undoubtedly acceptable example of gapping:

(55) I eat oysters, and Emily ~~eats~~ clams

The relevant point here is that the agreement morphology on the verb ‘eat’ is able to vary with no apparent penalty. Examples like this seem to demonstrate that there are particular mismatches that are ignored or inconsequential, principally those involving inflectional morphology, and that a matching criterion based on morphological or phonological form is untenable. There are certain exceptions in English, however, specifically involving the verb ‘be’ and the auxiliary ‘have’, which appear to have a unique status. Consider “* I am here and Mary will too”, which is unacceptable in contrast to cases of VPE with main verbs. It has been argued (Lasnik, 1995) that in English these two verbs are fully inflected in the lexicon, and as such cannot be separated from their inflectional morphology under ellipsis. In the case of pronouns, it does seem intuitive that given their special linguistic status (Jakobson, 1957; Kaplan, 1989) and since a pronoun’s reference can shift from utterance to utterance that the ellipsis process would do well to ignore certain features such as person, number, and gender. However, there does appear to be a sensitivity for a pronoun’s referential status (Fiengo & May, 1994; Kitagawa, 1991; McCawley, 1988; Sag & Hankamer, 1984), as in (56).

(56) Bill: You never returned **my** call!

Jane: Yes, I did ~~return **your** call~~

In (56), only the strict interpretation is available, and the mismatch between the morphological form of the pronoun appears to impose no burden. The coreference or *index*, meanwhile, is maintained. Whether this index is retrieved or recreated at the point of ellipsis resolution is unclear, and in general there does not seem to be consensus (nor a great deal of

interest) about whether indices or binding relations are maintained as part of the memory representation of anaphora. However, returning back to Chapter 2 where I observe parallelism of coreference across conjoined clauses, it appears more likely that this type of information is represented in some capacity in memory.

While the high degree of acceptability of sentences demonstrating phenomena like strict/sloppy identity and vehicle change are interesting examples of acceptable mismatch, little attention has been paid to investigating them in the realm of online processing, which may provide us with a more finely honed lens to observe potential processing disruptions due to mismatch. Furthermore, the observation of such disruption can illuminate what information is ultimately retrieved and is required to match with the retrieval cues. Finally, investigations of this nature may call into question accounts that propose matching a subset of the material (e.g. Kim et al., 2011) or semantic propositions. To this end, I present a series of experiments exploring offline and online processing of the interpretation of verb phrase ellipsis and pronouns, as well as vehicle change.

Experiments

The following experiments are constructed to probe costs associated with judging and processing certain types of elliptical constructions which involve non-isomorphism between the antecedent and the ellipsis site. More specifically, these experiments will employ examples of verb phrase ellipsis in which a pronoun or referent within the ellipsis site does not strictly match that which is within the antecedent. Experiments 5a and 5b examine cases of strict and sloppy identity in VPE using acceptability judgments and eye tracking, and Experiments 6a and 6b use identical tasks to examine cases of vehicle change in VPE.

Experiment 5a

Experiment 5a was designed to test whether comprehenders show any kind of offline preference for interpretations of VPE that demonstrate either a strict or sloppy interpretation. In this design, the first clause establishes the antecedent and binding relationship, the second clause involves VPE and the interpretation of the elided reflexive, and finally the third clause negates one of the available interpretations. A baseline condition is also included in which a proper name is used in the first clause instead of a reflexive, and the third clause causes a direct contradiction. An example of the stimuli used is shown in (57), and a full set of stimuli is included in the Appendix.

(57)

Strict

a. Bill punished himself, and Susan did too, but Susan didn't punish herself since it was unnecessary.

Sloppy

b. Bill punished himself, and Susan did too, but Susan didn't punish him since it was unnecessary.

Baseline

c. Bill punished Kendra, and Susan did too, but Susan didn't punish Kendra since it was unnecessary.

As in previous experiments, the acceptability judgement experiment is informative inasmuch as the processes being performed during online processing affect offline judgements about these interpretations. A failure to observe an offline effect is uninformative as to whether

the hypothesized processes occur online, however evidence of the predicted effects in offline processing may provide additional support for any observations during online processing. If we assume that a binding relation is part of the information retrieved by the parser when resolving VPE and pronouns which exist in the antecedent, then there may be a preference for a strict interpretation as it preserves the original (and presumably retrieved) binding of the referent and the reflexive. In opposition, the sloppy interpretation may require that the pronoun be 're-linked' to the local referent in the VPE containing clause, which involves an extra step of processing post-retrieval. This preference could emerge in Experiment 5a as an advantage for the strict interpretation condition over the sloppy condition, as in the absence of other pressures the parser may in a sense 'default' to a strict interpretation. In the sloppy condition, in which the possibility for a strict interpretation of VPE has been eliminated in the third clause, the acceptability may be degraded due to reanalysis of the VPE. If, however, binding information is not accessed when retrieving the antecedent, there is no obvious reason to predict a preference for either the strict or sloppy condition. Finally, the baseline condition aims to elicit judgements which reflect outright contradiction between the second and third clause without the availability of an alternative interpretation.

Methods and Materials

Participants

41 self-reported native English speakers from the United States were recruited on Amazon Mechanical Turk and compensated for their participation. No subjects were excluded from analysis.

Materials

The stimuli for this experiment consisted of 24 experimental items, and 76 unrelated fillers. As shown in (57), stimuli consisted of one condition which forced a strict interpretation, one which forced a sloppy interpretation, and one baseline condition. Stimuli were counterbalanced across 4 lists, and no subject saw more than one condition from each item.

Procedure

Prior to viewing experimental items subjects were shown 3 example sentences ranging from grammatical to ungrammatical, and demonstrating plausible vs. implausible scenarios. Subjects were instructed to rate each sentence on a 7 point scale based on the possibility of it being a sentence of English, even if implausible.

Analysis and Results

Responses for all items and subjects were retained, however as in previous analyses, raw judgments of the remaining experimental items were z-transformed within subjects to account for individual differences in usage of the rating scale.

Z-scored ratings served as the dependent measure for a linear mixed effect model with a maximal random effects structure. The model was performed using the lme4 package in R, and p-values were evaluated using model comparison between the maximal model and a reduced one.

Figure 13 presents the average z-score ratings for each condition of Experiment 5a. Pairwise comparisons indicate significant differences between judgments for each condition, such that the strict interpretation condition was rated significantly higher (mean z-score = .23)

than the sloppy interpretation condition (mean z-score = -.04), and the sloppy interpretation was also rated significantly higher than the baseline condition (mean z-score = -.19).

Table 20 presents the estimated coefficients, standard error, and t-value for the linear mixed effect model. A main effect of condition was found (χ^2 (df) = 15.81 (2), p <.0005).

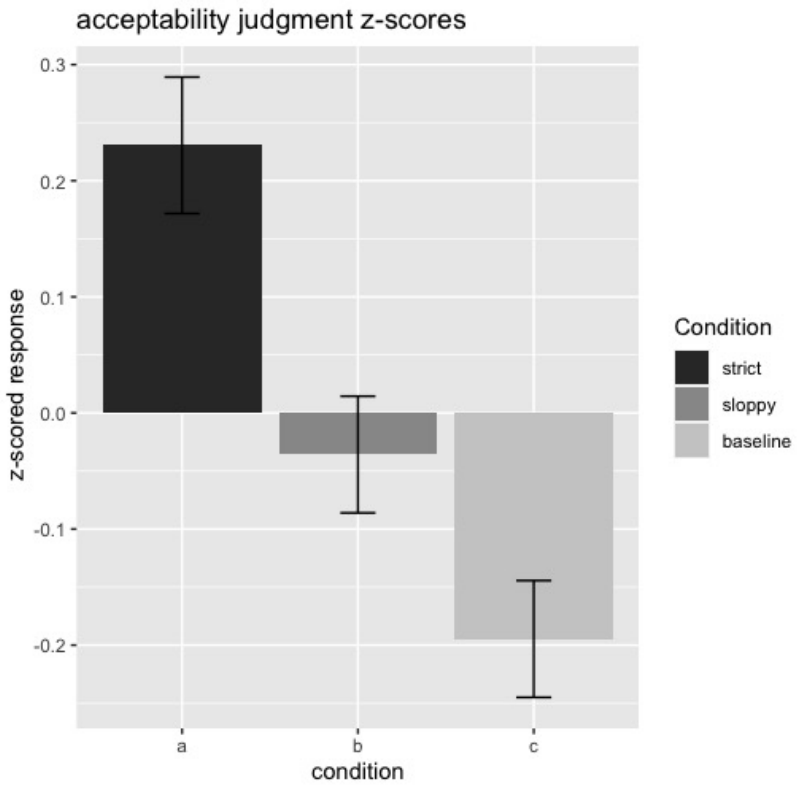


Figure 13:Bar plot of z-scores for Experiment 5a. Condition is on the x-axis, mean z-score is on the y-axis. Error bars represent standard error.

condition	**estimate**	**standard error**	**t-value**
strict	.43	.09	4.71
sloppy	.15	.08	1.97
baseline	-.19	.05	7

Table 20:Estimates, standard error, and t-values from the linear mixed effect model for Experiment 5a.

Discussion

The results from Experiment 5a are quite clear – the experimental manipulation resulted in significant differences in comprehenders' judgements. The most interesting (and somewhat surprising) result is the difference in mean ratings between the strict and sloppy conditions, such that there seemed to be a preference for the strict interpretation, in contrast to many previous claims and findings (Frazier & Clifton, 2000; Shapiro & Hestvik, 1995). These previous works claimed that the sloppy reading is preferred during comprehension either because (1) copying of a reflexive creates the preference for a local antecedent, (2) the non-local subject is more distant than the local one, and is thus harder to access, or (3) parallelism encourages parallel dependencies across the conjuncts. However, Shapiro and Hestvik (1995) only *assume* the sloppy preference, and in fact provide evidence that the strict and sloppy reading are both considered in real time as evidenced in a cross-modal priming paradigm (Shapiro & Hestvik, 1995; 2003). Furthermore, in the study performed by Frazier and Clifton (2000) using self-paced reading, it is unclear whether their stimuli was sufficiently biased towards either a strict or sloppy interpretation, and their measure of interest was both quite large ('Anne does too') and occurred at the end of the sentence, possibly conflated with wrap up effects. One explanation of the current study's novel finding is that in the strict condition, participants more often establish a strict interpretation of the VPE in the second clause, and so upon encountering the negation of the sloppy interpretation no revision of that interpretation is needed. In the sloppy condition, however, the comprehender must revise their original strict interpretation so that the sentence does not result in a clear contradiction. This revision may be costly, and therefore result in a

degraded judgment. The question that remains is *why* the strict interpretation is preferred initially.

As speculated upon in the predictions for this experiment, the answer may reside in the content of the material that is recovered during resolution of the ellipsis site. If the binding relation between the subject and the reflexive is part of the memory representation of the first clause, and if this relation is also recovered during the processing of the VPE, then there is no a priori reason for the parser to consider the alternative interpretation. Furthermore, arriving at the alternative interpretation likely requires additional steps of processing, namely the re-linking of the recovered pronoun to a new referent, which may also dissuade the parser from achieving this interpretation initially. It is important to note that these factors do not make the sloppy interpretation unreachable, contextual and pragmatic considerations may obviously exert pressures that make the sloppy interpretation more plausible or attractive as an option. Additionally, as indicated by the significant difference between the sloppy condition and the baseline, the sloppy interpretation, whether arrived at via revision or simply as an initial interpretation which requires additional processes, is markedly better than outright contradiction.

While the explanation given above may be the most straightforward one, it is largely unpredicted by a cue-based model in which the binding information is (likely) not part of what is retrieved, and straightforward retrieval of the isomorphic form of the reflexive is performed, resulting in the sloppy interpretation. It may be that there exists other alternatives as to why the sloppy interpretation is dispreferred in this experiment. One possible alternative is that the construction of the stimuli was such that the gender of the subject in the first clause and the gender of the subject in the second clause always mismatched. It could be that the binding relation in the first clause is not in fact recovered during VPE, but rather that changing the

gender and morphology of the recovered reflexive is itself costly. It could be this process that ultimately creates the bias I see in Experiment 5a, or it could be that both factors influence the interpretation of VPE in the experimental sentences. Another possibility is that rather than the processing of the VPE being the source of the strict preference, it is instead related to the ease of processing in the third clause (e.g. 'Susan didn't punish herself'), where the reflexive 'herself' may limit the search for its local antecedent in a more efficient manner than the pronoun 'him', which can be resolved by a more distant antecedent, but also can be open to an interpretation where 'him' is a third party not present in the immediate linguistic context. Indeed, it may be that both of these alternative factors are responsible for the strict preference.

In regards to the factors discussed above, a more temporally sensitive measure such as eye tracking may give us more detail about which of these possibilities is more likely. At the very least, looking more closely at the time course of anaphora processing should shed light on the processes involved in real-time dependency resolution within the ellipsis site. To further investigate the factors discussed above, an eye tracking while reading experiment was constructed to more precisely target the potential online processing differences in strict and sloppy interpretations as well as gender mismatch between the antecedent and recovered pronoun.

Experiment 5b

Experiment 5b further probes and expands the factors discussed in Experiment 5a by employing an eye tracking while reading paradigm in which interpretation and gender are manipulated. If the offline results seen in Experiment 5a are due to processing considerations related to the retrieval of binding information, then an observation that sentences with sloppy

interpretation impact eye tracking measures is predicted. Furthermore, if changing the gender value on the reflexive is also costly for the parser, then mismatching gender information between the antecedent and the recovered reflexive is predicted to slow down processing. More specifically, a main effect is predicted for strict vs. sloppy interpretation in the case of the retrieval of binding information, and an interaction is predicted for gender such that mismatching gender is only significant in the sloppy conditions.

Materials and Methods

Participants

52 participants from the Northwestern University undergraduate population participated in this task for course credit. Participants received 1 unit of course credit required by an undergraduate level Linguistics course. Participants had normal or corrected-to-normal vision, and were self-reported native English speakers. All participants provided informed consent.

Materials

The experimental materials were modeled on the items for Experiment 5a, however included several key changes. To be able to more concretely look at the interpretation of VPE and gender information, Interpretation (strict vs. sloppy) and Gender (match vs. mismatch) were manipulated in a 2x2 design. Furthermore, to be able to examine the VPE region directly, negation of one the interpretations needed to occur *prior* to encountering the VPE. An example item is shown in (58). In these items, the first clause contains the negation of either the strict or sloppy interpretation, the second clause contains the antecedent for ellipsis, and the third clause contains VPE and an additional spillover region.

(58) **Strict/Match**

a. Although Mike didn't punish himself, Bill punished himself and Mike did too since it was necessary.

Strict/Mismatch

b. Although Mary didn't punish herself, Bill punished himself and Mary did too since it was necessary.

Sloppy/Match

c. Although Mike didn't punish him, Bill punished himself and Mike did too since it was necessary.

Sloppy/Mismatch

Although Mary didn't punish him, Bill punished himself and Mary did too since it was necessary.

In addition to the experimental items, 84 unrelated fillers were also included. Each sentence was followed by a comprehension question, which were designed to target diverse areas of the sentences. Lists were pseudo-randomized according to a standard Latin square design to create 12 lists of items.

Procedure

Procedure for this Experiment was identical to the other eye tracking while reading experiments previously described. Participants first participated in practice trials, and performed the task using a tower mounted Eyelink Eyetracker. Participants were offered several breaks throughout the experiment, and recalibrations were performed as necessary.

Analysis and Results

Prior to analysis 1 item was excluded due to a typo, and data from 2 subjects were excluded due to low (< 2.5 sd less than mean accuracy) comprehension question accuracy. Fixations of 40 milliseconds or less were combined with fixations within one character of the fixation. As in previous experiments, first fixation duration, first pass time, regression path time, and total fixation time were analyzed. Analysis focused on 3 regions, as shown in (59): the critical region containing the VPE, and two subsequent spillover regions.

(59) Although Mike didn't punish himself, Bill punished himself and Mike **did too$_1$ since$_2$ it was$_3$** necessary.

As before, raw fixation values were log-transformed to meet normality assumptions, and the log-transformed values served as the dependent measure in mixed effects regression models. Intercepts were permitted to vary across subjects and items, and slopes of Interpretation, Gender, and the interaction between these factors were also allowed to vary across subject and item. If the maximal model failed to converge, random effects were reduced until convergence occurred. As before, significance was determined via model comparison using ANOVA. Raw means for each measure and region of interest are shown in Table 21, and (β) and standard error (S.E.) from the maximal model from each test are shown in Table 21, accompanied by the corresponding χ^2 value and significance from the model comparisons.

Gender	Interpretation	Critical region	Spillover 1	Spillover 2
FIRST FIXATION				

match	strict	242 (7)	240 (5)	256 (12)
mismatch	strict	237 (9)	234 (6)	238 (9)
match	sloppy	235 (8)	257 (8)	224 (8)
mismatch	sloppy	234 (8)	241 (7)	231 (7)
FIRST PASS				
match	strict	250 (8)	281 (8)	277
mismatch	strict	250 (9)	264 (8)	257
match	sloppy	243 (9)	283 (10)	235
mismatch	sloppy	246 (9)	281 (12)	252
REGRESSION PATH				
match	strict	266 (11)	350 (19)	696 (88)
mismatch	strict	266 (11)	313 (14)	590 (74)
match	sloppy	253 (10)	370 (32)	543 (63)
mismatch	sloppy	277 (15)	340 (17)	689 (94)
TOTAL TIME				
match	strict	340 (17)	381 (17)	367 (19)
mismatch	strict	347 (18)	390 (19)	387 (22)
match	sloppy	329 (16)	411 (18)	343 (19)
mismatch	sloppy	353 (16)	426 (21)	380 (20)

Table 21:Mean fixation values for each condition, region, and eye tracking measure for Experiment 5b.

Region	Effect	Estimate	Std error	χ^2 (df)	p-value
First Fixation					

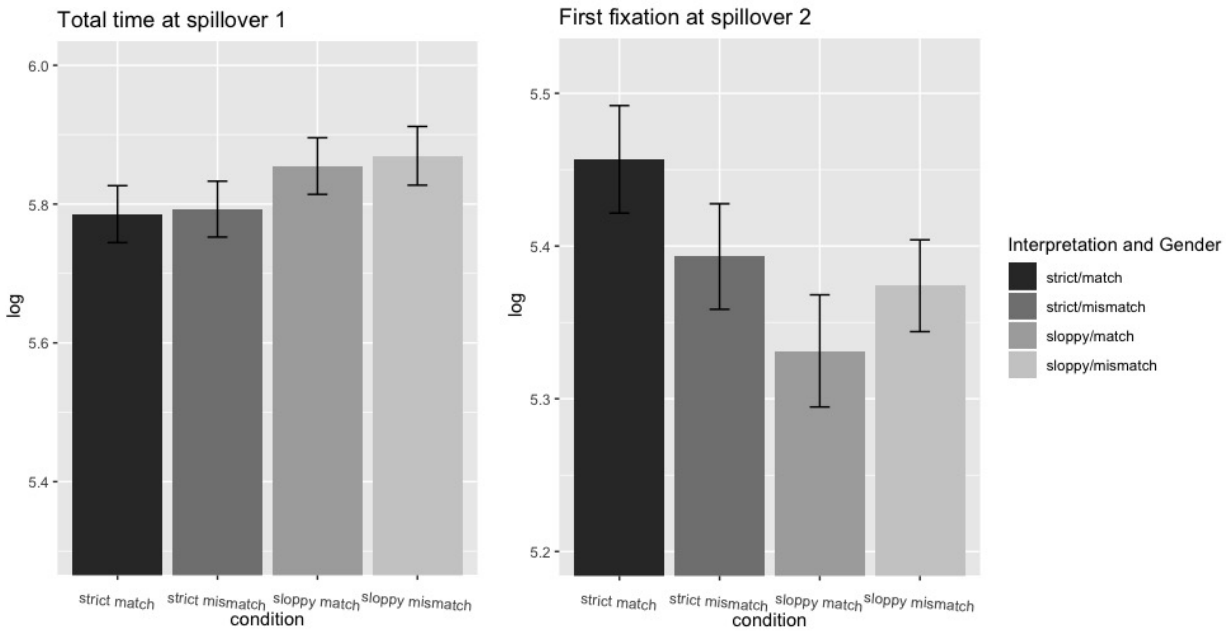
Critical	gender	-.01	.03	.3 (1)	>.1
	interpretation	.01	.03	.3 (1)	>.1
	interaction	-.02	.05	.12 (1)	>.1
Spillover 1	gender	-.05	.03	2.55 (1)	>.1
	interpretation	-.03	.04	1.04 (1)	>.1
	interaction	.02	.06	1.99 (1)	>.1
Spillover 2	gender	-.02	.04	.35 (1)	>.1
	interpretation	.06	.04	1.88 (1)	.17
	interaction	**-.13**	**.06**	**3.73 (1)**	**.05**
First Pass					
Critical	gender	0	.03	0 (1)	>.1
	interpretation	.02	.03	.54 (1)	>.1
	interaction	-.02	.05	.18 (1)	>.1
Spillover 1	gender	-.04	.03	1.56 (1)	>.1
	interpretation	-.01	.04	.06 (1)	>.1
	interaction	-.05	.06	0 (1)	>.1
Spillover 2	gender	-.01	.04	.09 (1)	>.1
	interpretation	.05	.05	0 (1)	>.1
	interaction	**-.19**	**.07**	**6.39**	**.01**
Regression Path					
Critical	gender	.03	.03	1.05 (1)	>.1
	interpretation	.01	.03	.23 (1)	>.1
	interaction	-.06	.06	.93 (1)	>.1
Spillover 1	gender	-.04	.04	.86 (1)	>.1

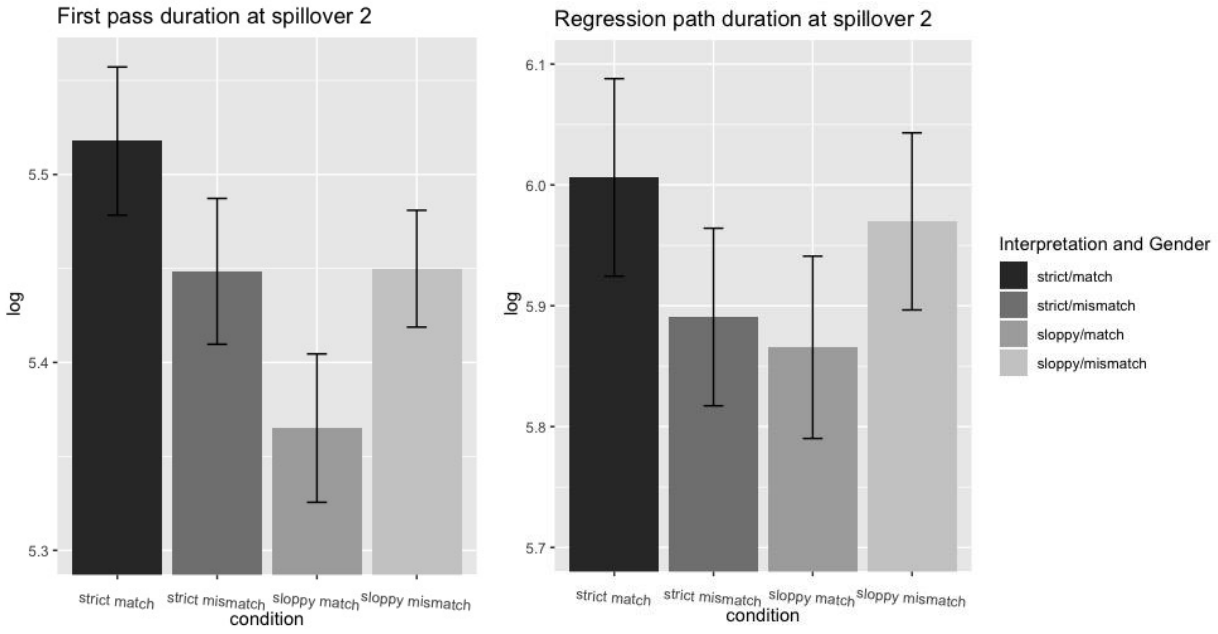
	interpretation	-.02	.05	.16 (1)	>.1
	interaction	-.1	.07	1.8 (1)	>.1
Spillover 2	gender	-.03	.09	.13 (1)	>.1
	interpretation	.03	.08	.63 (1)	>.1
	interaction	**-.26**	**.13**	**3.52 (1)**	**.06**
Total Time					
Critical	gender	.03	.04	.42 (1)	>.1
	interpretation	0	.04	0 (1)	>.1
	interaction	-.07	.08	.79 (1)	>.1
Spillover 1	gender	.01	.04	.02 (1)	>.1
	interpretation	**-.07**	**.04**	**3.10 (1)**	**.08**
	interaction	.01	.09	0 (1)	>.1
Spillover 2	gender	.07	.05	1.69 (1)	>.1
	interpretation	.04	.05	.85 (1)	>.1
	interaction	-.02	.09	.05 (1)	>.1

Table 22:Estimates, standard error, χ^2 values, and p-values for Experiment 5b.

No significant effects were observed at the critical region. At the first spillover region, a marginal main effect of Interpretation (p =. 07) was revealed in total fixation time, such that the sloppy interpretation had longer total fixation times than the conditions with a strict interpretation. At the second spillover region, a significant interaction was witnessed in first pass time ($p<.05$), such that mismatching gender resulted in longer first pass times only in the sloppy conditions. Furthermore, a marginally significant interaction (p = .05) was witnessed in first fixation time such that the strict interpretation exhibited longer first fixations than the sloppy

interpretation, but only when the gender was matching. Finally, the interaction of Interpretation and Gender approached significance in regression path duration, such that a gender mismatch effect was seen in the sloppy conditions, but a gender *match* penalty was seen in strict conditions.





Discussion

The results from Experiments 5b are lamentably not as straightforward as those from Experiment 5a. To reiterate, the main prediction that I made was concerned with the interpretation of the VPE, and how that relates to the information we believe is retrieved. If the parser retrieves not only the reflexive and the referent in the antecedent, but also the binding relation between them, then I expected to see a processing cost for sentences in which the binding information needed to be 'relinked' to a new referent, e.g. sentences which have a sloppy identity reading of VPE. The reasoning behind this prediction may be responsible for the effect seen in total times in the first spillover region, however this effect only approached significance ($p = .07$), and a main effect of Interpretation was not observed in any other measures or regions.

In regard to the surprising preference for the strict interpretation in the offline experiment but the general lack of facilitation for the strict interpretation online, the alternatives discussed for the results of Experiment 5a may now appear more likely. It would be useful in future experiments to probe this offline result more closely. One straightforward way to do so would be

to perform an additional acceptability rating task using the stimuli materials from the online portion to enable a more direct comparison between the offline and online experiments of Experiment 5.

In addition to the possibility that the binding information needed to be amended after retrieval occurred, my stimuli were also designed to investigate whether changing the gender feature on the retrieved pronoun is costly. While there is little question that gender is a feature assumed to be retrieved during the resolution of certain anaphora, it is unclear whether having non-isomorphism between the gender of a pronoun in the antecedent and the ellipsis site will result in a disruption. In this case, it was a possibility that a penalty associated with mismatching gender would surface in the sloppy conditions of the experiment. This prediction is indeed borne out in the first pass durations in the second spillover region, and is also somewhat supported by the marginal effect I saw in regression path durations.

Adding on to the anticipated results, there are some results from Experiment 5b that are entirely unanticipated and difficult to account for. The largest of these is the interaction witnessed in first fixation time such that the s*trict* condition elicited longer first fixations than the sloppy condition, and only in the gender matching condition. This appears to be instance of a gender match effect, which is typically attributed to similarity-based interference. If the retrieved reflexive initiates the search for a local antecedent within its binding domain, then such an effect is unsurprising in the strict conditions. This, however, would require ignorance (at least in initial stages) of the binding relation established in the antecedent. This point is addressed in more specificity in the general discussion for Chapter 3.

As a final note, I refrain from making any conclusions regarding the regions in which I see these results. While no effects emerge at the critical region, it may be that the presence of

ellipsis is not entirely obvious until encountering the conjunction/preposition following ‘did too’. The construction of the sentences used allowed for the possibility of a continuation that did not contain ellipsis, such as ‘did too many dishes’. While there is some evidence that the parser may actively posit ellipsis when available (Kim et al., 2020; Yoshida, Dickey, et al., 2013), it could be that for at least some portion of the sentences that participants waited for bottom-up information to confirm the ellipsis, which would effectively move the critical region one region to the right. It is unfortunate that more spillover area was not included in the original material, as the third potential spillover region is also the end of the sentence, and thus wrap-up effects might obscure any lingering or ongoing effects from the recognition of the ellipsis site.

Experiment 6

Experiment 6 again investigates VPE, however in this instance the phenomenon of vehicle change is examined. As discussed in the background section of this chapter, vehicle change refers to another case of non-isomorphism between the antecedent and the interpretation of ellipsis where an antecedent R-expression appears to be interpreted as a pronoun. The design for this (and the following) experiment pivots on the experimental findings of Kazanina et al (2007), who demonstrate that during the processing of cataphora the parser does not consider antecedents that are in structurally illicit positions, e.g. when the antecedent is c-commanded by the pronoun and therefore violates Condition C of Binding Theory (Chomsky, 1981b). Using the design in (60), the authors observed that no gender mismatch effect arose at ‘quarterback’ when the pronoun ‘she’ c-commanded the potential antecedent.

(60)

a/b. *constraint (match/mismatch)*

He/She chatted amicably with some fans while the talented, young quarterback signed autographs for the kids…

c/d. *no constraint (match/mismatch)*

His/Her managers chatted amicably with some fans while the talented, young quarterback signed autographs for the kids…

I adapt this design for Experiment 6 by combining cataphora and ellipsis to create the potential for an overt pronoun to c-command an antecedent within an ellipsis site, setting the stage for the possibility of vehicle change. Experiments 6a and 6b both use a gender match/mismatch paradigm to gauge whether vehicle change has occurred (6a/b), and whether that change is computed during online processing (6b).

Experiment 6a

Experiment 6a uses an acceptability task to discover whether comprehenders are sensitive to a gender mismatch between a pronoun and a referent in instances when vehicle change could remedy a Condition C violation, and therefore allow for the linkage between a pronoun and an elided referent. As noted, the design of the stimuli for this and the following experiment are adapted from the design of Kazanina et al. (2007). In (61), the first clause contains an overt cataphoric pronoun and VPE, the second clause contains a VP that can resolve the ellipsis in the first clause as well as an R-expression within that VP which can resolve the pronoun, and the final clause provides an additional referent which can ultimately resolve the pronoun if needed. A 2x2 design crossing Position (c-command/no c-command) of the pronoun, and Gender of the

pronoun and referent in the second clause (match/mismatch) results in the four conditions shown in (61).

(61) **c-command/match**

a. Since he didn't loudly, the students said Norman's name clearly, and Nathan heard it.

no c-command/match

b. Since his teachers didn't loudly, the students said Norman's name clearly, and Nancy heard it.

c-command/mismatch

c. Since she didn't loudly, the students said Norman's name clearly, and Nancy heard it.

no c-command/mismatch

d. Since her teachers didn't loudly, the students said Norman's name clearly, and Nancy heard it.

The predictions for this and the following eye tracking experiment are complex, and rely on a number assumptions about when and how the parser processes pronoun and ellipsis resolution, and about how these processes may interact. One large assumption is that the parser attempts to resolve dependencies as soon as possible in an incremental fashion and does not wait for the entirety of bottom-up information before linking controlling and dependent elements. As discussed in this and other chapters, there is substantial experimental evidence that demonstrates this kind of eagerness in parsing. Another key assumption is the conclusion reached by Kazanina et al: during the interpretation of cataphora the parser will not consider a coreference relation

between the pronoun and an R-expression c-commanded by that pronoun. Finally, there is the assumption that the search for the pronoun's antecedent will consider elided material that has been recovered. This assumption is supported by the very constructions that support the idea of vehicle change in the first place, namely the appearance of a Condition B violation in sentences like '* John$_1$ loves Mary$_2$, and she$_2$ does [~~love her$_2$~~] too', which indicates that the parser has attempted to construct a coreferential relation between overt 'she' and elided 'her'.

As an illustration of how these factors may play out, let us examine the examples in (16) piece by piece, diagrammed in Figure 14. In the first clause, when the parser encounters the pronoun, an active search for an antecedent for that pronoun is launched. Shortly thereafter, minimally at 'the students' but perhaps as early as 'loudly', the parser recognizes the presence of an ellipsis site (VPE) which initiates a (forwards) search for an antecedent of the elided material. In the second clause, the parser encounters the verb 'said' and after processing the VP links that VP to the ellipsis site. At this point several things (may) occur. The first concerns the phenomenon of vehicle change, which may happen immediately and automatically, or there may be a stage where the name 'Norman' is recovered in the ellipsis site, and subsequently is replaced by a pronoun. The second concerns the search that has been activated by the pronoun. Up to this point the parser has been actively searching for an antecedent outside of the adjunct clause, however with the resolution of the VPE the parser is now possibly able to include the interior of the adjunct clause as part of the search space. Whether or not the parser considers this space is somewhat at issue. Strictly following the conclusions from Kazanina et al.'s study, the material within the domain c-commanded by the cataphoric pronoun by hypothesis should not be evaluated as potentially containing the antecedent for that pronoun. Yet it may be that the

occurrence of vehicle change itself, such that there is a pronoun in the ellipsis site, may trigger the linking of that pronoun to the local potential antecedent.

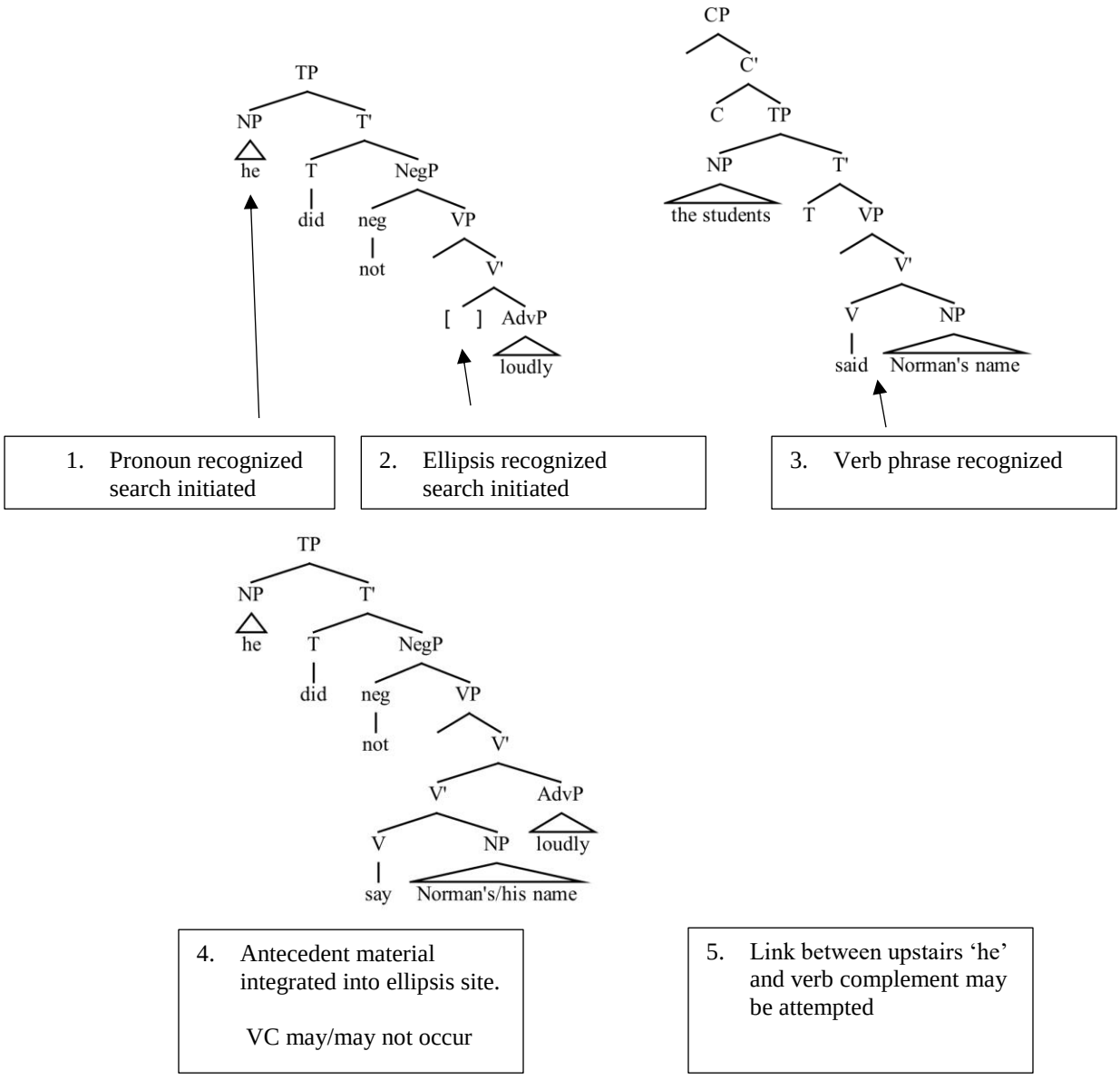


Figure 14:Time course and series of (potential) events during the processing of sentences in Experiment 6a-b.

Keeping these myriad factors and assumptions in mind, manipulating the gender of the first overt pronoun while maintaining the gender of the R-expression in the second clause can give some insight as to what is happening and when. If vehicle change is indeed occurring, I expect to see a gender mismatch effect regardless of the structural position of the pronoun, which may surface as degraded judgments for conditions c and d. In contrast, if vehicle change is not occurring, the parser should not attempt to link the pronoun to the R-expression in conditions 61a/c since the pronoun c-commands the R-expression, which is structurally illicit. In this case the prediction is that only condition d will result in a mismatch effect and degraded judgements.

Methods and Materials

Participants

61 self-reported native English speakers from the United States were recruited on Amazon Mechanical Turk and compensated $3.00 for their participation. Due to a presentation error, 10 subjects saw some item/condition combinations more than once, and were subsequently excluded from analysis. 3 other subjects were excluded due to extremely low standard deviation throughout the experiment (<.5 sd).

Materials

The stimuli for this experiment consisted of 36 experimental items, and 92 unrelated fillers. As shown in (61), stimuli consisted of four conditions in which Position (c-command/no c-command) and Gender (match/mismatch) were crossed. As one may note, the third clause did not always contain the same proper name (e.g. ‘Nathan’ vs. ‘Nancy’ in (61)). This design choice was made so that in no cases would the pronoun be left unresolved upon sentence completion. In (61a), if the vehicle change reading is unavailable, then ‘Nathan’ is able to ultimately resolve the

pronoun. In (61b) an additional referent is unnecessary, and having a stereotypically male one would only create ambiguity, therefore 'Nancy' is used. For (61c-d), a stereotypically female referent is required so that 'she' can finally be resolved. 8 counterbalanced lists were created using a standard Latin square design.

Procedure

Prior to participating, subjects were shown examples of 'natural' and 'unnatural' sentences, accompanied by sample ratings. They were instructed to consider how 'natural' a sentence sounded, and assign ratings on a 7 point scale, where 7 was the most natural. All subjects provided informed consent.

Analysis and Results

Identical to previous acceptability judgement analyses performed, raw judgements were z-transformed to account for individual differences in rating scale usage. Those z-scores were then used as the dependent variable in a linear mixed effect model with a maximal random effects structure using the lme4 package in R. Significance of the fixed effects was evaluated using model comparison between the maximal model and a model with a reduced term.

Mean raw values and standard error are shown in Table 23 , and Figure 15 presents a bar graph of the average z-score ratings for each condition of Experiment 6a. Results from the linear model are shown are shown in Table 24, and includes the estimated coefficients, standard error, χ^2 value, and corresponding p-values.

condition	mean (standard error)
match/c-command	3.32 (.08)
match/no c-command	3.09 (.07)

mismatch/c-command	3.25 (.08)
mismatch/no c-command	3.13 (.08)

Table 23:Raw means and standard error for each condition in Experiment 6a.

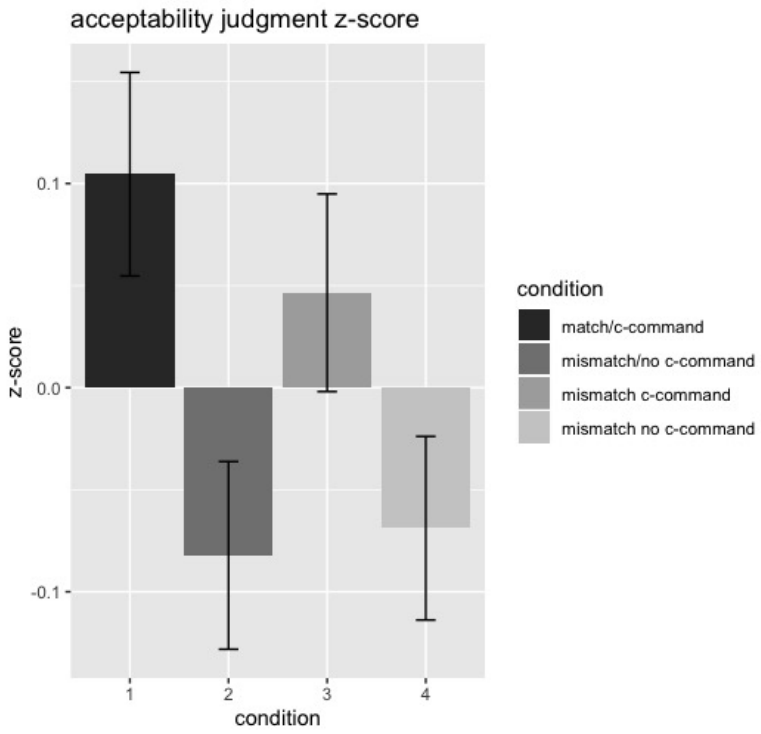


Figure 15:Bar plot of z-scores for Experiment 6a. Condition is on the x-axis, mean z-score is on the y-axis. Error bars represent standard error.

factor	**estimate**	**standard error**	**χ^2 (df)**	**p-value**
gender	.01	.02	.22 (1)	.64
pronoun	.07	.02	10.07 (1)	.002
interaction	.02	.02	.57 (1)	.45

Table 24:Estimates, standard error, χ^2 values, and p-values from the linear mixed effect model for Experiment 6a.

The statistical analysis reveals a main effect of Position ($\chi^2 = 10.07$, $p < .005$), such that conditions where the cataphoric pronoun c-commands the elided R-expression were judged as more natural than conditions where the pronoun was embedded in a more complex noun phrase. No other effect or interaction reached significance.

Discussion

The results from this experiment did not confirm any of my predictions, and furthermore had an entirely unpredicted but significant result. However, I believe that there may be several explanations for what we see. Regarding the lack of any gender mismatch effect in either of the mismatching conditions, it may simply be that the presumed online processes are not reflected in offline measures. While it is difficult to predict precisely which types of online processes impact final sentence judgements, it's possible that the processes under investigation are too subtle to emerge in such a relaxed task, or that other processing considerations obscure them.

As to the unpredicted main effect of position, and one which impacts judgements in a counterintuitive fashion, I believe two non-mutually exclusive explanations are available. The first is a simple matter of complexity – the sentences in which the possessive pronoun is embedded as part of the noun phrase are more structurally complex than those which have a bare pronoun, and this may account for the preference for the c-command conditions. An alternative, or perhaps additional, explanation is that in the no c-command conditions another animate entity is introduced, e.g. ' his *teachers'*. It was necessary for this set of stimuli that all the objects of possession were animate so that as the head noun they had the same qualities as the pronoun in terms of animacy and thematic role. Potentially the inclusion of an additional referent in the no c-command conditions also impacted the difficulty comprehenders had with these sentences.

In sum, the results of this initial experiment are uninformative to the issue I am probing, however this result does not eliminate the potential to see an effect of vehicle change in online processing. While it is beneficial (and reassuring) to see an offline effect bolstering any results which I may see in online measures, as noted these two routes of investigation are not always linked, or the effect may be small enough to be overwhelmed by other considerations. I turn now to the other route, and explore whether effects of vehicle can be observed by a more temporally sensitive measure.

Experiment 6b

Experiment 6b is an eye tracking while reading experiment which utilizes the design and materials of Experiment 6a to probe online effects of vehicle change in sentence processing, using the same gender match paradigm as in 6a. Although the acceptability judgement experiment failed to show an effect of vehicle change, I predict that using more temporally sensitive measures, that I might observe the presence of vehicle change as evidenced by a gender mismatch effect in both the c-command and no c-command conditions. No main effect of Position is predicted, as the point of interest is fairly far downstream from this conflation with complexity, and no lingering effect is predicted.

Methods and Materials

Participants

40 participants from the Northwestern University undergraduate population received 1 unit of course credit required by an undergraduate level Linguistics course for participation in

this experiment. Participants had normal or corrected-to-normal vision, and were self-reported native English speakers. All participants provided informed consent.

Materials

The materials for this experiment were identical to those used in Experiment 6a, and are shown again below in (62).

(62) **c-command/match**

a. Since he didn't loudly, the students said Norman's name clearly, and Nathan heard it.

no c-command/match

b. Since his teachers didn't loudly, the students said Norman's name clearly, and Nancy heard it.

c-command/mismatch

c. Since she didn't loudly, the students said Norman's name clearly, and Nancy heard it.

no c-command/mismatch

d. Since her teachers didn't loudly, the students said Norman's name clearly, and Nancy heard it.

92 unrelated fillers were also included with the experimental items. Each sentence was followed by a comprehension question, which were designed to target diverse areas of the sentences. Lists were pseudo-randomized according to a standard Latin square design to create 8 lists of items.

Procedure

Procedure for this Experiment was identical to the other eye tracking while reading experiments previously described.

Analysis and Results

Prior to analysis 1 item was excluded due to a typo, and data from 2 subjects were excluded due to low (< 65%) comprehension question accuracy. Fixations of 40 milliseconds or less were combined with fixations within one character of the fixation. First fixation duration, first pass time, regression path time, and total fixation time were analyzed, as in previous analyses. Analysis focused on 3 regions, as shown in (63): the critical region containing the point at which the VP can serve to resolve the ellipsis in the first clause, and two subsequent spillover regions.

(63) Since he didn't loudly, the students said Norman's **name$_1$ clearly$_2$, and Nathan$_3$** heard it.

As before, log-transformed fixation values were used as the dependent measure in mixed effects regression models. Intercepts were permitted to vary across subjects and items, and slopes of Position, Gender, and the interaction between these factors were also allowed to vary across subject and item. Model comparison was used as before to evaluate the significance of the fixed effects. Table 25 shows raw means for each measure and region of interest are shown, and (β) and standard error (S.E.) from the maximal model from each test are shown in Table 26, accompanied by the corresponding χ^2 value and significance from the model comparisons.

Gender	Pronoun	Critical region	Spillover 1	Spillover 2
FIRST FIXATION				
match	c-command	283 (21)	272 (11)	231 (7)
match	no c-command	263 (10)	271 (10)	235 (9)
mismatch	c-command	299 (32)	265 (7)	249 (9)
mismatch	no c-command	257 (8)	272 (8)	238 (8)
FIRST PASS				
match	c-command	313 (21)	323 (13)	275 (11)
match	no c-command	285 (12)	331 (12)	274 (12)
mismatch	c-command	328 (33)	316 (10)	302 (12)
mismatch	no c-command	282 (10)	336 (11)	284 (11)
REGRESSION PATH				
match	c-command	423 (32)	518 (36)	414 (34)
match	no c-command	369 (20)	577 (40)	296 (53)
mismatch	c-command	421 (37)	564 (36)	467 (40)
mismatch	no c-command	381 (27)	522 (34)	483 (60)
TOTAL TIME				
match	c-command	514 (27)	556 (26)	461 (21)
match	no c-command	492 (35)	550 (23)	465 (24)
mismatch	c-command	540 (35)	564 (23)	506 (23)
mismatch	no c-command	450 (21)	561 (22)	431 (2)

Table 25:Means and standard error for each measure and region in Experiment 6b.

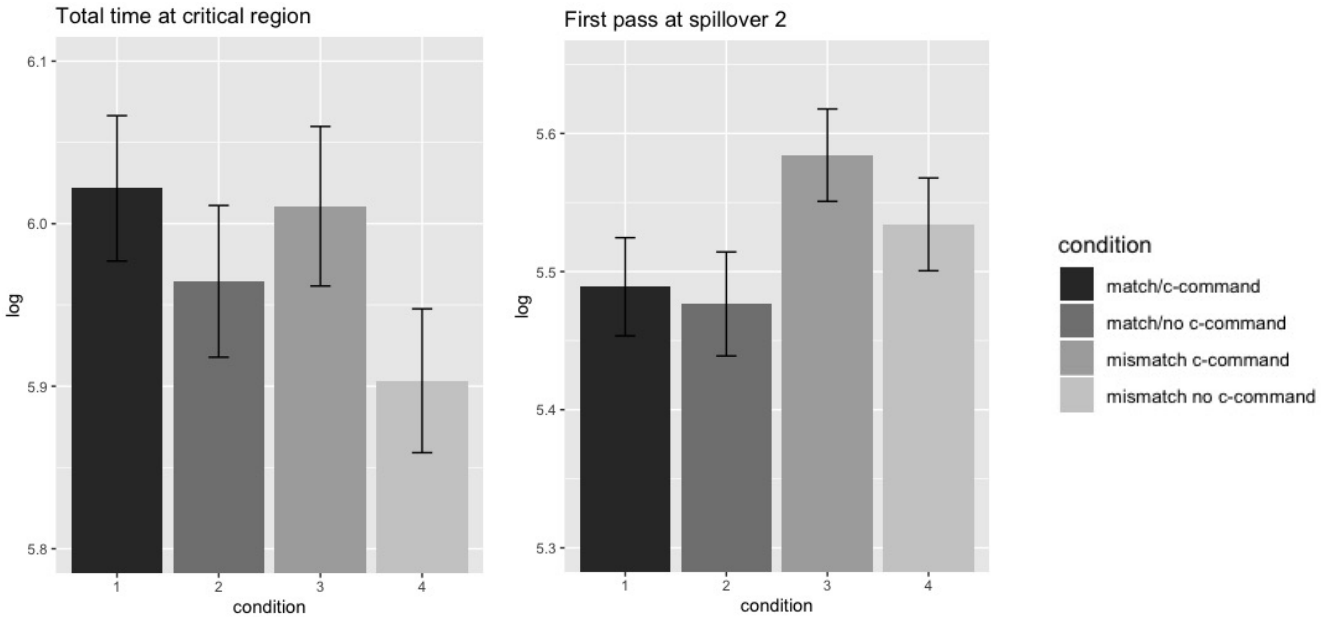
Region	Effect	Estimate	Std error	χ^2 (df)	p-value
First Fixation					
Critical	gender	-.01	.04	.07 (1)	>.1
	pronoun	-.02	.04	.31 (1)	>.1
	interaction	.07	.06	1.28	>.1
Spillover 1	gender	.008	.03	.08	>.1
	pronoun	.008	.03	.07	>.1
	interaction	.04	.06	.52	>.1
Spillover 2	gender	.05	.03	2.74 (1)	.1
	pronoun	-.03	.03	.79 (1)	>.1
	interaction	-.06	.05	.96 (1)	>.1
First Pass					
Critical	gender	-.02	.04	.10 (1)	>.1
	pronoun	-.04	.04	.3 (1)	>.1
	interaction	.08	.07	.22 (1)	>.1
Spillover 1	gender	.008	.03	.05	>.1
	pronoun	.04	.04	1.14	>.1
	interaction	.06	.06	.79	>.1
Spillover 2	**gender**	**.09**	**.03**	**7.36**	**.007**
	pronoun	-.03	.03	.85 (1)	>.1
	interaction	-.04	.06	.45 (1)	>.1
Regression Path					
Critical	gender	-.03	.04	.49 (1)	>.1
	pronoun	-.04	.05	.48 (1)	>.1

	interaction	.07	.09	.7 (1)	>.1
Spillover 1	gender	.01	.04	.08 (1)	>.1
	pronoun	.02	.05	.22 (1)	>.1
	interaction	-.12	.09	1.84 (1)	>.1
Spillover 2	gender	.05	.05	.96	>.1
	pronoun	-.001	.05	0	>.1
	interaction	-.11	.1	1.09	>.1
Total Time					
Critical	gender	-.02	.05	.2 (1)	>.1
	pronoun	**-.07**	**.04**	**2.91 (1)**	**.08**
	interaction	-.04	.08	.29 (1)	>.1
Spillover 1	gender	.03	.04	.47 (1)	>.1
	pronoun	.001	.04	0 (1)	>.1
	interaction	.03	.08	.18 (1)	>.1
Spillover 2	gender	.04	.04	.95 (1)	>.1
	pronoun	**-.09**	**.04**	**5.05 (1)**	**.03**
	interaction	-.13	.08	2.56 (1)	>.1

Table 26:Estimates, standard error, χ^2 values, and p-values for Experiment 6b.

At the critical region, a main effect of Position approached significance ($\chi^2 = 2.91$, $p = .08$), such that conditions where the pronoun was in a c-commanding position had longer total times. No effects were witnessed in the first spillover region, however in the second spillover region a main effect of Gender reached emerged ($\chi^2 = 7.36$, $p < .01$) in first pass duration such that mismatching gender elicited longer first pass duration. Additionally, a main effect of Position in total time also reached significance ($\chi^2 = 5.05$, $p < .05$) such that conditions where the

pronoun c-commanded the elided material had longer total reading times. Bar graphs of the measures with significant effects are shown in Figure 16.



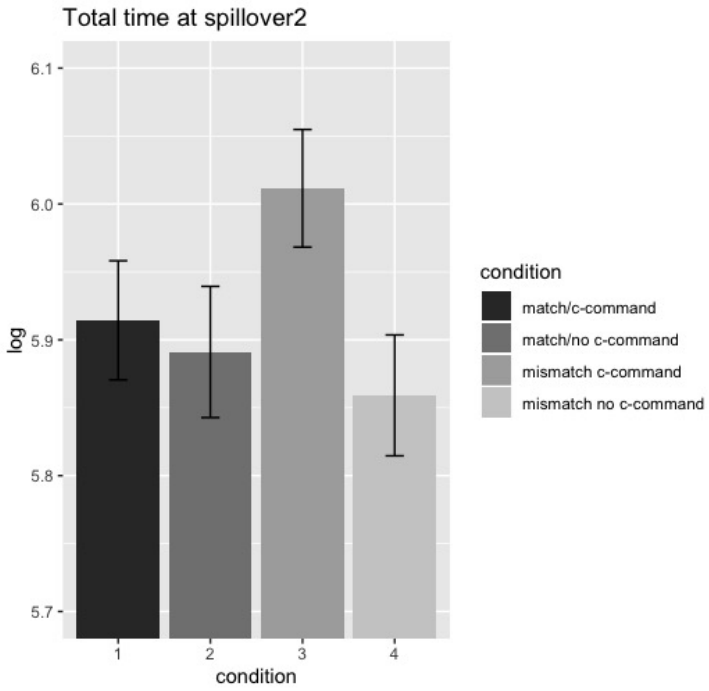


Figure 16:Bar plots of total reading times in the critical region, first pass duration at the second spillover region, and total time at the second spillover region. Error bars represent standard error.

Discussion

The results from Experiment 6b are more in line with the predictions made prior to analyzing the results than those from Experiment 6a, however it is difficult to say that my hypothesis about the effects we would see in this paradigm was strongly confirmed. The result that most supports that idea that vehicle change is in fact computed during parsing is the significant processing difficulty induced by mismatching gender in both the c-command and no c-command conditions in first pass duration at the second spillover region. Based on the assumption (and evidence) that the parser will not consider an R-expression that is c-commanded by a pronoun as a contender to resolve that pronoun, the main effect of gender mismatch

indicates that the proper name in the pronounced VP is not isomorphic to the referent in the elided VP.

The second main finding from this eye tracking experiment is that the structural position of the pronoun also had an effect on processing, although in the opposite direction of the effect seen in the offline task. I believe there are two main ways to explain this effect. The first possibility is that upon resolving the ellipsis, the parser recognizes that the R-expression is in a position to be c-commanded by an unresolved pronoun, and furthermore recognizes that this creates a Binding Condition C violation. Although potentially ameliorated by the ellipsis, this violation itself, regardless of whether we believe vehicle change to be taking place, could adversely affect processing. However, given the evidence from the gender manipulation which indicates that vehicle change has taken place, the process of converting the proper name into a pronoun may be a costly process. Akin to what I concluded in Experiment 5, having to amend or change features on material in order to achieve a viable interpretation is not cost-free, and is an example of how non iso-morphism in ellipsis can be tolerated and interpreted by the parser.

A critical point to consider is that in the case of the vehicle change experiments, the phenomenon in question is not occurring during or related to the act of retrieval. The implementation of the backwards VPE construction flips the resolution of ellipsis from a backwards retrieval problem into a forward search mechanism. While it is possible that the first clause which contains both the pronoun and the VPE site is stored in memory and then later retrieved when a suitable VP is encountered to resolve the ellipsis, such a scenario would minimally require some trigger for retrieval. There is not an obvious or logical source of such a trigger. A more likely hypothesis is that faced with open dependencies, the parser maintains the first clause and incorporates, perhaps incrementally, the VP in the second clause when it is

encountered. It is during the course of this incorporation that VC must occur. It is the fact that alteration to material interpreted at an ellipsis site *occurs* which is important, not whether that material needed to be retrieved or not.

Two additional points should be made about the observations from this data. As helpfully pointed out by a reviewer, the effect that is seen in the second spillover region could also be attributed to the resolution of the pronoun in the gender mismatch conditions. The second spillover region contains the final referent, which in the mismatch conditions is the first (and only) opportunity to successfully close that dependency. It is unfortunate that additional material was not included to expand the spillover area prior to the introduction of the last referent, and in future studies this oversight will be corrected. However, if this was indeed the cause of the effect, we would have to assume that prior to this region the pronoun in the gender match conditions had already been resolved. I see no equivalent effect in any prior region that would indicate a similar resolution cost in the gender match conditions.

Finally, there may an issue of power concerning the total time result in spillover 2. Although comparison of mixed effect models with maximal random effects structure revealed a *main* effect of Position, looking at the data (for example in the bar graphs) gives the strong sense of an interaction such that position mattered very much in the mismatch conditions, but did not so robustly affect processing in the match conditions. A neat explanation which falls in line with my general predictions for this experiment is that the possible interaction is due to an additive effect of (1) the processing costs associated with the computation of vehicle change, and (2) the gender mismatch between the pronoun and the elided referent. However, in this scenario I would also expect to see an effect of gender in the no c-command condition, which does not appear to be the case. The lack of a significant interaction, as well as a lack of clarity regarding why such

an interaction would occur, give cause to avoid drawing any strong conclusions about this issue at this point.

General Discussion

The conclusions of Experiments 5 and 6 are not as straightforward as desired, however there are some intriguing results which warrant more exploration. At a very broad level, these two sets of experiments sought to find evidence of a penalty due to the mismatch of features associated with referents such as names and pronouns (potentially including binding relations), and operated under an assumption that repair processes may be responsible for the ultimate acceptability of these sentences as well as associated processing disruptions. In acceptability judgements, I found a clear preference for the strict interpretation of pronouns in verb phrase ellipsis, however found weaker indications of facilitation for the strict interpretation in online processing. Additionally, an effect of gender was found in the strict/sloppy experiments, which is unexpected under many accounts in which gender features are irrelevant during reconstruction. In the vehicle change experiments, conditions in which vehicle change was expected to occur were ultimately rated as *more* acceptable, however this may be due to other considerations involving the complexity of the conditions in which vehicle change was not expected to occur. However, in the online portion of the vehicle change experiments a gender mismatch effect surfaced, which was only predicted if vehicle change had indeed taken place, and there also seemed to be a processing penalty associated with conditions in which vehicle change occurred.

How can these results be accommodated under existing theories of ellipsis-antecedent mismatch, and are additional mechanisms or heuristics required to explain them? To the extent that the dominant models that I consider most probable, e.g. those that rely on syntactic

considerations and identity, and in particular those that are able to address the gradience of acceptability, these particular cases are difficult to explain. In the case of strict and sloppy identity, syntactic form is upheld in both cases, however morphological and indexical information may vary. There is an argument to be made that the strict interpretation does match at some morphological level if we take the composition of a reflexive to be [+ pronoun] [+self]: the pronoun interpreted at the ellipsis site matches the [+pronoun] feature, and merely lacks [+self]. This type of subset match is precisely what is proposed by Fiengo and May, and is akin to the proposal made by Kim et al. in that there is some step in the derivation where forms are isomorphic. If the mismatch in the surface form is thus irrelevant in these cases, then one can say that the binding information is retrieved during VPE, and the preference for the strict interpretation is straightforward.

While the story above aligns with the results from the offline strict and sloppy experiment, the online experiment is more difficult to explain away in this fashion, or at the very least requires more detail about what additional information is retrieved. The preference for the strict interpretation is somewhat supported by total fixation times, but the more robust observation is that there appears to be a gender match penalty in the strict interpretation conditions, and a (predicted) gender mismatch effect in the sloppy conditions. A gender match effect has typically been attributed to similarity based interference when there is more than one potential antecedent that matches the gender cues of the probe (e.g. Badecker & Straub, 2002). For a gender match effect to surface in the strict conditions, it must be that antecedent search is initiated during resolution of VPE and the pronoun internal to the ellipsis, regardless of whether indexical information is also retrieved. When the retrieved pronoun is a reflexive, as it is in all conditions in the online experiment, it spurs the search for a local antecedent within the

reflexive's binding domain that c-commands the reflexive. In the strict interpretation sentences, resolution to this local antecedent is untenable – the context of the sentence creates a strong bias for resolution to the non-local antecedent. Additionally or alternatively, retrieval of the indexical relation between the pronoun and NP in the antecedent clause encourages the strict interpretation. Taking this online result into consideration in light of the results of the offline experiment, in which context *did not* strongly encourage a strict reading yet the strict reading was robustly preferred, I conclude that not only is the surface form of the reflexive retrieved, but that the binding relation is as well.

However, these different pieces of information may not be available to the parser at the same time. The interaction observed in the online experiment indicates that the reflexive spurs a search before the indexical information is recognized. Somewhat parallel findings for this idea come from Shapiro et al. (2003), who in a cross modal priming paradigm find that even when a verb is incongruous with either a strict or sloppy interpretation during VPE (e.g. * The fireman$_1$ perjured himself$_1$, and the police officer$_2$ did ~~perjure him$_1$~~ too), comprehenders appear to reactivate both potential antecedents at or near the point of ellipsis. They take this as evidence that the parser ignores lexical and probabilistic information during initial processing of the ellipsis, and builds parallel syntactic analyses of the elided verb phrase which entertain the strict and sloppy readings, respectively. Although this is not precisely what I claim here, it is additional evidence that the time course of availability for different kinds of information may not be identical.

This conclusion is also supported when looking at the results from the sloppy conditions. In these conditions search for a local antecedent in the reflexive's binding domain is satisfied, and the context encourages that resolution. What then of the retrieved binding relation? It is

difficult to discern an online effect of retrieved indexical information in the sloppy conditions, which I presume would be disruptive, however as stated there is a trend in total times indicating more overall difficulty in the sloppy conditions. Given the similarity-based interference I seem to be observing in the strict cases and the gender mismatch effect I see in the sloppy conditions, there is not an obvious opportunity to directly assess a main effect in the absence of other factors. However, once again the offline observations most strongly support the conclusion that binding information is present during the resolution of pronominals within VPE. A more straightforward observation from the sloppy conditions is the emergence of a gender mismatch effect, which indicates that gender information of the retrieved reflexive is present as well, and is serving as a cue during antecedent search (evidenced as well in the interference effect in the strict conditions). Clearly this gender mismatch during VPE is much more acceptable than its overt pronounced correlate, and so it seems unlikely that the effect I see is simply a result of cue mismatch (in the sense outlined by Parker, 2018), which would presumably result in dramatically degraded acceptability (e.g. * *Bill punished himself, and Mary punished himself too*). A process that repairs the antecedent by changing its gender value to salvage the interpretation seems a more likely explanation.

Turning now to the results of the vehicle change experiment, we see another instance where a gender match/mismatch paradigm illuminates the presence of certain information in the ellipsis site. In this case, I observe an effect that can be explained by positing that an R-expression has assumed the form of a pronoun during integration into the ellipsis site. This hypothesis is not new, however evidence of this phenomenon occurring in real time processing is. The results of the offline vehicle change experiments were not terribly illuminating due to the conflation of certain factors, however the online experiment demonstrated the ability of a

cataphoric pronoun to link to a NP in the ellipsis site that would be unavailable if that NP was a name. The apparent availability of the NP supports the idea that the R-expression in question has changed during reconstruction. There was, additionally, a processing cost reflected in reading time measures that indicated a difficulty associated with sentences in which vehicle change was presumed to have taken place. As in the cases of strict and sloppy identity, the overt forms of these sentences are unacceptable, but are allowed during ellipsis, however not without incurring certain penalties.

To summarize, I witness penalties associated with gender mismatch, indexical mismatch, and name/pronoun mismatch. In opposition to much of the literature in this area, in these cases the elided material demonstrates more acceptability than if the overt form occurred without ellipsis. In other instances, voice mismatch for example, a penalty seems to arise specifically in cases in ellipsis, which has been used to argue that isomorphism is a condition on ellipsis and is not due to overall preferences of parallelism. In my experiments, the reverse is true – the elided form is tolerated but the pronounced isomorphic form is not. How does this arise? Accounts that rely on matching syntactic form seem unable to accommodate these types of results, and semantic accounts that rely on matching propositions also seem insufficient. The data then leads us to consider accounts that incorporate processing considerations or repair processes to account for the acceptability of these mismatches. The cue-mismatch account of Parker (2018) is attractive in that it relies on mechanisms already at play during the resolution of various dependencies, however it fails to account for the attenuation of a strong mismatch effect like we would witness in the pronounced forms. For example, it is unclear why or if the cue probes of an elided pronoun would have a weaker call on the antecedent than an un-elided pronoun.

A more probable account of the phenomena I present in this chapter is a post retrieval process that alters the material of the controlling element to create an acceptable and grammatical parse, and is more in line with the proposals made by Arregui et al. (2006) and Frazier and Duff (2019). In this scenario, the material incorporated into the ellipsis site undergoes repair to match the required interpretation based on the cues at the ellipsis site. In the case of vehicle change, this involves allowing the pronominal feature on the noun to change, but this feature is not allowed to freely vary in the sense of Fiengo and May, but is rather a step performed by the parser which incurs real processing costs, as seen in Experiment 6b. In the case of sloppy identity, the gender feature is amended in instances where the retrieved pronoun mismatches in gender with the local antecedent, which again incurs a processing cost.

Chapter 5: Conclusion

In this thesis I have aimed to illuminate the prominent and crucial role that structural information and relations play during the resolution and interpretation of long-distance dependencies in language comprehension. At the outset I asked three primary questions: (1) Are relations in previously processed material able to affect subsequent processing? (2) Does processing of a retrieved element vary as a result of antecedent complexity? And (3), is there evidence that the parser alters or amends material in real time? The nature of scientific inquiry rarely allows an unequivocal 'yes' or 'no', however the observations made in Chapter 2-4 point to the affirmative for all three issues.

Based off the observations made in previous chapters, there are strong indications that hierarchical structural information is encoded and stored with generous specificity. This conclusion is not inherently incompatible with dominant retrieval models, however general claims in this area of research have leaned to embrace feature-based models wholeheartedly, leaving the structural output of parsing on the sidelines. In the previous chapters I have demonstrated via acceptability judgment and eye tracking experiments that relations that reference relative structural position can impact subsequent dependency formation, and that the structural content of an antecedent has an observable effect on retrieval. Additionally, I have shown that the material that is ultimately retrieved and/or linked to the ellipsis site demonstrates a remarkable degree of faithfulness, such that while certain features and designations are able to

vary under ellipsis, this variation is not free and indicates retrieval of detailed information on various levels.

In Chapter 2, I examined the role of parallelism in dependency formation, and found a robust effect in both coordinate and subordinate environments of a previously processed dependency on a following dependency. This result was apparent in both offline and online measures. While the ultimate source of the effects was presumed to be a type of priming, the more curious finding concerned *what* precisely was being primed. In the types of constructions I used, resolution of a pronoun in one conjunct to a subject of either a matrix or embedded clause impacted whether subsequent pronoun resolution preferred resolution to an antecedent in the same position. It is crucial here that the antecedent options are identical in almost every respect (vis-à-vis their phrasal category, thematic role, etc.) except in their structural position relative to the pronoun itself. The relation established in the first conjunct must be stored in memory in structural terms, and the representation of that relation is able to facilitate or direct the resolution of a succeeding pronoun to an antecedent in an identical structural relation.

In Chapter 3, I probed the effect that structure might have on ellipsis resolution, testing the prediction that an increase in structural complexity would negatively impact the processing of an ellipsis site. Using sluicing and sprouting constructions with monoclausal vs. biclausal antecedents in Experiment 3, I found that not only were sentences with longer antecedents judged more harshly, but there was an observable delay in processing as indexed by various fixation measures during eye tracking while reading. In certain models of retrieval, namely direct access accounts such as outlined in Martin and McElree (2008), this result is unexpected. In these types of models, retrieval times should not vary as a function of antecedent complexity or length, and this claim has been used to advocate for the absence of detailed structural

information in the memory representation of the antecedent retrieved during dependency resolution. The observations I make in Experiments 3 and 4 directly contradict this claim.

Experiment 3 also replicated the finding that sprouting is both dispreferred in offline judgments and disrupts processing in online measures. As noted before, this is not a novel finding, and has been explained as a penalty for having to amend retrieved material to include the structural source of the wh-element during ellipsis, as no correlate was included in the structural representation of the antecedent. However, these results provide corroborating support for the presence of syntactic structure in both the antecedent and the ellipsis site.

Chapter 3 also examined the interaction of the antecedent complexity effect witnessed in Experiment 3 and island constraints. The persistence of island effects in certain ellipsis contexts has long been a strong argument for the presence of structural information in elided material, and in Experiment 4 we see an effect of islands in a novel way. In this experiment I observed a main effect of antecedent length in offline measures, however in the online eye tracking portion of the experiment I observed an effect of length only during argument extraction. It appeared as if the parser was only retrieving the embedded clause in the case that retrieval of the biclausal structure containing the factive verb would result in an island violation. I attributed this effect to the notion of scope parallelism, such that the height of wh-element extracted out of the elided material was under pressure to mirror the scope that the correlate phrase was able to take in the antecedent clause. Again, we see an instance where the mechanism involved in retrieval is able to target certain structures, and that the structural content of that retrieval has a notable effect on processing times.

Finally, in Chapter 4 I turned to the interpretation of material used to resolve ellipsis, specifically in cases in which the correlate and the interpretation were non-isomorphic. The

investigation of cases such as strict and sloppy identity and vehicle change can help clarify the content of what is retrieved or linked to the dependent element as well as the possibility of processes or operations that can occur after such resolution. In Experiment 5, I examined whether in the absence of strong contextual pressure there is a preference for the strict vs. sloppy interpretation of VPE, and found a robust effect such that the strict interpretation was preferred compared to sloppy or outright contradictory interpretations. This result led to the hypothesis that binding information is part of the content of retrieval, and a prediction was made that the strict interpretation of VPE would show facilitation in online measures. However, in an eye tracking while reading experiment, it appeared that indexical information had a minimal and/or late effect on the processing of VPE. The online portion of this experiment also manipulated the gender match between the elided reflexive pronoun and the subject of the clause containing the ellipsis site. Several interesting points arose from this manipulation, notably the observation that a gender match effect surfaced in the strict interpretation cases, but a gender mismatch effect surfaced in the sloppy conditions. I took this as evidence that retrieval of the surface form of the reflexive initiated the search for a local antecedent, resulting in interference in the strict interpretation and a gender mismatch in the sloppy conditions. This combined with the weak and late evidence of facilitation in the strict conditions was taken to indicate that while indexical information may be part of the information accessed during retrieval, it does not impact initial structure building and interpretation.

Experiments 5 and 6 extended the examination of VPE to include cases of vehicle change, in which an R-expression within an ellipsis site appears to behave as a pronoun under ellipsis. The offline acceptability judgment task was ultimately deemed uninformative due to methodological concerns, however in the online experiment there appeared to be evidence for the

real-time computation of vehicle change. In the paradigm I employed, which involved both cataphora and VPE, a gender mismatch effect during pronoun resolution was predicted only in the case that vehicle change had indeed occurred when resolving the VPE. Such an effect was observed in fixation measures, and furthermore there was some general difficulty for sentences in which vehicle change was hypothesized to have transpired. While other alternatives were discussed, I ultimately made the argument that this general difficulty was likely indexing the cost of the performance of vehicle change itself.

The importance of parallelism

One underlying theme to many of the conclusions reached in this thesis regards the importance of parallelism in linguistic domains. The pervasiveness of parallelism is impressive, and we observe it here affecting material and relations both global and quite specific. In this dissertation parallelism is most overtly examined in Chapter 2, however we see its effect in Chapter 3 as well, and perhaps in some limited way in Chapter 4. In the realm of comprehension, parallelism appears to affect overall 'goodness' of an interpretation, expectations about the interpretation of material, and the processing of linguistic content. Its ubiquity certainly tells us something about cognition and the human language system. In broad and informal terms, the reuse of familiar representations is easier than the use of novel representations, and both the production and comprehension systems seem eager to take advantage of such a benefit (Bock, 1986; Branigan et al., 2005; Branigan, Pickering, McLean, & Stewart, 2006; Frazier & Duff, 2019; Thothathiri & Snedeker, 2008; Tooley & Traxler, 2010; Xiang et al., 2019, inter alia).

The source of a parallelism advantage has been modelled in various ways, most modernly as fallout from higher activation levels for recently processed or produced representations,

underlaid by baseline activation that increases as a result of frequency of activation. Simply attributing the prevalence of parallelism to high baseline activation, however, avoids the question of whether it is so very frequent as a result of any kind of special cognitive status. Regardless, the propensity for parallel elements can be leveraged to spotlight the type of material held in memory. In Chapter 2, this spotlight was able to illuminate the representation of a pronoun-antecedent relation, and demonstrate that this relationship exists as an element in its own right, with the ability to affect processing far downstream. It may be most appropriate to refer to this relation as the indices of coreference, but where exactly do they reside? I believe two main alternatives are available, that the indices live on the syntactic structure that is parsed, or that they are represented as features of the relevant chunks in memory. Given the results of Experiment 5 in which the binding information of a relationship in VPE did not have a strong effect on online processing despite the apparent presence of structural information, it may be more likely that this relation is represented featurally, such that when an antecedent is selected it is updated to reflect this relation. However, even in such a scenario the relative structural positions of these chunks must be referenced to compute the parallelism of dependencies across conjuncts.

Structural parallelism also surfaced strongly in Chapter 3, in which we saw a penalty associated with non-parallelism between the antecedent and the elided material (sprouting), however I might make the case that parallelism played a role in the sluicing sentences as well. One aspect of Experiments 3 and 4 that I did not discuss specifically in Chapter 3 was that when faced with two potential antecedents, one a larger biclausal structure and the other a monoclausal structure contained within that biclausal structure, participants seemed to retrieve the larger antecedent when available, as indexed by fixation times. However, given that there is apparent

difficulty associated with retrieving a larger, more complex structure why would the parser not opt for the simpler embedded clause? The answer here may lie in parallelism – parallelism is maximized when the larger structure is retrieved in that there is simply just more material to parallel.

We also see parallelism guiding the retrieval process in Experiment 4, where I reasoned that scope parallelism was responsible for retrieval of only the embedded clause in instances where a factive verb blocked covert raising of the adjunct clause in the antecedent. Scope parallelism is of course not my own novel invention, but its appearance in this experiment only reinforces the impression that parallelism has some degree of influence over a wide range of linguistic contexts. In this light, then, it is interesting to consider the results from the strict and sloppy experiments in Chapter 4. In the offline judgments we see a marked effect of interpretation, such that the strict interpretation is greatly preferred. But considering the parallelism of relations we saw in Chapter 1, we would predict quite the opposite: it is the sloppy interpretation that demonstrates identical relations in terms of the distance and relative structural positions of the pronoun and the antecedent. It may be that offline the retention of the binding information overrides a parallelism preference, although this effect is not so clear online. Indeed, during online processing it seemed that consideration of the local subject (the 'parallel relations' case) was pursued, perhaps before indexical information was available to the parser.

The importance of structure

The other critical conclusion reached in this thesis concerns the prevalence of structural information at all stages of comprehension. This conclusion is not meant to minimize the attention researchers have paid to the impact of structure in parsing, however with some few

exceptions widely accepted cue-based models have not focused on how to represent hierarchical structural information beyond the local environment, and studies investigating structure in dependency resolution often avoid explicitness when highlighting the precise mechanism of retrieval or representation of structure in memory. In this dissertation we see evidence that there is not only structure at the ellipsis site, but also in the memory representation of the processed material (Kim, Brehm, & Yoshida, 2019; Parker, 2017). Most often the claim regarding hierarchical structure in cue-based models is to minimize its primacy, leaning on evidence from grammaticality illusions and the time course of processing. In the previous chapters I have presented evidence that contradicts this claim – structure is respected during retrieval and resolution, and it has a noticeable effect on processing.

In my investigation of parallelism in pronoun resolution, we witness a hearty effect which most probably references the relative structural positions of the key elements. In these studies, discernment of the parallel antecedent required that it be distinguishable from the non-parallel antecedent, but in all cases the only distinguishing factor was whether the antecedent was the subject of the matrix or embedded clause. It is possible that a cue indicating something along the lines of [+/- matrix] could accomplish the task, however as previously noted if we move beyond one level of embedding this cue is no longer very informative. The alternative is to assume that the relation between the two items is stored in structural terms, and that when faced with the resolution of an additional pronoun that relation increases the likelihood of a structurally identical relation, regardless of whether one frames this as priming, parallelism, or increased activation levels.

The parallelism studies are corroborated by the observations made in sluicing and sprouting. It is conceivable that in the offline studies in Chapter 3, the dispreference for longer

antecedents is a simple result of overall length of the sentences, a factor which is known to affect judgments. However, no such explanation exists in the online experiments. In these studies I see a fairly clear effect of structural complexity on retrieval times, something which up to this point has not been witnessed in processing experiments, with the exception of the sprouting results of Dickey and Bunger (2011) . Potentially one could make the argument that it is simply the quantity (rather than complexity) of material to be retrieved that affects processing, however (1) direct-access models do not predict that length or amount of material should matter to the retrieval mechanism, and (2) a structural explanation seems more likely in consideration of other experiments in this thesis. Additional support comes from the factive island results, where the targeting of a particular structure to satisfy scope parallelism is rapid and shows no sign of reanalysis.

Finally, structural effects surface in Chapter 4, although perhaps in a more subtle fashion than in previous chapters. In the strict and sloppy interpretation experiments, I see a preference for the strict interpretation in offline judgments, but in the online experiment I fail to observe a strong effect of the original binding configuration. Instead, I see an effect of the surface morphological form of the retrieved pronoun – in all cases the isomorphic form is a reflexive, which appears to spur a search for a local antecedent. In the conditions in which a sloppy interpretation is encouraged I do not observe evidence which support an interference effect, and in fact observe quite the opposite. It is only when there is overriding pressure to consider the strict interpretation that a gender match (similarity-based interference) effect emerges. This is very much in line with other research that has demonstrated a structurally constrained search for reflexive resolution (e.g. Dillon et al., 2013).

The comprehension mechanism

What conditions do we need to place on the nature of memory representations used in language processing and how must our comprehension mechanism function to account for the observations and claims made in this thesis? One central claim here is that the output of parsing, the syntactic structure achieved by the parser after the assignation of category information and phrase structure rules, is not merely released after an interpretation is achieved, contrary to previous claims to this effect (e.g. Gibson, 1998, 2000; Martin & McElree, 2008). To account for the influence of particular structures and relations from previously processed material we must assume that a representation for this material has been encoded and stored in declarative memory. Furthermore, this representation must be linked in some fashion to the chunks containing featural information. To accommodate the presence of structural material in ellipsis as well as the influence of parallelism on relations we must also have some way to reactivate this structure.

In the ACT-R framework, grammatical knowledge is represented procedurally in a set of specific production rules, which are the essence of parsing (Lewis & Vasishth, 2005). As elements are encountered during comprehension, they are held in terms of value attribute pairs in buffers where they can interact with procedural rules and elements retrieved from declarative memory. As more input is encountered, elements in buffers are passed to declarative memory and the chunks and their corresponding features can be stored for future use. Importantly, in this architecture the chunks are 'flat', they do not contain other chunks, they can only reference them. While there is substantial evidence to support these memory representations as feature bundles, it is obvious that they themselves *cannot* encode hierarchical information (beyond the local context) that we see evidence of in this thesis, among other work. One possibility is to assume

that larger structures are maintained during the encoding of this material, and that these chunks are linked to positions of this larger structure during processing, as diagrammed below in Figure 17. After completion of the current parse, the entire structure is passed to declarative memory, and the positions in the tree are connected to the feature bundles in some fashion, perhaps something like indices. This type of proposal suffers from the issue of what kind of boundedness is imposed on the size of the structure we are able to maintain, something discussed in more detail on pages 207-208.

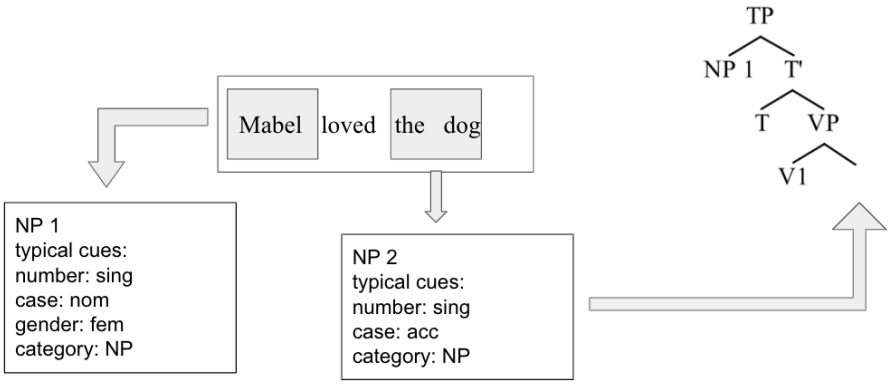


Figure 17: Diagram of how items are encoded and linked to a (simplified) syntactic tree. As material is parsed, the feature bundles are linked to positions in the syntactic tree. This tree with indices (or some linking function) corresponding to the chunks is then passed to declarative memory.

Another possible avenue which avoids the need for separate storage of syntactic structure altogether would be a process of reparsing the retrieved information. Although linearization information is not explicitly encoded in feature bundles, one might imagine a scenario in which activation of the relevant chunks is passed to the parser for reconfiguration, resulting in a reconstruction of the original structure. This process could be aided by the recent use of certain procedural rules, in essence a priming of phrase structure rules.

There is an attractive quality to this proposal – we avoid the creation of new memory representations that must be linked to other representations, and we account for the antecedent complexity effects observed in retrieval as a byproduct of additional structure building. However, we are faced with a question of parsimony: is it simpler to store and retrieve the product of parsing, or is it easier to reparse? Which is more error prone? The answer to this first question is more open to speculation, however considering the impressive capacity of human memory and the propensity to reuse information when possible, storage and retrieval might seem the more likely option. Furthermore, while clearly the retrieval of structural information could be vulnerable to similar errors as we see in other domains, reparsing material could easily introduce new errors related to the unremitting ambiguity of human language. This has in fact been investigated in some areas, for example Paape et al. (2018) do *not* observe evidence that comprehenders are (re)garden-pathed at an ellipsis site after resolving the structural ambiguity in the antecedent clause, a result that is incongruous with a reparsing account.

Another way to think about how structural information can be stored and subsequently activated is a parallel chunking system that encodes structural and linearization information for items passed from the buffer to declarative memory. This idea has been explored as well, although in the realm of priming in language production, by Reitter et al. (2011). In this work, the authors present a computational model of syntactic priming within the ACT-R architecture. The main innovation from the model detailed by Lewis and Vasishth (2005) is to shift grammatical knowledge from procedural memory to declarative memory in form of a combinatory categorial grammar (CCG) which allows for more complex syntactic types (including subcategorization frames) than traditional syntactic categories. These syntactic types are represented as ‘syntax chunks’ which are accessed via spreading activation from a lexical

item in a buffer. Although in their proposal they are modeling production and therefore speak little of how these pieces of information may be stored in memory as a function of parsing and comprehension, something akin to this type of idea may allow for the inclusion of structural information associated with the more 'conventional' chunks commonly referenced. If items in a buffer are passed to declarative memory as two linked elements, one which contains the traditional set of features and one which contains structural and linearization information, we have provided a method of representing hierarchical information. In this scenario, activation of a chunk would trigger activation of its paired chunk, and both types of information would be available to the parser.

The theory above is conjecture, as most theories are, however some system like this might be able to explain a number of phenomena described in this thesis and elsewhere. Furthermore, it addresses some issues that would arise when assuming structural output is stored in memory. One such issue is the size of the structural representation stored – at what point does one storable structure end and another begin? One could propose something along the lines of a clause, however given the recursive nature of language this kind of boundedness could quickly become unwieldy. Without direct evidence from some source, any division proposed is somewhat arbitrary, however the architecture of cue-based models give us some discreteness: an element that enters into relations with other elements. In the case of my parallelism results, there may be two ways to think about facilitation for pronoun resolution under this proposal. The repetition of structural information activates the structures previously used in the sentence, and the structure associated with the matrix subject, for example, would have a higher level of activation compared to the local subject if it had been reactivated in the previous conjunct due to dependency resolution. Essentially (and wholly expected under cue-based models) retrieval of an

element boosts its activation level, only now that retrieval also boosts the structural information activation as well. An alternative to this is that indices live on the structural representation and are syntactic in nature. The reactivation of structural information includes indexical information, and resolution to a node that hosted a previous index is facilitated.

The storage component very obviously helps explain the presence of structural information in ellipsis (Gibson, 1998; Kaan & Stowe, 2002; Wanner & Maratsos, 1978), however any theory of storage falls short of explaining the arrival of that information in the ellipsis site. As discussed, our three main options are sharing, copying, and rebuilding. I have already argued against the rebuilding idea (although it does retain a certain appeal), however sharing vs. copying has been difficult to discern in past experiments. The observations in Chapter 3, however, would rule in favor of a copying account. If simple reactivation of this information was sufficient, there should be no cost associated with larger structures. If instead the parser reintegrates the structural information into the ellipsis site, then the cost I witness can be explained. Referring to the final set of experiments, this model is not strongly contradicted or confirmed, except in the sense that there appeared to a primacy for structural information in the strict and sloppy experiment, and the retrieval of the reflexive seemed to initiate a structurally informed search regardless of whether that interpretation would ultimately be incongruous.

Some Remaining Issues and Musings

Parallelism and maintenance

There are some remaining topics to discuss, that are in ways more nebulous than the previous discussions, and illuminate some of the difficulties of making explicit statements about the nature of the memory representations involved in sentence processing and memory retrieval. The

first of these concerns the topic of parallelism, and how the parser creates parallel structure and calculates parallelism between the current structure and the one that has been previously processed. It is clear that no parallelism can hold between two elements until the second element exists, and so it can be stated that parallelism is dependent on the second element and its comparison to the first. The propensity for parallelism in comprehension thus emerges during the parsing of the second element, and would appear to involve a heavily top-down parsing process that is predictive and hypothesis driven. If the opposite were true, and the parser merely evaluated the left corner and waited for bottom-up information before building a structural representation then it is unclear how parallelism effects would arise in parsing.

But upon what information are these hypotheses made? The most viable answer is that the structure built in the first conjunct is held in working memory and highly specific hypotheses are made on the basis of this structure. However, there is a classic question embedded in this explanation – how much structure can be held? The answer to this may lay in more classic proposals of the working memory system (e.g. Baddeley & Hitch, 1974; Just & Carpenter, 1992; Miller & Chomsky, 1963), and it might be that the complexity of the first conjunct used in Experiments 1 and 2 was 'simple' enough to maintain (Hale, 2016). This issue is in some ways orthogonal to the question of how structural information is stored for later retrieval, however it is at odds with cue-based models in which the focus of working memory is extremely limited, consisting of a single chunk buffer. Upon encountering a connective like 'and' or 'whereas', presumably the entire contents of the first conjunct are now represented as a single chunk that enters into a relation with this connective, and resets the local control state. Although the first conjunct chunk is not entirely opaque, its internal hierarchical structure is. How predictive structure building follows from the information provided by this chunk is similarly opaque.

While unrelated to the issue of parallelism, the experiments examining vehicle change in Chapter 4 also point to the requirement of some sort of maintenance component. The construction of these experiments was such that two dependent elements are immediately encountered in the first clause, and are followed (rather than preceded) by the elements that will ultimately resolve them. I have framed these 'backwards' dependencies as initiating active searches for their controlling elements, and it is unclear how this would proceed if material from the first clause was not maintained in working memory until dependency resolution was complete. If instead the material in the first clause was chunked away into declarative memory, retrieval would be necessary in order to complete both the pronoun and VPE dependencies. Retrieval is triggered – encountering a VP in the second clause is itself not a trigger for retrieval. Without the presence of an open dependency awaiting eagerly to be resolved in a buffer I see no plausible way in which a successful interpretation of these sentences is achieved.

Dependencies

On a final note, I would like to touch on the issue of dependencies and dependency formation in general. Throughout this dissertation I have deployed the term 'dependency' in a fairly free and easy manner, referring in this way to the relations made via coreference, links made between ellipsis sites and their antecedents, and the connections made by wh-elements and their gaps. Are these diverse constructions instances of the same phenomenon, and are they formed and represented in the same way? In a very general sense, dependencies can be thought of as an association between two (or more) elements that is necessary in order to achieve an interpretation of a sentence. In this way, yes, there is a common denominator to these constructions.

In another way, however, these types of 'dependencies' are quite different. In essence, are they a product or a part of building a syntactic parse, or do they involve secondary processes? In the case of pronoun resolution the answer is clear - a perfectly acceptable parse can occur in the absence of a relation between a pronoun and an antecedent. Interpretation may ultimately fail if a link to an antecedent for the pronoun is not made, however this action of 'linking' occurs on top of a syntactic parse. In opposition, wh-filler gap dependencies are formed by syntactic operations on syntactic objects, for example chain creating movement of a wh-element to a higher position in the syntactic tree. In terms of the dependencies created in the ellipsis experiments of Chapter 3, I assume a standard 'move and delete' approach, where a wh-element has been extracted out of full-fledged structure which hosts the foot of the movement chain and the wh-element survives as a remnant after deletion. Besides this structurally defined wh-dependency, the very argument I make that there is structure present at the ellipsis site makes the resolution of ellipsis a structure building/copying operation, and therefore part of building a parse.

In terms of processing, we may want to consider whether the same mechanism is at work when resolving and comprehending these different dependency types. There is crucially the matter of forward predictive processing, such as that which occurs in backwards anaphora and in wh-filler gap dependencies, as opposed to a backwards search which must examine previously processed material. These seem to be two very different processes. Furthermore, we may want to give import to dependencies which are subject to syntactic constraints. For example, we do not appear to observe islands which block coreference for pronouns (with perhaps the exception that the antecedent cannot locally bind it), however indeed see numerous instances in which a wh-filler gap dependency cannot span over particular syntactic configurations. Ultimately, however,

if resolution requires retrieval from memory, then I assume that the same mechanism is employed to locate the relevant information. Similarly, if the process involves maintaining a dependent element in expectation of encountering its controlling element, then I believe the same mechanism applies whether the dependent element is a wh-element or a pronoun. However, I acknowledge that while the same mechanisms may be involved for different dependency types, that does not mean that the relevant information is the same. For example, I do not presume that during pronoun resolution that the complexity or size of the antecedent material has any import.

What More Do We Need?

I have presented some speculations on the nature of the structural memory representation that is stored after parsing, however I would also like to note some alternatives, difficulties, and dead-ends that I have encountered during my consideration of this issue. The crucial question here is what additional information must be encoded to account for successful retrieval of items that rely on structural information to be disambiguated from their competitors, and in the case of ellipsis what is needed for the parser to integrate the correct hierarchical structure post-retrieval? Models such as ACT-R already give us quite a lot to work with regarding representations: phi-features, syntactic category and local syntactic context, perhaps even if a clause has been embedded or not. We have seen, however, that this is probably not enough.

One option, as noted, is that a larger structural object is stored, and that positions in this structure are linked to the corresponding chunks. Something akin to this is what many accounts seem to implicitly assume, however problems arise. Where does one structure end and another begin? In the case of complex and multi-clausal sentences how might the relations between structures be represented? How are unbounded dependencies handled in such a framework?

These questions may be answerable, and perhaps this is indeed how things work, but we see conflicting evidence for the presence of such large scale structures, for example the different patterns of interference between various dependencies such as agreement and reflexives, and an overall lack of evidence for serial tree traversing search. Conversely, a larger structural representation of this type accounts for the presence of detailed and accurate hierarchical information at the ellipsis site, and also is compatible with observations that indicate that ambiguity resolution in the antecedent is retained in dependent element.

The larger of the issues noted above pertains to how unbounded dependencies are resolved, like those that have been investigated in this thesis. This is a longstanding issue, and in general is problematic for most accounts. Given the potential for unlimited material between an item and its dependent element, how can we represent, locate, and link these elements together? This is course part of what models of parsing are trying to capture, and cue-based models give us a way to represent an unlimited number of items after processing, but as extensively noted in this thesis are unable to solve this problem when the distinguishing information is based on hierarchical structure and relations. The problem was elegantly stated by Berwick and Weinberg (1984, pp. 153-154):

> *"Because the material between an antecedent and its trace can be arbitrarily long, how can the machine store all the context to the left of the token it is currently analyzed? This problem can be solved if we can guarantee that the left context required for parsing decisions is finitely representable."*

If we abandon the notion of an overarching structural representation, then we must decide how to divvy up the output of parsing. I have suggested something of the same size as those in the model of Lewis and Vasishth (2006), a phrasal constituent, however if that is the case there is not an obvious reason to separate the chunks at all – rather the relevant information can be

encoded as an additional feature on the 'original' chunk. But what is this information? Whatever it is, it must be sufficient to allow for the identification of otherwise structurally identical chunks, e.g. possibly the matrix vs. embedded subject position in Experiments 1 and 2 and the matrix vs. the embedded CP in Experiments 3 and 4.[16] Furthermore, following activation it must be sufficient to allow for the accurate copying or reassembly of material for the purposes of ellipsis resolution. I believe the most obvious choice is something that references order, but this too has problematic aspects. Purely serial order information of words seems unlikely, however including order information referencing the phrasal node could get us farther, and that combined with information about the local syntactic environment could certainly aid in recreating hierarchical material. However, if my observations regarding parallelism are generated in a cue-based architecture and not as an effect of maintaining a larger structure in working memory, then order information referencing XPs would be largely unhelpful – how would previous activation of DP[2] in the specifier of IP in the first conjunct facilitate activation to DP[10] rather than DP[13] in the second conjunct, for example? Perhaps the numbering is more specific, something like 'subject[1]', however then we are presented with the problem of needing to access the current state of number for a given category, a relational and retrieval problem in and of itself.

Ending with open questions is a hallmark of much research, and I hope that this assembly of remarks does not confuse more than it elucidates.

Conclusion

[16] As noted, a +/- matrix feature would be sufficient for these experiments, however I have a strong intuition that similar effects could be elicited in structure with more than one level of embedding.

The array of interest and approaches to the retrieval mechanism employed during the resolution of long-distance dependencies in recent research endeavors has provided critical evidence, and various theories have opened wide avenues of investigation in many domains. In this thesis I have shown that while we should not abandon the advances made by feature-based models, we should also not abandon the complex and informative output of the parsing process itself: structural material. An opportunistic system would make use of such a rich source of information, and such information working in tandem with other features commonly assumed to be accessible to the resolution process creates an optimal environment for the successful resolution and interpretation of pervasive linguistic phenomena.

References

Abrusán, M. (2007). *Contradiction and grammar: The case of weak islands.* Massachusetts Institute of Technology.

Abrusán, M. (2011). Presuppositional and negative islands: a semantic account. *Natural Language Semantics, 19*(3), 257-321

Abrusán, M. (2014). *Weak island semantics* (Vol. 3): OUP Oxford.

Adams, M. (1985). Government of empty subjects in factive clausal complements. *Linguistic Inquiry, 16*(2), 305-313

Altmann, G., Henstra, J., & Garnham, A. (1993). Processing coordinated structures *Poster presented at the Sixth CUNY Conference on Human Sentence Processing, Amherst, Massachusetts*.

Ambridge, B., & Goldberg, A. E. (2008). The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics, 19*(3), 357-389

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American psychologist, 51*(4), 355

Aoshima, S., Phillips, C., & Weinberg, A. (2002). Active filler effects and reanalysis: A study of Japanese Wh-scrambling constructions. *University of Maryland working papers in linguistics, 12*, 1-24

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of memory and language, 51*(1), 23-54

Apel, J., Knoeferle, P., & Crocker, M. W. (2007). Processing parallel structure: Evidence from eye tracking and a computational model *In Proceedings of the 2nd European cognitive science conference* (pp. 125-131): Taylor & Francis London.

Arai, M., Van Gompel, R. P., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive psychology, 54*(3), 218-250

Arregui, A., Clifton Jr, C., Frazier, L., & Moulton, K. (2006). Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of memory and language, 55*(2), 232-246

Asher, N., & Vieu, L. (2005). Subordinating and coordinating discourse relations. *Lingua, 115*(4), 591-610

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language, 59*(4), 390-412

Baddeley, A. D., & Hitch, G. (1974). Working memory *Psychology of learning and motivation* (Vol. 8, pp. 47-89): Elsevier.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 748

Beaver, D. (2010). 3: Have You Noticed that Your Belly Button Lint Color is Related to the Color of Your Clothing? *Presuppositions and discourse: Essays offered to Hans Kamp* (pp. 65-100): Brill.

Beck, S., & Tiemann, S. (2018). Towards a model of incremental composition *Proceedings of Sinn und Bedeutung* (Vol. 21, pp. 143-162).

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 317

Berwick, R. C., & Weinberg, A. (1984). The grammatical basis of linguistic performance: Cambridge, MA: MIT Press.

Bock, J. K. (1986). Meaning, sound, and syntax: lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(4), 575

Branigan, H. P., Pickering, M. J., & McLean, J. F. (2005). Priming prepositional-phrase attachment during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(3), 468

Branigan, H. P., Pickering, M. J., McLean, J. F., & Stewart, A. J. (2006). The role of local and global syntactic structure in language production: Evidence from syntactic priming. *Language and cognitive processes, 21*(7-8), 974-1010

Büring, D. (2005). *Binding theory*: Cambridge University Press.

Carlson, K., Dickey, M. W., & Kennedy, C. (2005). Structural economy in the processing and representation of gapping sentences. *Syntax, 8*(3), 208-228

Cattell, R. (1976). Constraints on movement rules. *Language*, 18-50

Chambers, C. G., & Smyth, R. (1998). Structural parallelism and discourse coherence: A test of centering theory. *Journal of Memory and Language, 39*(4), 593-608

Chomsky, N. (1972). Some Empirical Issues. *Chomsky, Studies in Semantics in Generative Grammar, Mouton, The Hague*

Chomsky, N. (1973). Conditions on transformations. *A festschrift for Morris Halle*

Chomsky, N. (1977). On wh-movement. *Formal syntax*, 71-132

Chomsky, N. (1981a). *Lectures on Government and Binding*. Berlin, Federal Republic of Germany: Mouton de Gruyter.

Chomsky, N. (1981b). Lectures on Government and Binding, Foris, Dordrecht. *ChomskyLectures on Government and Binding1981*

Chomsky, N. (1986). *Barriers* (Vol. 13): MIT press.

Christensen, K. R., Kizach, J., & Nyvad, A. M. (2013). Escape from the island: Grammaticality and (reduced) acceptability of wh-island violations in Danish. *Journal of Psycholinguistic Research, 42*(1), 51-70

Christensen, K. R., & Nyvad, A. M. (2014). On the nature of escapable relative islands. *Nordic Journal of Linguistics, 37*(1), 29-45

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive psychology, 42*(4), 368-407

Chung, S. (2013). Syntactic identity in sluicing: How much and why. *Linguistic Inquiry, 44*(1), 1-44

Chung, S., Ladusaw, W. A., & McCloskey, J. (1995). Sluicing and logical form. *Natural language semantics, 3*(3), 239-282

Cinque, G. (1990). *Types of Ā-dependencies*: MIT press Cambridge, MA.

Clark, H. H., & Sengul, C. (1979). In search of referents for nouns and pronouns. *Memory & Cognition, 7*(1), 35-41

Clifton Jr, C., & Ferreira, F. (1987). Discourse structure and anaphora: Some experimental results

Clifton, L. F. C. (1998). Comprehension of sluiced sentences. *Language and Cognitive Processes, 13*(4), 499-520

Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research, 19*(4), 245-264

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes, 28*(1-2), 188-219

Cunnings, I., Patterson, C., & Felser, C. (2014). Variable binding and coreference in sentence comprehension: Evidence from eye movements. *Journal of Memory and Language, 71*(1), 39-56

Dalrymple, M., Shieber, S. M., & Pereira, F. C. N. (1991). Ellipsis and Higher-Order Unification. *Linguistics and Philosophy, 14*(4), 399-452

De Cuba, C. (2006). The adjunction prohibition and extraction from non-factive CPs *Proceedings of the 25th West Coast Conference on Formal Linguistics* (pp. 123-131): Citeseer.

Deane, P. (1991). Limits to attention: A cognitive theory of island phenomena. *Cognitive linguistics, 2*(1), 1-64

Derrick, D., & Archambault, D. (2010). TreeForm: Explaining and exploring grammar through syntax trees. *Literary and linguistic computing, 25*(1), 53-66

Dickey, M. W., & Bunger, A. C. (2011). Comprehension of elided structure: Evidence from sluicing. *Language and Cognitive Processes, 26*(1), 63-78

Dillon, B. (2011). *Structured access in sentence comprehension.*

Dillon, B. (2014). Syntactic memory in the comprehension of reflexive dependencies: an overview. *Language and Linguistics Compass, 8*(5), 171-187

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language, 69*(2), 85-103

Dubey, A., Keller, F., & Sturt, P. (2008). A Probabilistic Corpus-based Model of Syntactic Parallelism. *Cognition, 109*(3), 326-344

Dubey, A., Sturt, P., & Keller, F. (2005). Parallelism in coordination as an instance of syntactic priming: Evidence from corpus-based modeling *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 827-834): Association for Computational Linguistics.

Ehrlich, K. (1980). Comprehension of pronouns. *The Quarterly Journal of Experimental Psychology, 32*(2), 247-255

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of verbal learning and verbal behavior, 22*(1), 75-87

Elbourne, P. (2001). E-type anaphora as NP-deletion. *Natural Language Semantics, 9*(3), 241-288

Elman, J. L., Kehler, A., & Rohde, H. (2006). Event structure and discourse coherence biases in pronoun interpretation *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 28).

Erteschik-Shir, N. (1973). *On the nature of island constraints.* Massachusetts Institute of Technology.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language, 47*(2), 250-272

Fiengo, R., & May, R. (1994). *Indices and identity* (Vol. 24): MIT press.

Fodor, J. D. (1989). Empty categories in sentence processing. *Language and Cognitive Processes, 4*(3-4), SI155-SI209

Fodor, J. D. (1993). Processing empty categories: A question of visibility *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 351-400): Lawrence Erlbaum East Sussex.

Fox, D. (2000). *Economy and semantic interpretation* (Vol. 35): MIT press.

Franck, J., Colonna, S., & Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in psychology, 6*, 349

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and cognitive processes, 17*(4), 371-404

Frazier, L. (2018). Ellipsis and Psycholinguistics *The Oxford Handbook of Ellipsis*.

Frazier, L., & Clifton, C. (2000). On bound variable interpretations: The LF-only hypothesis. *Journal of Psycholinguistic Research, 29*(2), 125-140

Frazier, L., & Clifton, C. (2006). Ellipsis and discourse coherence. *Linguistics and Philosophy, 29*(3), 315-346

Frazier, L., & Clifton, J., Charles. (2001). Parsing coordinates and ellipsis: Copy α. *Syntax, 4*(1), 1-22

Frazier, L., & Clifton Jr, C. (2005). The syntax-discourse divide: processing ellipsis. *Syntax, 8*(2), 121-174

Frazier, L., & Duff, J. (2019). Repair or accommodation? Split antecedent ellipsis and the limits of repair. *Glossa: a journal of general linguistics, 4*(1)

Frazier, L., Munn, A., & Clifton, C. (2000). Processing coordinate structures. *Journal of Psycholinguistic Research, 29*(4), 343-370

Frazier, L., Taft, L., Roeper, T., Clifton, C., & Ehrlich, K. (1984). Parallel structure: A source of facilitation in sentence comprehension. *Memory & cognition, 12*(5), 421-430

Frazier, M. (2015). Morphological recoverability in Gapping. *PhD diss., Northwestern University*

Frazier, M., Ackerman, L., Baumann, P., Potter, D., & Yoshida, M. (2015). Wh-filler-gap dependency formation guides reflexive antecedent search. *Frontiers in psychology, 6*, 1504

Gazdar, G. (1981). Unbounded dependencies and coordinate structure *The Formal Complexity of Natural Language* (pp. 183-226): Springer.

Geis, M. L. (1970). *Adverbial subordinate clauses in English.* Massachusetts Institute of Technology.

Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of memory and language, 27*(6), 699-717

Gernsbacher, M. A., Hargreaves, D. J., & Beeman, M. (1989). Building and accessing clausal representations: The advantage of first mention versus the advantage of clause recency. *Journal of memory and language, 28*(6), 735-755

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1-76

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain, 2000*, 95-126

Gibson, E., & Warren, T. (2004). Reading Time Evidence for Intermediate Linguistic Structure in Long-Distance Dependencies. *Syntax, 7*(1), 55-78

Ginzburg, J., & Sag, I. (2000). *Interrogative investigations*: Stanford: CSLI publications.

Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive science, 22*(4), 389-424

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology-Learning Memory and Cognition, 27*(6), 1411-1423. doi: Doi 10.1037//0278-7393.27.6.1411

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*(1), 97-114. doi: 10.1016/j.jml.2004.02.003

Grant, M., Sloggett, S., & Dillon, B. (2020). Processing ambiguities in attachment and pronominal reference. *Glossa: a journal of general linguistics, 5*(1)

Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational linguistics, 21*(2), 203-225

Hale, J. (2016). Information-theoretical complexity metrics. *Language and Linguistics Compass, 10*(9), 397-412

Hall, K., & Yoshida, M. (2020). Coreference and parallelism. *Language, Cognition and Neuroscience*, 1-24

Hardt, D. (1993). Verb phrase ellipsis: Form, meaning, and processing

Harris, J. A. (2015). Structure modulates similarity-based interference in sluicing: An eye tracking study. *Frontiers in psychology, 6*, 1839

Hestvik, A. (1995). Reflexives and ellipsis. *Natural Language Semantics, 3*(2), 211-237

Hobbs, J. R. (1979). Coherence and coreference. *Cognitive science, 3*(1), 67-90

Hofmeister, P., Jaeger, T. F., Sag, I. A., Arnon, I., & Snider, N. (2007). Locality and accessibility in wh-questions. *Roots: Linguistics in search of its evidential base*, 185-206

Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language, 86*(2), 366

Hooper, J. B. (1975). On assertive predicates *Syntax and Semantics volume 4* (pp. 91-124): Brill.

Huddleston, R., & Pullum, G. (2005). The Cambridge grammar of the English language. *Zeitschrift für Anglistik und Amerikanistik, 53*(2), 193-194

Hudson, R. A. (1980). Constituency and Dependency. *Linguistics, 18*, 179-198

Hudson, R. A. (1993). Recent Developments in Dependency Theory. In J. J. a. A. v. S. a. W. S. a. T. Vennemann (Ed.), *Syntax: Ein Internationales Handbuch Zeitgenossischer Forschung [An International Handbook of Contemporary Research* (pp. 329-338). Berlin: de Gruyter.

Jakobson, R. (1957). *Shifters, verbal categories and the Russian verb*: Harvard University, Department of Slavic Languages and Literatures, Russian ….

Jasinskaja, K., & Karagjosova, E. (2015). Rhetorical relations. *The companion to semantics. Oxford: Wiley*

Johnson, K. (2001a). Sluicing and constraints on quantifier scope. *Glot International, 5*(2)

Johnson, K. (2001b). *What VP ellipsis can do, and what it can't, but not why*: Citeseer.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review, 99*(1), 122

Kaan, E., Overfelt, C., Tromp, D., & Wijnen, F. (2013). Processing gapped verbs. *Journal of psycholinguistic research, 42*(4), 307-338

Kaan, E., & Stowe, L. (2002). Storage and computation in sentence processing *Storage and computation in the language faculty* (pp. 257-295): Springer.

Kaan, E., Wijnen, F., & Swaab, T. Y. (2004). Gapping: Electrophysiological evidence for immediate processing of "missing" verbs in sentence comprehension. *Brain and Language, 89*(3), 584-592

Kaplan, D. (1989). Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals.

Karttunen, L. (1971). Implicative verbs. *Language*, 340-358

Kazanina, N., Lau, E., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The Effect of Syntactic Constraints on the Processing of Backwards Anaphora *Journal of Memory and Language, 56*, 384-409

Keenan, E. L. (1971). Names, quantifiers, and the sloppy identity problem. *Research on Language & Social Interaction, 4*(2), 211-232

Kehler, A. (1993). A discourse copying algorithm for ellipsis and anaphora resolution *Sixth Conference of the European Chapter of the Association for Computational Linguistics*.

Kehler, A. (2000). Coherence and the resolution of ellipsis. *Linguistics and Philosophy, 23*(6), 533-575

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*: CSLI publications Stanford, CA.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2007). Coherence and Coreference Revisited. *Journal of Semantics, 25*, 1-44

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics, 39*(1-2), 1-37

Kertz, L., Kehler, A., & Elman, J. (2006). Grammatical and coherence-based factors in pronoun interpretation *Proceedings of the 28th annual conference of the Cognitive Science Society* (pp. 1605-1610): Lawrence Erlbaum Associates Mahwah, NJ.

Kim, C. (2006a). Structural and thematic information in sentence production *Proceedings of the 37th Annual Meeting of the North East Linguistic Society* (pp. 59-72): Citeseer.

Kim, C. S. (2006b). *Structural priming and non-surface representations.* University of California, Los Angeles.

Kim, C. S., Carbary, K. M., & Tanenhaus, M. K. (2014). Syntactic priming without lexical overlap in reading comprehension. *Language and speech, 57*(2), 181-195

Kim, C. S., Kobele, G. M., Runner, J. T., & Hale, J. T. (2011). The acceptability cline in VP ellipsis. *Syntax, 14*(4), 318-354

Kim, N., Brehm, L., Sturt, P., & Yoshida, M. (2019). How long can you hold the filler: Maintenance and Retrieval. *Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2019.1626456*

Kim, N., Brehm, L., & Yoshida, M. (2018). The online processing of noun phrase ellipsis and mechanisms of antecedent retrieval. *Language, Cognition and Neuroscience, 34*(2), 190-213

Kim, N., Brehm, L., & Yoshida, M. (2019). The online processing of noun phrase ellipsis and mechanisms of antecedent retrieval. *Language, Cognition and Neuroscience, 34*(2), 190-213

Kim, N., Carlson, K., Dickey, M., & Yoshida, M. (2020). Processing gapping: Parallelism and grammatical constraints. *Quarterly Journal of Experimental Psychology, 73*(5), 781-798

Kimball, J. P. (1973). *The formal theory of grammar*. Englewood Cliffs, N.J.,: Prentice-Hall.

Kiparsky, P., & Kiparsky, C. (1970). Fact. *Progress in linguistics*, 143-147

Kitagawa, Y. (1991). Copying identity. *Natural Language & Linguistic Theory, 9*(3), 497-536

Kluender, R., & Kutas, M. (1993a). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*(2), 196-214

Kluender, R., & Kutas, M. (1993b). Subjacency as a processing phenomenon. *Language and cognitive processes, 8*(4), 573-633

Knoeferle, P. (2007). Eye tracking parallelism effects in coordinate constructions: The effects of “and”,“but” and “while”. *Poster presented at the 20th Annual CUNY. La Jolla, USA*

Knoeferle, P. (2014). Conjunction meaning can modulate parallelism facilitation: Eye-tracking evidence from German clausal coordination. *Journal of Memory and Language, 75*, 140-158

Knoeferle, P., & Crocker, M. W. (2009). Constituent order and semantic parallelism in online comprehension: Eye-tracking evidence from German. *The Quarterly Journal of Experimental Psychology, 62*(12), 2338-2371

Knott, A., & Sanders, T. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of pragmatics, 30*(2), 135-175

Konieczny, L. (2000). Locality and parsing complexity. *Journal of psycholinguistic research, 29*(6), 627-645

Konieczny, L. (2005). The psychological reality of local coherences in sentence processing *Proceedings of the 27th annual conference of the cognitive science society* (pp. 1178-1183).

Kratzer, A., & Heim, I. (1998). *Semantics in generative grammar* (Vol. 1185): Blackwell Oxford.

Kroch, A. (1989). Amount quantification, referentiality and long wh-movement

Kuno, S., & Li, C. N. (1976). Subject, Theme, and the Speaker's Empathy. A Reexamination of Relativization Phenomena in Subject and Topic

Kush, D., Lidz, J., & Phillips, C. (2015). Relation-sensitive retrieval: evidence from bound variable pronouns. *Journal of Memory and Language, 82*, 18-40

Kush, D., Lohndal, T., & Sprouse, J. (2018). Investigating variation in island effects. *Natural Language & Linguistic Theory, 36*(3), 743-779. doi: 10.1007/s11049-017-9390-z

Kush, D., Omaki, A., & Hornstein, N. (2013). 11 Microvariation in islands? *Experimental syntax and island effects*, 239

Kwon, N. (2008). *Processing of syntactic and anaphoric gap-filler dependencies in Korean: Evidence from self-paced reading time, ERP and eye-tracking experiments.* UC San Diego.

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language, 82*, 133-149

Lappin, S. (1992). *The syntactic basis of ellipsis resolution*: IBM Thomas J. Watson Research Division.

Larson. (1995). Semantics. In Gleitman, Lieberman & Osherson (Eds.), *Invitation to Cognitive Science: Language* (Vol. 1, pp. 361-380): The MIT Press.

Lasnik, H. (1995). Verbal morphology: Syntactic structures meets the Minimalist Program. *Evolution and revolution in linguistic theory: Studies in honor of Carlos P. Otero*, 251-275

Lasnik, H. (2001). When can you save a structure by destroying it? *PROCEEDINGS-NELS* (Vol. 31, pp. 301-320).

Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1545

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science, 29*(3), 375-419

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*(10), 447-454. doi: 10.1016/j.tics.2006.08.007

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language, 58*(3), 879-906

Martin, A. E., & McElree, B. (2011). Direct-access retrieval during sentence comprehension: evidence from sluicing. *Journal of memory and language, 64*(4), 327-343

Mauner, G., Tanenhaus, M. K., & Carlson, G. N. (1995). A note on parallelism effects in processing deep and surface verb-phrase anaphora. *Language and Cognitive Processes, 10*(1), 1-12

McCawley, J. D. (1988). *The syntactic phenomena of English*. Chicago: University of Chicago Press.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of psycholinguistic research, 29*(2), 111-123

McElree, B. (2006). Accessing recent events. *Psychology of learning and motivation, 46*, 155-200

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: the time course of recognition. *Journal of Experimental Psychology: General, 118*(4), 346

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension (Vol. 48, pp. 67-91).

Melvold, J. (1991). Factivity and definiteness. *MIT working papers in linguistics, 15*, 97-117

Merchant, J. (2001). *The syntax of silence: Sluicing, islands, and the theory of ellipsis*: Oxford University Press on Demand.

Merchant, J. (2008). An asymmetry in voice mismatches in VP-ellipsis and pseudogapping. *Linguistic Inquiry, 39*(1), 169-179

Merchant, J. (2013). Voice and ellipsis. *Linguistic Inquiry, 44*(1), 77-108

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users

Murphy, G. L. (1985). Processes of understanding anaphora. *Journal of Memory and Language, 24*(3), 290-303

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*(3), 251-269

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of psycholinguistic research, 18*(1), 5-19

Nicol, J. L. (1988). *Coreference processing during sentence comprehension.* Massachusetts Institute of Technology.

Nicol, J. L., Fodor, J. D., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigation sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1229-1238

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of memory and language, 55*(4), 601-626

Orth, W., Yoshida, M., & Sloggett, S. (2020). Negative polarity item (NPI) illusion is a quantification phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*

Oshima, D. Y. (2006). On factive islands: pragmatic anomaly vs. pragmatic infelicity *Annual Conference of the Japanese Society for Artificial Intelligence* (pp. 147-161): Springer.

Paape, D., Hemforth, B., & Vasishth, S. (2018). Processing of ellipsis with garden-path antecedents in French and German: Evidence from eye tracking. *PloS one, 13*(6), e0198620

Paape, D., Nicenboim, B., & Vasishth, S. (2017). Does antecedent complexity affect ellipsis processing? An empirical investigation. *Glossa: a journal of general linguistics, 2*(1)

Parker, D. (2017). Processing multiple gap dependencies: Forewarned is forearmed. *Journal of Memory and Language, 97*, 175-186

Parker, D. (2018). A memory-based explanation of antecedent-ellipsis mismatches: New insights from computational modeling. *Glossa: a journal of general linguistics, 3*(1)

Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language, 94*, 272-290. doi: https://doi.org/10.1016/j.jml.2017.01.002

Partee, B. H. (2014). A brief history of the syntax-semantics interface in western formal linguistics. *Semantics-Syntax Interface, 1*(1), 1-21

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41*(3), 427-456. doi: DOI 10.1006/jmla.1999.2653

Phillips, C. (2006). The real-time status of island phenomena. *Language*, 795-823

Phillips, C. (2013). On the nature of island constraints I: language processing and reductionist accounts. *Experimental syntax and island effects*, 64-108

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological bulletin, 134*(3), 427

Piñango, M. M., Finn, E., Lacadie, C., & Constable, R. T. (2016). The localization of long-distance dependency components: Integrating the focal-lesion and neuroimaging record. *Frontiers in psychology, 7*, 1434

Postal, P. (1969). On so-called "pronouns" in English. Modern studies in English, ed. by DA Reibel and S. Schane. 19-37: Englewood Cliffs, NJ: Prentice-Hall.

Potter, D. (2017). *The island (in) sensitivity of stripping.* Northwestern University.

Potter, D., & Görgülü, E. (2019). Fragments and Island (In) sensitivity in Turkish

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive science, 35*(4), 587-637

Roberts, L., Matsuo, A., & Duffield, N. (2013). Processing VP-ellipsis and VP-anaphora with structurally parallel and nonparallel antecedents: An eye-tracking study. *Language and Cognitive Processes, 28*(1-2), 29-47

Rohde, H., Kehler, A., & Elman, J. L. (2007). Pronoun interpretation as a side effect of discourse coherence *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 29).

Romero, M. (1998). Focus and reconstruction effects in wh-phrases.

Rooryck, J. (1992). Negative and factive islands revisited. *Journal of Linguistics, 28*(2), 343-374

Ross, J. R. (1967). Constraints on variables in syntax

Ross, J. R. (1969a). 2012. Guess who *Proceedings of the Fifth Regional Meeting of the Chicago Linguistic Society* (pp. 252-286).

Ross, J. R. (1969b). Guess Who. In et al. Binnick, RI, editor *Fifth Regional Meeting of the Chicago Linguistics Club*.

Sag, I. A. (1976). *Deletion and logical form.* Massachusetts Institute of Technology.

Sag, I. A., & Hankamer, J. (1984). Toward a theory of anaphoric processing. *Linguistics and philosophy*, 325-345

Scheepers, C., & Crocker, M. W. (2004). Constituent order priming from reading to listening: A visual-world study. *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond*, 167-185

Schneider, D., & Phillips, C. (2001). Grammatical search and reanalysis. *Journal of Memory and Language, 45*(2), 308-336

Shapiro, L. P., & Hestvik, A. (1995). On-line comprehension of VP-ellipsis: Syntactic reconstruction and semantic influence. *Journal of Psycholinguistic Research, 24*(6), 517-532

Shapiro, L. P., Hestvik, A., Lesan, L., & Garcia, A. R. (2003). Charting the time-course of VP-ellipsis sentence comprehension: Evidence for an initial and independent structural analysis. *Journal of Memory and Language, 49*(1), 1-19

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of verbal learning and verbal behavior, 13*(3), 272-281

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language, 69*(2), 104-120

Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research, 23*(3), 197-229

Sprouse, J., & Hornstein, N. (2013). *Experimental syntax and island effects*: Cambridge University Press.

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language*, 82-123

Staub, A., & Clifton Jr, C. (2006). Syntactic prediction in language comprehension: evidence from either... or. *Journal of experimental psychology: Learning, memory, and cognition, 32*(2), 425

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and cognitive processes, 1*(3), 227-245

Streb, J., Hennighausen, E., & Rösler, F. (2004). Different anaphoric expressions are investigated by event-related brain potentials. *Journal of Psycholinguistic Research, 33*(3), 175-201

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48*(3), 542-562

Szabolcsi, A., & Den Dikken, M. (1999). Islands. *Glot International, 4*(6), 3-8

Szabolcsi, A., & Lohndal, T. (2017). Strong vs. weak islands. *The Wiley Blackwell Companion to Syntax, Second Edition*, 1-51

Szabolcsi, A., & Zwarts, F. (1993). Weak islands and an algebraic semantics for scope taking. *Natural language semantics, 1*(3), 235-284

Takahashi, S., & Fox, D. (2005). MaxElide and the re-binding problem *Proceedings of SALT* (Vol. 15, pp. 223-240).

Tanenhaus, M. K., & Carlson, G. N. (1990). Comprehension of deep and surface verbphrase anaphors. *Language and Cognitive Processes, 5*(4), 257-280

Thothathiri, M., & Snedeker, J. (2008). Give and take: Syntactic priming during spoken language comprehension. *Cognition, 108*(1), 51-68

Tooley, K. M., & Traxler, M. J. (2010). Syntactic priming effects in comprehension: A critical review. *Language and Linguistics Compass, 4*(10), 925-937

Traxler, M. J. (2008). Lexically independent priming in online sentence comprehension. *Psychonomic bulletin & review, 15*(1), 149-155

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35*(3), 454-475

Traxler, M. J., Pickering, M. J., & Clifton Jr, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*(4), 558-592

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*(3), 285-316

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language, 55*(2), 157-166

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language, 65*(3), 247-263. doi: 10.1016/j.jml.2011.05.002

Van Gompel, R. P., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(1), 128

Van Gompel, R. P., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language, 55*(3), 335-362

Van Gompel, R. P., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 52*(2), 284-307

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language, 45*(2), 225-258

von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of memory and language, 94*, 119-133

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*(2), 206-237

Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology, 67*(7), 1274-1304

Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. Linguistic theory and psychological reality, ed. by Morris Halle, Joan Bresnan and George Miller, 119–61: Cambridge, MA: MIT Press.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85*(1), 79-112

Williams, E. (1978). Across-the-board rule application. *Linguistic Inquiry, 9*(1), 31-43

Williams, E. S. (1977). Discourse and logical form. *Linguistic inquiry*, 101-139

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language, 108*(1), 40-55

Xiang, M., Grove, J., & Merchant, J. (2019). Structural priming in production through 'silence': An investigation of verb phrase ellipsis and null complement anaphora. *Glossa: a journal of general linguistics, 4*(1)

Yoshida, M. (2018). Parsing Strategies *The Oxford Handbook of Ellipsis*.

Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes, 28*(3), 272-302

Yoshida, M., Kazanina, N., Pablos, L., & Sturt, P. (2014). On the origin of islands. *Language, Cognition and Neuroscience, 29*(7), 761-770

Yoshida, M., Lee, J., & Dickey, M. W. (2013). 16 The island (in) sensitivity of sluicing and sprouting. *Experimental Syntax and Island Effects*, 360

Yoshida, M., Potter, D., & Hunter, T. (2019). Condition C reconstruction, clausal ellipsis and island repair. *Natural Language & Linguistic Theory, 37*(4), 1515-1544

Appendix

Stimuli: Experiment 1a

(1) (Bill said Jane loves her/his mother and) Susan said David hates __ sister. (her/his)

a. Baseline:	Susan said David hates __ sister. (her/his)
b. Matrix:	Bill said Jane loves his mother and Susan said David hates __ sister. (her/his)
c. Local	Bill said Jane loves her mother and Susan said David hates __ sister. (her/his)

(2) (Amy said Sean likes his/her house and) Eric said Ann detests ____ car. (her/his)
(3) (Bob said Molly adores her/his brother and) Sue said Ken dislikes ____ father. (her/his)
(4) (Jen said Ken trusts his/her mechanic and) Larry said Erica appreciates ___ gardener. (her/his)
(5) (Joe thinks Kara admires her/his teacher and) Maggie thinks John respects ___ boss. (her/his)
(6) (Marsha thinks Bill pities his/her uncle and) Ethan thinks Sophia resents ___ aunt. (her/his)
(7) (Michael thinks Emma misses her/his dog and) Sarah thinks Max despises ____ cat. (her/his)
(8) (Beth thinks David fears his/her manager and) Lucy thinks Evan tolerates ___ job. (her/his)
(9) (Rich believes Susan cherishes her/his house and) Ella believes Joseph prizes __ car. (her/his)
(10) (Nancy believes Thomas supports his/her family and) Daniel believes Karen values __ friends. (her/his)
(11) (Paul believes Betty abhorred her/his haircut and) Sandra believes Mark loathes __ outfit. (her/his)
(12) (Ruth believes Kevin dreads his/her vacation and) Jason believes Sharon rues ___ job. (her/his)
(13) (Tony believes Ellen aggravates her/his son and) Maria believes Zach dazzles ___ daughter. (her/his)
(14) (Allison believes Colin frightens his/her pets and) Andrew believes Alice repels ___ friends. (her/his)
(15) (Jake believes Cora delights her/his students and) Kelly believes Noah enthralls ___ audience. (his/her)
(16) (Chloe believes Jack offends his/her parents and) Charlie believes Anna enchants ___ in-laws. (his/her)
(17) (Robert said Linda embarrasses her/his partner and) Barbara said William irritates ___ co-worker. (his/her)
(18) (Michelle said Jeff pleases his/her mentor and) Pablo said June dismays ___ nephew. (his/her)
(19) (Marty said Sheila saddens her/his grandchild and) Tara said Miguel entertains ___ father. (his/her)
(20) (Maura said Ira fascinates his/her pupils and) George said Lily mesmerizes ___ clients. (his/her)
(21) (Ryan thinks Sally scares her/his puppies and) Kayla thinks Jim startles ___ kittens. (his/her)
(22) (Claire thinks Jeremey teases his/her cousins and) Joey thinks Ruby amuses ___ nieces. (his/her)

(23) (Keith thinks Jessica enrages her/his sponsor and) Naomi thinks Harry shames ___ psychiatrist. (his/her)
(24) (Gwen thinks Mac alienates his/her colleagues and) Edward thinks Amanda impresses ___ associates. (his/her)
(25) (Earl thinks Agnes loves her/his cooking and) Jenna thinks Ely hates___ baking. (his/her)
(26) (Juliet thinks Nick likes his/her garden and) Josh thinks Delia adores ___ patio. (his/her)
(27) (Oliver thinks Katie frightens her/his terrier and) Audrey thinks Brett repels ___ bulldog. (his/her)
(28) (Cate thinks Dale impresses his/her customers and) Louie thinks Cathy delights ___ fans. (his/her)
(29) (Harry said Gloria misses her/his snake and) Leila said Hugh tolerates ___ lizard. (his/her)
(30) (Peggy said Roger supports his/her hobbies and) Seth said Kirsten loathes ___ habits. (his/her)
(31) (Shane said Tina offended her/his neighbor and) Tanya said Brent embarrassed ___ teammate. (his/her)
(32) (Megan said Eddie enchanted his/her toddler and) Johnny said April irritated ___ teenager. (his/her)
(33) (Jonas believes Dora abhors her/his position and) Lila believes Joel despises ___ responsibility. (her/his)
(34) (Piper believes Nate prizes his/her boat and) Toby believes Zoe enjoys ___ scooter. (her/his)
(35) (Thomas believes Christie dismays her/his priest and) Todd believes Elsa shocks ___ pastor. (her/his)
(36) (Abbie believes Wilfred calms his/her dogs and) Wally believes Denise excites ___ ferrets. (her/his)

Stimuli: Experiment 1b

(1) Michael/Emma thought Emma/Michael missed her dog and Sarah/Max thought Max/Sarah despised his cat with a great hatred.

a. Local/Parallel:	Michael thought Emma missed her dog and Sarah thought Max despised his cat with a great hatred.
b. Matrix/Parallel:	Emma thought Michael missed her dog and Max thought Sarah despised his cat with a great hatred.
c. Matrix/Non-parallel:	Michael thought Emma missed her dog and Max thought Sarah despised his cat with a great hatred.
d. Local/Non-parallel:	Emma thought Michael missed her dog and Sarah thought Max despised his cat with a great hatred.

Question: Did someone miss a dog?

(2) Beth/David thought David/Beth feared his manager and Evan/Lucy thought Lucy/Evan accepted her job with a resigned attitude.
Question: Did someone protest something?

(3) Rich/Susan thought Susan/Rich cherished her house and Ella/Joseph thought Joseph/Ella prized his car with a great pride.
Question: Did someone prize a car?

(4) Nancy/Thomas believed Thomas/Nancy supported his family and Daniel/Karen believed Karen/Daniel encouraged her friends with a nice compliment.
Question: Did someone support a co-worker?

(5) Ruth/Kevin believed Kevin/Ruth dreaded his vacation and Jason/Sharon believed Sharon/Jason rued her job with a bitter regret.
Question: Did someone dread a vacation?

(6) Tony/Ellen said Ellen/Tony aggravated her son and Maria/Zach said Zach/Maria dazzled his daughter with a lavish party.
Question: Did someone aggravate a friend?

(7 Allison/Colin believed Colin/Allison frightened his pets and Andrew/Alice believed Alice/Andrew insulted her friends with a mean comment.
Question: Did someone insult some friends?

(8) Jake/Cora believed Cora/Jake delighted her students and Kelly/Noah believed Noah/Kelly enthralled his audience with an eloquent speech.
Question: Did someone dismay some students?

(9) Chloe/Jack said Jack/Chloe offended his parents and Charlie/Anna said Anna/Charlie enchanted her in-laws with a natural charm.
Question: Was something done with natural charm?

(10) Robert/Linda said Linda/Robert embarrassed her partner and Barbara/William said William/Barbara irritated his co-worker with an annoying joke.
Question: Was a clever story told?

(11) Michelle/Jeff said Jeff/Michelle pleased his mentor and Pablo/June said June/Pablo dismayed her nephew with an outrageous proposal.
Question: Did an outrageous proposal dismay someone?

(12) Marty/Sheila said Sheila/Marty saddened her grandchild and Tara/Miguel said Miguel/Tara disappointed his father with a bad report-card.
Question: Did a detention slip disappoint someone?

(13) Maura/Ira said Ira/Maura fascinated his pupils and George/Lily said Lily/George mesmerized her clients with an engaging presentation.
Question: Were Ira's pupils fascinated?

(14) Ryan/Sally thought Sally/Ryan scared her puppies and Kayla/Jim thought Jim/Kayla startled his kittens with a loud alarm.
Question: Were Ryan's puppies scared?

(15) Claire/Jeremey thought Jeremey/Claire bored his cousins and Joey/Ruby thought Ruby/Joey amused her nieces with a creative story.
Question: Were Ruby's nieces amused?

(16) Earl/Agnes thought Agnes/Earl preferred her cooking and Jenna/Ely thought Ely/Jenna relished his baking with a refined palette.
Question: Was Jenna's baking relished?

(17) Juliet/Nick thought Nick/Juliet liked his garden and Josh/Delia thought Delia/Josh adored her patio with a great zest.
Question: Was Nick's garden liked?

(18) Harry/Gloria said Gloria/Harry missed her snake and Leila/Hugh said Hugh/Leila tolerated his lizard with a patient attitude.
Question: Was Hugh's snake missed?

(19) Peggy/Roger said Roger/Peggy supported his hobbies and Seth/Kirsten said Kirsten/Seth loathed her pastimes with a true disgust.
Question: Were Kirsten's pastimes loathed?

(20) Jonas/Dora believed Dora/Jonas abhorred her position and Lila/Joel believed Joel/Lila despised his responsibility with a true passion.
Question: Was Lila's responsibility despised?

(21) Piper/Nate believed Nate/Piper prized his boat and Toby/Zoe believed Zoe/Toby enjoyed her scooter with a real excitement.
Question: Was the word 'enjoyed' mentioned?

(22) Thomas/Christie believed Christie/Thomas dismayed her priest and Elsa/Todd believed Todd/Elsa shocked his pastor with a sinful action.
Question: Was the word 'displeased' mentioned?

(23) Abbie/Wilfred believed Wilfred/Abbie calmed his dogs and Wally/Denise believed Denise/Wally excited her ferrets with a tasty treat.
Question: Was the word 'tasty' mentioned?

(24) Carl/Amy thought Amy/Carl reviewed her paper and Emily/Matt thought Matt/Emily edited his essay with a thoughtful tutor.
Question: Was the word 'thankful' mentioned?

Stimuli: Experiment 2a

(1) (Michael thought Emma missed his/her dog whereas) Sarah thought Max despised ___ cat with a great hatred. (his/her)

a. Baseline:	Sarah thought Max despised ___ cat with a great hatred. (his/her)
b. Matrix:	Michael thought Emma missed his dog whereas Sarah thought Max despised ___ cat with a great hatred. (his/her)
c. Local	Michael thought Emma missed her dog whereas Sarah thought Max despised ___ cat with a great hatred. (his/her)

(2) (Beth thought David feared her/his manager whereas) Evan thought Lucy accepted ___ job with a resigned attitude. (his/her)

(3) (Rich thought Susan cherished his/her house whereas) Ella thought Joseph prized ___ car with a great pride. (his/her)

(4) (Nancy believed Thomas supported her/his family whereas) Daniel believed Karen encouraged ___ friends with a nice compliment. (his/her)

(5) (Ruth believed Kevin dreaded her/his vacation whereas) Jason believed Sharon rued ___ job with a bitter regret. (his/her)

(6) (Tony said Ellen aggravated his/her son whereas) Maria said Zach dazzled ___ daughter with a lavish party. (his/her)

(7) (Allison believed Colin frightened her/his pets whereas) Andrew believed Alice insulted ___ friends with a mean comment. (his/her)

(8) (Jake believed Cora delighted his/her students whereas) Kelly believed Noah enthralled ___ audience with an eloquent speech. (his/her)

(9) (Chloe said Jack offended her/his parents whereas) Charlie said Anna enchanted ___ in-laws with a natural charm. (his/her)

(10) (Robert said Linda embarrassed his/her partner whereas) Barbara said William irritated ___ co-worker with an annoying joke. (his/her)

(11) (Michelle said Jeff pleased her/his mentor whereas) Pablo said June dismayed ___ nephew with an outrageous proposal. (his/her)

(12) (Marty said Sheila saddened his/her grandchild whereas) Tara said Miguel disappointed ___ father with a bad report-card. (his/her)

(13) (Maura said Ira fascinated her/his pupils whereas) George said Lily mesmerized ___ clients with an engaging presentation. (her/his)

(14) (Ryan thought Sally scared his/her puppies whereas) Kayla thought Jim startled ___ kittens with a loud alarm. (her/his)

(15) (Claire thought Jeremey bored her/his cousins whereas) Joey thought Ruby amused ___ nieces with a creative story. (her/his)

(16) (Earl thought Agnes preferred his/her cooking whereas) Jenna thought Ely relished ___ baking with a refined palette. (her/his)

(17) (Juliet thought Nick liked her/his garden whereas) Josh thought Delia adored ___ patio with a great zest. (her/his)

(18) (Harry said Gloria missed his/her snake whereas) Leila said Hugh tolerated ___ lizard with a patient attitude. (her/his)

(19) (Peggy said Roger supported her/his hobbies whereas) Seth said Kirsten loathed ___ pastimes with a true disgust. (her/his)

(20) (Jonas believed Dora abhorred his/her position whereas) Lila believed Joel despised ___ responsibility with a true passion. (her/his)

(21) (Piper believed Nate prized her/his boat whereas) Toby believed Zoe enjoyed ___ scooter with a real excitement. (her/his)

(22) (Thomas believed Christie dismayed his/her priest whereas) Elsa believed Todd shocked ___ pastor with a sinful action. (her/his)

(23) (Abbie believed Wilfred calmed her/his dogs whereas) Wally believed Denise excited ___ ferrets with a tasty treat. (her/his)

(24) (Carl thought Amy reviewed his/her paper whereas) Emily thought Matt edited ___ essay with a thoughtful tutor. (her/his)

Stimuli: Experiment 2b

(1) Michael/Emma thought Emma/Michael missed her dog whereas Sarah/Max thought Max/Sarah despised his cat with a great hatred.

a. Local/Parallel:	Michael thought Emma missed her dog whereas Sarah thought Max despised his cat with a great hatred.
b. Matrix/Parallel:	b. Emma thought Michael missed her dog whereas Max thought Sarah despised his cat with a great hatred.
c. Matrix/Non-parallel:	Michael thought Emma missed her dog whereas Max thought Sarah despised his cat with a great hatred.
d. Local/Non-parallel:	Emma thought Michael missed her dog whereas Sarah thought Max despised his cat with a great hatred.

Question: Did someone miss a dog?

(2) Beth/David thought David/Beth feared his manager whereas Evan/Lucy thought Lucy/Evan accepted her job with a resigned attitude.
Question: Did someone protest something?

(3) Rich/Susan thought Susan/Rich cherished her house whereas Ella/Joseph thought Joseph/Ella prized his car with a great pride.
Question: Did someone prize a car?

(4) Nancy/Thomas believed Thomas/Nancy supported his family whereas Daniel/Karen believed Karen/Daniel encouraged her friends with a nice compliment.
Question: Did someone support a co-worker?

(5) Ruth/Kevin believed Kevin/Ruth dreaded his vacation whereas Jason/Sharon believed Sharon/Jason rued her job with a bitter regret.
Question: Did someone dread a vacation?

(6) Tony/Ellen said Ellen/Tony aggravated her son whereas Maria/Zach said Zach/Maria dazzled his daughter with a lavish party.
Question: Did someone aggravate a friend?

(7 Allison/Colin believed Colin/Allison frightened his pets whereas Andrew/Alice believed Alice/Andrew insulted her friends with a mean comment.
Question: Did someone insult some friends?

(8) Jake/Cora believed Cora/Jake delighted her students whereas Kelly/Noah believed Noah/Kelly enthralled his audience with an eloquent speech.
Question: Did someone dismay some students?

(9) Chloe/Jack said Jack/Chloe offended his parents whereas Charlie/Anna said Anna/Charlie enchanted her in-laws with a natural charm.
Question: Was something done with natural charm?

(10) Robert/Linda said Linda/Robert embarrassed her partner whereas Barbara/William said William/Barbara irritated his co-worker with an annoying joke.
Question: Was a clever story told?

(11) Michelle/Jeff said Jeff/Michelle pleased his mentor whereas Pablo/June said June/Pablo dismayed her nephew with an outrageous proposal.
Question: Did an outrageous proposal dismay someone?

(12) Marty/Sheila said Sheila/Marty saddened her grandchild whereas Tara/Miguel said Miguel/Tara disappointed his father with a bad report-card.
Question: Did a detention slip disappoint someone?

(13) Maura/Ira said Ira/Maura fascinated his pupils whereas George/Lily said Lily/George mesmerized her clients with an engaging presentation.
Question: Were Ira's pupils fascinated?

(14) Ryan/Sally thought Sally/Ryan scared her puppies whereas Kayla/Jim thought Jim/Kayla startled his kittens with a loud alarm.
Question: Were Ryan's puppies scared?

(15) Claire/Jeremey thought Jeremey/Claire bored his cousins whereas Joey/Ruby thought Ruby/Joey amused her nieces with a creative story.
Question: Were Ruby's nieces amused?

(16) Earl/Agnes thought Agnes/Earl preferred her cooking whereas Jenna/Ely thought Ely/Jenna relished his baking with a refined palette.
Question: Was Jenna's baking relished?

(17) Juliet/Nick thought Nick/Juliet liked his garden whereas Josh/Delia thought Delia/Josh adored her patio with a great zest.
Question: Was Nick's garden liked?

(18) Harry/Gloria said Gloria/Harry missed her snake whereas Leila/Hugh said Hugh/Leila tolerated his lizard with a patient attitude.
Question: Was Hugh's snake missed?

(19) Peggy/Roger said Roger/Peggy supported his hobbies whereas Seth/Kirsten said Kirsten/Seth loathed her pastimes with a true disgust.
Question: Were Kirsten's pastimes loathed?

(20) Jonas/Dora believed Dora/Jonas abhorred her position whereas Lila/Joel believed Joel/Lila despised his responsibility with a true passion.
Question: Was Lila's responsibility despised?

(21) Piper/Nate believed Nate/Piper prized his boat whereas Toby/Zoe believed Zoe/Toby enjoyed her scooter with a real excitement.
Question: Was the word 'enjoyed' mentioned?

(22) Thomas/Christie believed Christie/Thomas dismayed her priest whereas Elsa/Todd believed Todd/Elsa shocked his pastor with a sinful action.
Question: Was the word 'displeased' mentioned?

(23) Abbie/Wilfred believed Wilfred/Abbie calmed his dogs whereas Wally/Denise believed Denise/Wally excited her ferrets with a tasty treat.
Question: Was the word 'tasty' mentioned?

(24) Carl/Amy thought Amy/Carl reviewed her paper whereas Emily/Matt thought Matt/Emily edited his essay with a thoughtful tutor.
Question: Was the word 'thankful' mentioned?

Stimuli: Experiment 3a/3b.
Comprehension questions were only included for 3b.

(1) (Bill thinks that) Mary, (for some reason)/(after the meeting), quit her job, but I don't know why specifically, although I hope to find out soon.

a. Long/Sluice:	Bill thinks that Mary, for some reason, quit her job, but I don't know why specifically, although I hope to find out soon.
b. Short/Sluice:	Mary, for some reason, quit her job, but I don't know why specifically, although I hope to find out soon.
c. Long/Sprout:	Bill thinks that Mary, after the meeting, quit her job, but I don't know why specifically, although I hope to find out soon.
d. Short/Sprout:	Mary, after the meeting, quit her job, but I don't n\know why specifically, although I hope to find out soon.

Question: Did (Bill think that) Mary quit her job?

(2) (Jill believed that) John, (for some reason)/(under great stress), left his wife, but I don't know why exactly, since I've been out of town.
Question: Did (Jill believe that) John got/get married?

(3) (Jane said that) Bob, (for some reason)/(without much thought), bought a house, but I don't know why precisely, however I'll find out later.
Question: Did (Jane say that) Bob bought/buy a house?

(4) (Mike confirmed that) Lucy, (for some reason)/(during last week), adopted a dog, but I don't know why specifically, nevertheless I'd really like to know.
Question: Did (Mike confirm that) Lucy adopt(ed) a cat?

(5) (Sean thought that) Leila, (for some reason)/(in Central Park), ran a marathon, but I don't know why exactly, because I'm out of the loop.
Question: Did (Sean think that) Leila ran/run a marathon?

(6) (Kelly believed that) Aaron, (for some reason)/(around nine yesterday), went to Cuba, but I don't know why precisely, considering I lost my phone.
Question: Did (Kelly believe that) Aaron go/went to France?

(7) (Joey said that) Maria, (for some reason)/(despite some concerns), sold her car, but I don't know why specifically, though I'm not surprised.
Question: Did (Joey say that) Maria sold/sell her bike?

(8) (Ann confirmed that) Jimmy, (for some reason)/(in the evening), went to the doctor, but I don't know why exactly, despite trying to find out.
Question: Did (Ann confirm that) Jimmy went/go to the doctor?

(9) (Susan thinks that) Bobby, (for some reason)/(in the afternoon), talked to a priest, but I don't know why precisely, although I hope to find out soon.
Question: Did Susan say that Bobby talked to a therapist?

(10) (Martha swears that) Emma, (for some reason)/(on halloween night), wore a bear costume, but I don't know why specifically, since I've been out of town.
Question: Did (Marth swear that) Emma wore/wear a cat costume?

(11) (Emily admitted that) Josh, (for some reason)/(at the end of the year), failed his classes, but I don't know why exactly, however I'll find out later.
Question: Did (Emily admit that) Josh fail(ed) his classes?

(12) (Melissa regretted that) Peter, (for some reason)/(at the last minute), abandoned his children, but I don't know why precisely, nevertheless I'd really like to know.
Question: Did (Melissa regret that) Peter adopt(ed) his children?

(13) (Jeremy presumed that) Jenny, (for some reason)/(in the winter), moved to Aspen, but I don't know why particularly, because I'm out of the loop.
Question: Did (Jeremy presume that) Jenny move(d) to Aspen?

(14) (Karen stated that) Kim, (for some reason)/(at the mall), pierced her nose, but I don't know why particularly, considering I lost my phone.
Question: Did (Karen state that) Kim pierc(ed) her nose?

(15) (Matthew claimed that) Laura, (for some reason)/(in the evening), got a tattoo, but I don't know why particularly, though I'm not surprised.
Question: Did (Matthew claim that) Laura go/get a piercing?

(16) (Jessica mentioned that) David, (for some reason)/(in the town square), punched a man, but I don't know why particularly, despite trying to find out.
Question: Did (Jessica mention that) David punch(ed) a man?

(17) (Bill remembered that) Kyle, (for some reason)/(at the dance), dumped his girlfriend, but I don't know why specifically, although I hope to find out soon.
Question: Did Kyle kiss his girlfriend?

(18) (Ellen suggested that) Jerry, (for some reason)/(back in January), lost his apartment, but I don't know why exactly, since I've been out of town.
Question: Did Jerry lose his dog?

(19) (Janice insisted that) Jack, (for some reason)/(at the university), argued with his instructor, but I don't know why precisely, however I'll find out later.
Question: Did (Janice insist that) Jack agree(d) with his instructor?

(20) (Erica explained that) Molly, (for some reason)/(in her youth), committed a crime, but I don't know why particularly, nevertheless I'd really like to know.
Question: Did (Erica explain that) Molly commit(ted) a crime?

(21) (Ben confessed that) Eric, (for some reason)/(out of desperation), ruined his reputation, but I don't know why specifically, because I'm out of the loop.
Question: Did (Ben confess that) Eric ruin(ed) his reputation?

(22) (Jan felt that) Kerry, (for some reason)/(after the conference), hated her boss, but I don't know why exactly, considering I lost my phone.
Question: Did (Jan feel that) Kerry hate(d) her husband?

(23) (Joyce argued that) William, (for some reason)/(against his judgment), disliked his profession, but I don't know why precisely, though I'm not surprised.
Question: Did (Joyce argue that) William dislike(d) his profession?

(24) Bryan complained that Sheila, (for some reason)/(at the party), ignored all her friends, but I don't know why particularly, despite trying to find out.
Question: Did (Bryan complain that) Sheila ignore(d) all her friends?

Stimuli: Experiment 4a/4b.

Comprehension questions were only included in 4b.

(1) (Bill hated that) Mary, (for some reason)/(in the morning), bought (clothes)/(something), and somebody should know why specifically, although they're keeping it a secret.

a. Long/Adjunct: Bill hated that Mary, for some reason, bought clothes, and somebody should know why specifically, although they're keeping it a secret.
b. Short/Adjunct: Mary, for some reason, bought clothes, and somebody should know why specifically, although they're keeping it a secret.
c. Long/Argument: Bill hated that Mary, in the morning, bought something, and somebody should know what specifically, although they're keeping it a secret.
d. Short/Argument: Mary, in the morning, bought something, and somebody should know what specifically, although they're keeping it a secret.
Question: Did (Bill hate that) Mary bought/buy clothes/something?

(2) (Jill detested that) John, (for some reason)/ (after the meeting), wrote (letters)/(something), and someone will know (why)/(what) exactly, since it wasn't a secret.
Question: Did Jill believe that John got married?/Did John eat something?

(3) (Jane loved that) Bob, (for some reason)/(under great stress), broke (mirrors)/(something), and everyone might remember (why)/(what) precisely, since it seemed strange.
Question: Did (Jane love that) Bob broke/break something?

(4) (Mike resented that) Lucy, (for some reason)/(without much thought), discarded (cans)/(something), and somebody should remember (why)/(what) specifically, but they aren't telling anyone.
Question: Did Mike love that Lucy discarded something?/Did Lucy love something?

(5) (Sean liked that) Leila, (for some reason)/(just last week), learned (Spanish)/(something), and someone will ask (why)/(what) exactly, because everyone wants to know.
Question: Was the word 'someone' mentioned?

(6) (Kelly worried that) Aaron, (for some reason)/(last Friday night) wanted (advice) /(something), and everyone might ask (why)/(what) precisely, considering it seems unusual.
Question: Was a computer mentioned?

(7) (Joey admired that) Maria, (for some reason)/(in Central Park), sold (flowers) /(something), and somebody figured out (why)/(what) specifically, though no-one is telling.
Question: Were flowers mentioned?/Was a park mentioned?

(8) (Jimmy doubted that) Ann, (for some reason)/(despite some concerns), donated (blood) /(something), and someone figured out (why)/(what) exactly, despite the attempts to hide it.
Question: Was a nurse mentioned?

(9) (Susan denied that) Bobby, (for some reason)/(in the morning), baked (cookies)/(something), and everyone found out (why)/(what) precisely, although it didn't make much sense.
Question: Did everyone find out why/what precisely?

(10) (Martha loathed that) Emma, (for some reason)/(as a favor), offered (money) /(something), and somebody found out (why)/(what) specifically, since people love to talk.
Question: Did somebody find out when specifically?

(11) (Emily adored that) Josh, (for some reason)/(during last semester), failed (tests) /(something), and someone will recall (why)/(what) exactly, however it's probably not important.

Question: Will someone recall why/what exactly?

(12) (Peter distrusted that) Melissa, (for some reason)/(after the crime), destroyed (evidence) /(something), and everyone might recall (why)/(what) precisely, since it seemed like a big deal.
Question: Might everyone forget why/what exactly?

(13) (Clara condemned that) George, (for some reason)/(in the morning), attended (class)/(something), but I don't know (why)/(what) specifically, although I hope to find out soon.
Question: Did the word ‘specifically’ appear in this sentence?

(14) (Tom disliked that) Sally, (for some reason)/(after the meeting), won (prizes) /(something), but no-one will know (why)/(what) exactly, since they've been out of town.
Question: Did the word ‘city’ appear in this sentence?

(15) (Kim respected that) Josh, (for some reason)/(under great stress), fostered (dogs) /(something), but nobody could remember (why)/(what) precisely, however it doesn't really matter.
Question: Did the word ‘fostered’ appear in this sentence?

(16) (Laura abhorred that) Dan, (for some reason)/(without much thought), rented (cars) /(something), but I don't remember (why)/(what) specifically, nevertheless I'd really like to know.
Question: Did the word ‘motorcycle’ appear in this sentence?

(17) (Bart despised that) Melissa, (for some reason)/(just last week), created (art) /(something), but I didn't ask (why)/(what) exactly, because it's nobody's business.
Question: Did somebody create something?

(18) (Carrie appreciated that) Bill, (for some reason)/(last Friday night), ruined (parties) /(something), but nobody could ask (why)/(what) precisely, considering they don't talk to anybody.
Question: Did somebody lose a book?

(19) (Stuart treasured that) Maura, (for some reason)/(in Central Park), killed (plants)/(something), but I didn't figure out (why)/(what) specifically, although I'm not surprised.
Question:Did somebody kill something?

(20) (Brent cherished that) Judy, (for some reason)/(in the morning), lost (pets)/(something), but no-one figured out (why)/(what) exactly, despite trying to find out.
Question: Did no-one figure out where?

(21) (Travis challenged that) Kelly, (for some reason)/(as a hobby), published (books)/(something), but nobody found out (why)/(what) precisely, although it will soon come to light.
Question: Did Travis challenge that somebody did something?/ Did Kelly publish something?

(22) (Allie disputed that) Ken, (for some reason)/(as a favor), told (stories)/(something), but no-one will find out (why)/(what) specifically, since gossiping is prohibited.
Question: Was a book mentioned?

(23) (Kendra deplored that) Joseph, (for some reason)/(during last semester), made (pies)/(something), but no-one will recall (why)/(what) exactly, because it wasn't very interesting.
Question: Did Kendra deplore that Joseph did something?/Did Joseph make something?

(24) (Brad regretted that) Sue, (for some reason)/(after the crime), returned (money)/(something), but nobody could recall (why)/(what) precisely, seeing that it was a long time ago.
Question: Did nobody recall why/what exactly?

Stimuli: Experiment 5a

(1) Mary criticized (herself)/(Bill), but John did not, but John did criticize (himself)/(her)/(Bill) for many good reasons.

a. Strict: *Mary criticized herself, but John did not, but John did criticize himself for many good reasons.*
b. Sloppy: *Mary criticized herself, but John did not, but John did criticize her for many good reasons.*
c. Baseline: *Mary criticized Bill, but John did not, but John did criticize Bill for many good reasons.*

(2) Bill punished (himself)/(Kendra), and Susan did too, but Susan didn't punish (herself)/(him)/(Kendra) since it was unnecessary.

(3) Jen judged (herself)/(Sophia), but Bob did not, but Bob did judge (himself)/(her)/(Sophia) for several poor decisions.

(4) Sean hated (himself)/(Gavin), but Angela did not, but Angela did hate (herself)/(him)/(Gavin) due to numerous considerations.

(5) Linda loved (herself)/(Molly), and Tom did too, though Tom didn't love (himself)/(her)/(Molly) because it was difficult.

(6) Joe adored (himself)/(Ken), but Sheila did not, though Sheila did adore (herself)/(him)/(Ken) since it was justified.

(7) Emily despised (herself)/(Chris), and Dave did too, although Dave didn't despise (himself)/(her)/(Chris) considering it was hateful.

(8) Aaron detested (himself)/(Vicky), but Jane did not, although Jane did detest (herself)/(him)/(Vicky) due to shameful behavior.

(9) Kelly admired (herself)/(Isaac), and Joey did too, although Joey didn't admire (himself)/(her)/(Isaac) because it was uncalled-for.

(10) Jimmy respected (himself)/(Julia), but Ann did not, although Ann did respect (herself)/(him)/(Julia) since it was appropriate.

(11) Emma complimented (herself)/(Tim), and Josh did too, though Josh didn't compliment (himself)/(her)/(Tim) considering it was irrelevant.

(12) Peter blamed (himself)/(Joel), but Clara did not, though Clara did blame (herself)/(him)/(Joel) due to poor planning.

(13) Sally disliked (herself)/(Kayla), but George did not, however George did dislike (himself)/(her)/(Kayla) because it was warranted.

(14) Dan challenged (himself)/(Glen), and Judy did too, however Judy didn't challenge (herself)/(him)/(Glen) since it was fruitless.

(15) Andrea deplored (herself)/(Marilyn), but Brad did not, however Brad did deplore (himself)/(her)/(Marilyn) considering it was deserved.

(16) Mike loathed (himself)/(Anthony), and Beth did too, however Beth didn't loathe (herself)/(him)/(Anthony) due to its senselessness.

(17) Lucy insulted (herself)/(Lisa), but Evan did not, yet Evan did insult (himself)/(her)/(Lisa) because it was effective.

(18) Paul indulged (himself)/(Jill), and Maria did too, yet Maria didn't indulge (herself)/(him)/(Jill) since it was contemptible.

(19) Chloe entertained (herself)/(Doug), but Jack did not, yet Jack did entertain (himself)/(her)/(Doug) considering it was easy.

(20) Charlie praised (himself)/(Sabrina), and Maura did too, yet Maura didn't praise (herself)/(him)/(Sabrina) due to it's frivolousness.

(21) Ira spoiled (himself)/(Justin), but Sally did not, still Sally did spoil (herself)/(him)/(Justin) because it was gratifying.

(22) Jessica bored (herself)/(Alicia), and Ryan did too, still Ryan didn't bore (himself)/ since it was impossible.

(23) Jeremey amused (himself)/(James), and Gwen did too, still Gwen didn't amuse (herself)/(him)/(James) considering it was tiring.

(24) Katie liked (herself)/(Allison), and Rich did too, still Rich didn't like (himself)/(her)/(Allison) due to it's indulgence.

Stimuli: Experiment 5b

(1) Although (Mike)/(Mary) didn't punish himself/herself, Bill punished himself and (Mike)/(Mary) did too since it was necessary.
Question: Did Bill punish himself?

a. Strict/Match: Although Mike didn't punish himself, Bill punished himself and Mike did too since it was necessary.

b. Strict/Mismatch: Although Mary didn't punish herself, Bill punished himself and Mary did too since it was necessary.

c. Sloppy/Match: Although Mike didn't punish him, Bill punished himself and Mike did too since it was necessary.

d. Sloppy/Mismatch: Although Mary didn't punish him, Bill punished himself and Mary did too since it was necessary.

(2) Although (Carla)/(John) didn't criticize herself/himself, Susan criticized herself, and (Carla)/(John) did too for many good reasons.
Question: Did Susan praise herself?

(3) Although (Bob)/(Kayla) didn't judge himself/herself, Ken judged himself, and (Bob)/(Kayla) did too for several poor decisions.
Question: Did Ken judge himself?

(4) Although (Angela)/(Andy) didn't hate herself/himself, Gwen hated herself, and (Angela)/(Andy) considering numerous causes.
Question: Did Gwen treat herself?

(5) Although(Peter)/(Molly) didn't love himself/herself, Tom loved himself, and (Peter)/(Molly) did too because it was easy.
Question: Did Tom love himself?

(6) Although (Linda)/(Joe) didn't adore herself/himself, Sheila adored herself, and (Linda)/(Joe) did too since it was justified.
Question: Did Sheila detest herself?

(7) Whereas (Sean)/(Emily) didn't despise himself/herself, Jimmy despised himself, and (Sean)/(Emily) did too considering all the reasons.
Question: Did Jimmy despise himself?

(8) Whereas (Kim)/(Aaron) didn't detest herself/himself, Jane detested herself, and (Kim)/(Aaron) did too for shameful behavior.
Question: Did Jane appreciate herself?

(9) Whereas (Joey)/(Kelly) didn't admire himself/herself, Aidan admired himself, and (Joey)/(Kelly) did too because it was called for.
Question: Was admiring someone called for?

(10) Whereas (Annie)/(Keith) didn't respect herself/himself, Julia respected herself, and (Annie)/(Keith) did too because it was deserved.
Question: Was respecting someone not deserved?

(11) Whereas (Josh)/(Diane) didn't compliment himself/herself, Gavin complimented himself, and (Josh)/(Diane)did too considering it was relevant.
Question: Was complimenting someone relevant?

(12) Whereas (Clara)/(Charles) didn't blame herself/himself, Katy blamed herself, and (Clara)/(Charles) did too considering the big failure.
Question: Was blaming someone done without cause?

(13) Though (George)/(Teresa) didn't dislike himself/herself, Dylan disliked himself, and (George)/(Teresa)did too for various flaws.
Question: Was someone disliked for various flaws?

(14) Though (Sally)/(Glen) didn't challenge herself/himself, Judy challenged herself, and (Sally)/(Glen) did too because it was motivating.
Question: Was challenging someone boring?

(15) Though (Brad)/(Andrea) didn't deplore himself/herself, Dan deplored himself, and (Brad)/(Andrea) did too considering it was deserved.
Question: Did someone deserve to be deplored?

(16) Though (Mac)/(Beth) didn't loathe himself/herself, James loathed himself, and (Mac)/(Beth) did too for being self-indulgent.
Question: Was someone loathed for being successful?

(17) Though (Lucy)/(Tim) didn't insult herself/himself, Emma insulted herself, and (Lucy)/(Tim) did too because it was effective.
Question: Was admiring someone called for?

(18) Though (Evan)/(Maria) didn't indulge himself/herself, Greg indulged himself, and (Evan)/(Maria) did too since it was enjoyable.
Question: Did Maria/Evan compliment someone?

(19) While (Jill)/Craig) didn't entertain herself/himself, Maura entertained herself, and (Jill)/Craig) did too since there was time.
Question: Did Craig/Jill entertain someone?

(20) While (Charlie)/(Zoe) didn't praise himself/herself, Daniel praised himself, and (Charlie)/(Zoe) did too for the gratification.
Question: Did Zoe/Charlie disgust someone?

(21) While (Sally)/(Ira) didn't spoil herself/himself, Jessica spoiled herself, and (Sally)/(Ira) did too because it was satisfying.
Question: Did Ira/Sally spoil someone?

(22) While (Ryan)/(Jenny) didn't bore himself/herself, Jeremy bored himself, and (Ryan)/(Jenny) did too since it was inevitable.
Question: Did Jenny/Ryan amuse someone?

(23) While (Alicia)/(Eric) didn't amuse herself/himselff, Vicky amused herself, and (Alicia)/(Eric) did too since it was entertaining.
Question: Did Eric/Alicia amuse someone?

(24) While (Rich)/(Ruth) didn't like himself/herself, Brent liked himself, and (Rich)/(Ruth) did too considering the circumstances.
Question: Did Ruth/Rich dislike someone?

Stimuli: Experiment 6a/6b.

Comprehension questions only included in 6b.

(1) Since (he/his teachers)/(she/her teachers) didn't loudly, the students said Norman's name clearly, and Nathan heard it.
Question: Did someone hear Norman's name?

a. C-Command/Match: Since he didn't loudly, the students said Norman's name clearly, and Nathan heard it.
b.No C-Command/Match: Since his teachers didn't loudly, the students said Norman's name clearly, and Nancy heard it.
c. C-Command/Mismatch: Since she didn't loudly, the students said Norman's name clearly, and Nancy heard it.
d. No C-Command /Mismatch: Since her teachers didn't loudly, the students said Norman's name clearly, and Nancy heard it.

(2) Since (he/his students)/(she/her students) didn't spiritedly, the choir sang Liam's song passionately, but Elliot hated it.

Question: Did the choir sing Liam's song unwillingly?\

(3) Whereas (he/his professors)/(she/her professors) didn't willingly, the TAs called Michael's parents persistently, and Carlos supported it.
Question: Did the TAs call someone's parents?\

(4) Because (he/his housekeepers)/(she/her housekeepers) didn't meticulously, the maids made Jacob's bed neatly, but Steve resented it.
Question: Was something done meticulously?\

(5) Whereas (he/his trainers)/(she/her trainers) didn't patiently, the employees trained William's dog strictly, but Victor despised it.
Question: Was training performed?\

(6) Because (he/his advisors)/(she/her advisors) didn't fully, the editors reviewed Ethan's thesis harshly, and Glenn rejected it.
Question: Was something done fully?\

(7) Although (he/his aides)/(she/her aides) didn't permanently, the senators revised James' proposal temporarily, and Lucas liked it.
Question: Was something done temporarily?\

(8) Although (he/his partners)/(she/her partners) didn't furiously, the friends attacked Daniel's nemesis violently, but Arnold stopped it.
Question: Did the friends attack Emma's nemesis?\

(9) Though (he/his secretaries)/(she/her secretaries) didn't elegantly, the cleaners organized Jackson's papers helpfully, and Martin destroyed them.
Question: Were secretaries mentioned?\

(10) Though (he/his daughters)/(she/her daughters) didn't irritably, the guests greeted Benjamin's family angrily, but Cesar corrected them.
Question: Did Carol agree with the guests?\

(11) Whenever (he/his assistants)/(she/her assistants) didn't thoroughly, the auditors reviewed Oliver's account rigorously, and Henry checked it.
Question: Was Oliver's account reviewed?\

(12) Whenever (he/his sisters)/(she/her sisters) didn't promptly, the coworkers ate Gabriel's meal selfishly, but Martin remade it.
Question: Was something done promptly?

(13) As (he/his decorators)/(she/her decorators) didn't swiftly, the teenagers painted Samuel's wall hastily, and Benson found out.
Question: Did the teenagers paint Samuel's wall?

(14) As (he/his friends)/(she/her friends) didn't eagerly, the campers told John's story quietly, but Eric didn't care.
Question: Was John's movie mentioned?

(15) When (he/his sons)/(she/her sons) didn't accurately, the mechanics fixed Luke's car flawlessly, and Trevor noticed it.
Question: Did the mechanics do something?

(16) When (he/his parents)/(she/her parents) didn't nicely, the students explained Henry's predicament politely, but Manuel ignored them.
Question: Did the students explain Henry's theory?

(17) While (he/his children)/(she/her children) didn't carelessly, the groomers patted Andrew's cat lightly, and Nicholas watched it.
Question: Does Andrew have a cat?

(18) While (he/his busboys)/(she/her busboys) didn't quickly, the waiters cleared Isaac's table hesitantly, but Louis stopped them.
Question: Did the waiters clear Irene's table?

(19) Since (she/her nephews)/(he/his nephews) didn't instantly, the attendants helped Emma's grandparent rapidly, and Gemma disregarded them.
Question: Was someone disregarded?

(20) Since (she/her cousins)/(he/his cousins) didn't skillfully, the baristas brewed Olivia's coffee expertly, but Paulina spilled it.
Question: Did someone spill Olivia's coffee?

(21) Whereas (she/her family)/(he/his family) didn't attentively, the newscasters watched Sophia's video kindly, and Kamila filmed it.
Question: Did the newscasters listen to Sophia's recordings?

(22) Whereas (she/her writers)/(he/his writers) didn't humorously, the comedians told Isabella's joke charmingly, but Irene hated it.
Question: Did comedians tell Isabella's secret?

(23) Because (she/her nieces)/(he/his nieces) didn't purposefully, the ladies exposed Charlotte's secret privately and Anika heard it.
Question: Was a secret exposed?

(24) Because (she/her proteges)/(he/his proteges) didn't neatly, the chefs cleaned Amelia's station meticulously, but Brenda was displeased.
Question: Were busboys mentioned?

(25) Although (she/her editors)/(he/his editors) didn't completely, the publishers edited Elizabeth's novel thoroughly, and Ashley appreciated it.
Question: Was something done thoroughly?

(26) Although (she/her colleagues)/(he/his colleagues) didn't personally, the professors presented Sofia's research remotely, but Claudia was angry.
Question: Was something done personally ?

(27) Though (she/her friends)/(he/his friends) didn't badly, the bellboys carried Evelyn's bags easily, and Cassie helped them.
Question: Did someone help the bellboys?

(28) Though (she/her maids)/(he/his maids) didn't habitually, the servants brushed Chloe's hair occasionally, but Zara knotted it.
Question: Did servants brush Chloe's hair habitually?

(29) Whenever (she/her brothers)/(he/his brothers) didn't carefully, the teammates drove Victoria's car safely, and Jenna parked it.
Question: Was a car parked?

(30) Whenever (she/her associates)/(he/his associates) didn't excitedly, the marketers promoted Lillian's business diligently, but Tiffani ruined it.
Question: Did the marketers ruin Lillian's business?

(31) As (she/her sponsors)/(he/his sponsors) didn't critically, the attorneys assessed Natalie's statement seriously, and Lynda documented it.
Question: Was the word 'critically' mentioned?

(32) As (she/her teammates)/(he/his teammates) didn't playfully, the players finished Hannah's game goofily, but Barbara regretted it.
Question: Was it Henry's game?

(33) When (she/her bosses)/(he/his bosses) didn't enthusiastically, the colleagues announced Alexa's news proudly, and Melissa was surprised.
Question: Was anyone surprised?

(34) When (she/her lawyers)/(he/his lawyers) didn't covertly, the officials hid Zoe's evidence secretly, but Viola discovered it.
Question: Was something done secretly?

(35) While (she/her uncles)/(he/his uncles) didn't calmly, the parents evaluated Penelope's actions peacefully, and Julianna justified them.
Question: Were Penelope's cousins involved?

(36) While (she/her coaches)/(he/his coaches) didn't confidently, the actors performed Rachel's play awkwardly, but Emma enjoyed it.
Question: Was something done confidently?