UNIVERSITY OF THE BASQUE COUNTRY (UPV/EHU)

DOCTORAL DISSERTATION

UNDERSTANDING DEPENDENCIES IN REAL TIME: A CROSSLINGUISTIC INVESTIGATION OF ANTECEDENT COMPLEXITY AND DEPENDENCY LENGTH

Author:
Sergio LÓPEZ SANCIO

Supervisor:
Prof. Itziar LAKA

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics

at the

Department of Linguistics and Basque Studies

2020

Abstract

Words in a sentence are dependent on each other for syntactic and semantic interpretation. The existence of dependencies between non-adjacent words implies our memory system needs to integrate information across intervening linguistic material by either actively maintaining information in memory and/or retrieving previously encoded representations. In this dissertation, I investigate the memory mechanisms involved in the processing of dependencies between non-adjacent words, exploring both local and unbounded dependencies in typologically diverse languages —English, Spanish and Basque. Specifically, I investigate two factors that affect the activation level of memory targets and, hence, their subsequent retrieval from memory in sentence comprehension: (a) the representational complexity of antecedents (Chapter 2), and (b) activation decay as a function of time (Chapters 3 & 4). I show that increasing the syntactic complexity of antecedents increases their activation level and facilitates their subsequent retrieval from memory. I also show that activation decay is a major determinant of comprehension difficulty in both local and unbounded dependencies in Spanish, suggesting locality effects are a general phenomenon in VO languages. Lastly, I find evidence that expectation-driven facilitation overrides activation decay in Basque, supporting the idea that all OV languages show antilocality effects, regardless of the degree of word order flexibility. In sum, this dissertation broadens the available evidence to further understand the mechanisms underlying the comprehension of linguistic dependencies in real time.

Acknowledgements

Finishing a PhD dissertation is always a great achievement in someone's life.[1] Finishing a PhD dissertation during a global pandemic is the next level. I must admit this has been tougher than I would have ever imagined. In the past five years, I've had the chance to meet amazing people who have helped me in different ways. It is impossible to sum up how grateful I am to all of you, but I'll try my best.

I'd like to thank my supervisor, Itziar Laka. When I met Itziar, I was a master's student and her psycholinguistics/neurolinguistics classes just blew my mind. Sometimes I felt like I was watching a documentary. She always encouraged me to think outside the box and gave me absolute freedom to explore new research lines. She also taught me science should be clear and concise. While I was writing this dissertation, I kept repeating to myself two things Itziar used to say to me: (a) what is the question?, and (b) ¡al grano! (to the point!).

None of this would have been possible without the assistance of Edurne Petrirena and Yolanda Acebo. Edurne is a superhero; she takes care of everything so we can focus on research. The lab would collapse without her! Yolanda is our go-to person when technology tries to sabotage our experiments. If you've ever been to our building, you'd probably hear someone yelling "Yolanda!!!" desperately. That was probably one of us struggling with computers. I would also like to thank my colleagues from Gogo Elebiduna at the University of the Basque Country (UPV/EHU): Adam Zawiszewski, Beatriz Gómez, Gillen Martínez de la Hidalga, Itziar Orbegozo, Kepa Erdozia, Marta De Pedis, Mikel Santesteban, Miren Urteaga, Nerea Egusquiza... Thanks to Paolo Lorusso for always being willing to have a beer after work. Especially, I'd like to thank Idoia Ros and Luis Pastor for their friendship and for all the time we shared doing research in a tiny basement. Idoia has been a great friend and a mentor to me. Her dedication and hard work are an inspiration to all of us. Thank you for your patience with my thousands of questions and your guidance! Luis is the kind of colleague you'd always want to have around. He always tries his best to help you —especially with all the paperwork! Oh, and his baking skills are insane! Eskerrik asko denoi!

I also wish to thank Colin Phillips for giving me the opportunity to spend three months at the Department of Linguistics of the University of Maryland. Special thanks to Ellen Lau for introducing to the now-not-so-scary world of ERPs and for trusting in me. Thanks to Anna Namyst for helping me set up my experiment and teaching me how to run experiments. I am grateful for all the amazing people I met in College Park, both faculty and students: Aaron Doliana, Alexander Williams, Annemarie van Dooren, Anouk Dieuleveut, Anton Malko, Geosel Mendes, Jeff Lidz, Mina Hirzel, Nick Huang, Paulina Lyskawa, Rodrigo Ranero, Sigwan Thivierge, Tonia Bleam, Tyler Knowlton... Especially, thanks to my QBH roommates: Andrew, Mina, Paulina and Sigwan. You guys really made me feel like home!

[1] This dissertation has been funded by the Spanish Ministry of Education, Culture and Sports (FPU15/05741).

I had the chance to spend three months at the Department of Linguistics of the University of Potsdam. I'd like to thank Shravan Vasishth for hosting me. Shravan is such a great scholar and a wonderful person. I owe to him everything I know about statistics, and I will always have him as an example of how to do honest and good quality research. I'd also like to thank the faculty and students for making me feel like a member of the group: Bruno Nicenboim, Daniela Mertzen, Dario Paape, Dorothea Pregla, Kate Stone, Paula Lissón, Titus von der Malsburg…

My interest for linguistics started when I was an undergraduate student at the University of Oviedo. I was so lucky to have passionate professors who made me want to learn more about language. I would like to thank Francisco Martín Miguel, Goretti Faya Ornia, José Luis González Escribano, Mónica Moro and Roberto Valdeón. Special thanks to Ana Ojea, who believed in me since the beginning and accepted to be my advisor for my honors thesis. Ana encouraged me to do research in linguistics and I will always be grateful for that.

Exactly six years ago, Aurora Troncoso-Ruiz and I took a train to Vitoria-Gasteiz and moved in together. Little did we know we'd be like family, even now in the distance. Thank you for your unconditional friendship. Thanks also to Blanca Arias-Badia for always being there, despite her influencer duties; and to Natalia Jardón, for all the time we spent together. I'd like to thank Asier Calzada Lizarraga. Asier is one of those rare people you meet and feel that you've known each other for years. Thank you for being such a good friend, for making me feel at home, and for so many *kafetxos* complaining about everything. Also, I'd like to thank Katherine Elizabeth Fraser (aka Katie) for always being there to have a pizza and a *crianza*. A big shout-out to Aurora, Blanca and Katie for reading this dissertation and giving me feedback! Thanks to everyone who's been part of this journey, both at the university or in Vitoria-Gasteiz in general: Arantza López Fernández, Dasha Grigoreva, Diego Llamazares de Miguel, Laura Vela-Plo, Lucía Masa, Maite de la Fuente, Marina Ortega Andrés, Olympia Simantiraki (aka Olina), Rubén Pérez Ramón, Tania Mibdo, etc. I would also like to thank my friends from Gijón/Xixón. Even in the distance, they always care about me. When we see each other, it feels like being in high school again!

Quiero dar las gracias a mi famila. Estoy seguro de que ni mi madre ni mis hermanos saben de qué va esta tesis, pero eso es precisamente lo bueno. Con mi familia, siempre he tenido un espacio seguro al que acudir y olvidarme de todo el estrés. Gracias a mi madre, Pili, que se desvive por todos nosotros y que es una persona fundamental en mi vida. Gracias a mi padre, Nico, que murió al poco de empezar este proyecto. Espero que allá donde estés te sientas orgulloso de mí. Gracias a mis hermanos, Paula y David, y a mi cuñada, Vero. Gracias a Mateo, porque inundó la familia de alegría y siempre me saca una sonrisa con su energía inagotable. Gracies a los mios güelos, que trabayaron dende que yeren guah.es pa que'l nuesu futuru fuera meyor. Especialmente, a les mios güeles, que tocó-yos vivir tiempos de fame nun mundu d'homes. Por último, quiero dar las gracias a Pelayo. Estos últimos años no han sido fáciles ni para ti ni para mí, pero siempre has estado al pie del cañón. Brindemos por el futuro que nos espera.

Contents

List of Figures

List of Tables

List of Abbreviations

3PL	Third person plural
3SG	Third person singular
ACC	Accusative
ACT-R	Adaptive Control of Thought–Rational
Adv	Adverb
AFS	Active Filler Strategy
ATN	Augmented Transitions Network
AUX	Auxiliary verb
CL	Clitic
D-linked	Discourse-linked
DAT	Dative
DLT	Dependency Locality Theory
ERP	Event-related potential
EU	Energy unit
GEN	Genitive
IO	Indirect object
LV05	Lewis & Vasishth (2005) (activation-based model)
MU	Memory unit
NP	Noun Phrase
O	Object
OV	Object-Verb
P600	ERP component characterized by a positive peak after 600 ms after stimulus onset
POST	Postcritical
PRE	Precritical
PTCP	Participle
S	Subject
SAN	Sustained anterior negativity
SOA	Stimulus-onset asynchrony
V	Verb
VO	Verb-Object

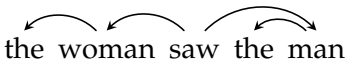
Para Amador Nicolás, Nico, Niquín, Amador, Madorín, Madorito.
Para papá.

Chapter 1

General Introduction

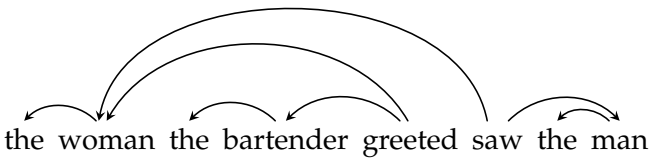
The words in a sentence are dependent on each other for a correct syntactic and semantic interpretation, thus, establishing a relationship of *dependency* between a head and a dependent (Tesnière, 1959). The head is the word that carries the features of a constituent and determines its function (Bloomfield, 1933), whereas dependents are modifiers. For example, in (1), *saw* —the head— is in a dependency relationship with both *woman* —the agent— and *man* —the theme. Likewise, both *woman* and *man* are the head of their respective noun phrase (NP), establishing a dependency with the determiner —*the*, in both cases.

(1) *Dependencies between adjacent words*



Because sentence comprehension is incremental and incoming words are integrated into a syntactic representation as the linguistic signal unfolds (Altmann & Kamide, 1999; Kaiser & Trueswell, 2004), one crucial challenge for the parser is that heads and dependents are not always adjacent. In fact, everyday communication involves both producing and understanding non-adjacent dependencies. In (2), *woman* is involved in two long dependencies: it is both the theme of the verb *greeted* and the agent of the verb *saw*.

(2) *Dependencies between adjacent and non-adjacent words*



The existence of dependencies between non-adjacent words implies our memory system needs to integrate information across intervening material by either actively maintaining information in memory and/or retrieving previously encoded representations. In this dissertation, I investigate the memory mechanisms involved in the processing of dependencies

between non-adjacent words, focusing on different dependency types, namely local and unbounded dependencies, and typologically diverse languages, namely English, Spanish and Basque. Specifically, I investigate two factors that affect the activation level of memory targets in sentence comprehension: (a) the representational complexity of antecedents, and (b) the linear distance between antecedents and heads.

Dependencies have played and still play a central role in both theoretical and experimental linguistics. From a theoretical perspective, this is more evident in approaches such as dependency grammar (Tesnière, 1959), which revolves around the notion of dependency and how words within a sentence are connected to one another. However, much work has been devoted to dependencies in other frameworks too. In generative linguistics, for example, long-distance dependencies are assumed to be the result of movement, a syntactic operation by which a constituent moves from one position to another leaving a trace or phonologically null copy in its original position (Chomsky, 1977). It is therefore assumed that constituents have been displaced from their canonical position, for example, in passivization (***The theater*** *was built* ***t*** *in 1956*), raising (***Katie*** *seems* ***t*** *to like wine*) topicalization (***That sandwich***, *I want to eat* ***t***) or *wh*-questions (***What*** *did you read* ***t***?).[1,2] Since Ross (1967), it is known that long-distance dependencies are constrained: constituents cannot move out of certain domains, metaphorically called *islands*. Core notions of different generativist models, such as *binding* and *barriers* in Government and Binding Theory (GB) (Chomsky, 1983) or *phases* in the Minimalist Program (MP) (Chomsky, 2000; Gallego, 2010), have tried to explain constraints on dependencies.

In psycholinguistics, researchers have tried to understand how humans integrate words or constituents that are further apart in a sentence in real time processing. There is a general consensus that understanding dependencies is taxing for our memory system. For example, it has long been noted that center-embeddings are difficult to understand, and having more than two center-embeddings in the same sentence hinders comprehension (compare 3a with 3b) (Chomsky & Miller, 1963).

(3) a. The man that [the cat scratched] went to the hospital.

b. The man that [the cat that [the dog chased] scratched] went to the hospital.

Increased processing difficulty has been tied to the assumption that the amount of information in our focus of attention, i.e. information that can be simultaneously processed, is highly limited (Miller, 1956; Baddeley & Hitch, 1974; Just & Carpenter, 1992; Cowan, 2001, 2010). For example, Cowan (2001) argues that focal attention capacity is limited to four elements, whereas McElree (2006) mantains that it is reduced to only one task-relevant

[1]Following standard notation in generative accounts, the trace or copy is represented in these examples with *t*.

[2]Generative accounts make a distinction between A-dependencies (e.g. passivization and raising) and A-bar dependencies (e.g. topicalization, *wh*-questions and relative clauses) depending on whether the displaced constituent occupies a structural argument position or not. For the purposes of this dissertation, I will not make such distinction.

representation. Given this limitation, it follows that our memory system must necessarily have a mechanism by which non-adjacent words can be related. On the one hand, it could be that the content of the antecedent —or some of its features— are stored in a separate memory state or memory buffer until it can be integrated into the syntactic representation (Baddeley, 1986). On the other hand, it could be that the antecedent is encoded in memory and brought back to focal attention via a retrieval operation (McElree, 2006).

The idea that the antecedent or filler is actively maintained in memory goes back to the *Augmented Transition Network* (ATN) model (Wanner & Maratsos, 1978), where the antecedent is believed to be held in a separate memory state —the HOLD cell— until it can be analyzed. Antecedent maintenance is also assumed in the *Active Filler Strategy* (hereafter, AFS) (Frazier, 1987; Frazier & Flores d'Arcais, 1989; Frazier & Clifton Jr, 1989), an influential account of *filler-gap* dependency processing.[3] The AFS states that the parser engages in an active search to find a suitable gap as soon as a filler is encountered and attempts to complete the dependency at the first available site. Evidence that the parser completes dependencies as soon as possible comes from the *filled-gap* effect (Stowe, 1986). For example, in (4a) the parser encounters the *wh*-pronoun *who* and attempts to complete the dependency at the verb *bring*, interpreting *who* as the object of the sentence. However, the next word (*us*) is the true object of the sentence, yielding a slowdown in reading times in comparison to the same word in (4b), where there is no dependency. Frazier (1987) argues that this eagerness to resolve the dependency may result from the cost associated with holding the filler in a memory buffer. However, early versions of the AFS do not specify whether a full representation of the antecedent is actively maintained. Recently, some researchers have argued that the parser only maintains syntactic category information, but not semantic content (Wagers & Phillips, 2013; Ness & Meltzer-Asscher, 2017).

(4) a. My brother wanted to know **who** Ruth will bring us home **to** at Christmas.

b. My brother wanted to know if Ruth will bring us home to Mom at Chrtistmas.

(Stowe, 1986, p. 234)

Not all accounts assuming memory maintenance assume the antecedent, or some of its features, is kept active in memory. For instance, *Dependency Locality Theory* (DLT) (Gibson, 1998, 2000) is a model of sentence comprehension that assumes dependency processing involves two mechanisms that make use of the same pool of memory resources (Just & Carpenter, 1992) and, therefore, increasing processing difficulty: memory *storage* and memory *integration*. In this model, memory storage refers to the cost of maintaining syntactic predictions in memory. That is, it is assumed that every antecedent triggers a prediction of a syntactic head. This means the parser must keep track of the number of incomplete dependencies while processing intervening words. Then, it is predicted that having more than one open dependency, as in (3a),

[3]The term *filler-gap* dependency is used in some psycholinguistic accounts influenced by generative grammar. Dependencies introduced by a *wh*-word, such as *wh*-questions or relative clauses, are assumed to be the result of a word or constituent —the *filler*— moving from one position to another, leaving a structural gap in its original position —the equivalent to a trace or copy.

is more costly than having simply one (3b). Memory integration refers to the use of cognitive resources in order to combine the antecedent with the verb. In this model, this cost depends on the number of intervening new discourse referents.

Retrieval models of sentence parsing (Van Dyke & Lewis, 2003; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke & McElree, 2006), on the contrary, assume that information is either in focal attention or not, without further assuming an intermediate memory state (McElree, 2006). Under this view, representations are shunted from focal attention in order to make way for incoming information and are encoded in a passive memory state. Information is assumed to be brought back to focal attention via a content-addressable, cue-based retrieval mechanism (M. C. Anderson & Neely, 1996). There is evidence that antecedents are retrieved at the verb coming from different experimental paradigms, such as cross-modal priming and speed-accuracy tradeoff (Nicol & Swinney, 1989; Osterhout & Swinney, 1993; McElree, 2000).

One influential sentence parsing model incorporating a content-addressable cue-based retrieval mechanism is the activation-based model of sentence comprehension (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006) (henceforth, LV05), which applies the cognitive architecture of Adaptive Control of Thought–Rational (ACT–R) (J. R. Anderson & Lebiere, 1998; J. R. Anderson, Qin, Stenger, & Carter, 2004; J. R. Anderson, 2005) to language processing. In this model, successful sentence comprehension depends on successful retrieval of previously accessed information and the fluctuating level of activation of memory targets.

In a cue-based retrieval event, the contents of memory encodings are matched in parallel against a set of retrieval cues (J. R. Anderson et al., 2004). Retrieval success is guaranteed as long as there is a match between the retrieval cues and the encoding of the target. In addition, the more distinct the retrieval cues are, the higher the probability of retrieving the right target. The advantage of this mechanism is that only encodings matching the retrieval cues are activated, whereas irrelevant representations are ignored. As a consequence, retrieval speed is not negatively affected by the set size, unlike in serial memory searches (Sternberg, 1966). However, when other encodings share cues with the target, all representations partially matching the retrieval cues are activated in memory, resulting in memory interference. That is, other encodings sharing cues with the target also receive activation. There is ample evidence of similarity-based interference in language comprehension, both proactive (Van Dyke & McElree, 2006, 2011) and retroactive (Gordon, Hendrick, & Johnson, 2001; Van Dyke & Lewis, 2003; Vasishth & Lewis, 2006; Van Dyke & McElree, 2011). For example, Van Dyke and McElree (2006) ran an experiment where participants had to remember a a three-word list (*table–sink–truck*), read a sentence and answer a comprehension question. Then, they were asked to recall the three words they had seen at the beginning. Results showed that the semantic fit of the verb with the words to be recalled negatively impacted sentence comprehension, yielding longer reading times at the verb when they were semantically compatible (5a) than when they were not (5b). These results suggest that the retrieval cues of *fixed* activated all the encodings in memory, resulting in interference at retrieval, whereas *sailed* only activated *boat*.

(5) a. It was **the boat** that the guy who lived by the sea **fixed** in two sunny days.

b. It was **the boat** that the guy who lived by the sea **sailed** in two sunny days.

Van Dyke and McElree, 2006, p. 160

The level of activation of memory items determines the retrieval success in LV05. Memory targets whose representation is more activated are retrieved faster (J. R. Anderson et al., 2004; Vasishth & Lewis, 2006). Activation is affected by different factors. First, the activation level of an item decays from memory as a function of time. This predicts that the longer the time between encoding and retrieval, the less active an item is and the more costly it is to recover it from memory. Second, activation is split between all the items sharing cues with the target, decreasing the activation of all of them. Third, repeated access to a representation boosts its activation in memory.

In this dissertation, I focus on the real time course of dependency processing in order to investigate how retrieval is affected by (a) the representational complexity of antecedents and (b) dependency length. Specifically, I investigate what features increase the level of activation of antecedents in memory by focusing on whether semantic and/or syntactic complexity facilitates memory retrieval in English. I also investigate the interplay between activation decay and predictability by examining the effect of increasing dependency length in different dependency types —unbounded and local— and languages with opposite basic word orders, namely Spanish (OV) and Basque (VO).

1 Antecedent complexity

Research on explicit memory recall has found that elaboration and distinctiveness enhance retrieval form memory of items in isolation (Von Restorff, 1933; J. R. Anderson & Reder, 1979; Bradshaw & Anderson, 1982; McDaniel, 1981; McDaniel, Dunay, Lyman, & Kerwin, 1988; Gallo, Meadow, Johnson, & Foster, 2008) —a phenomenon commonly known as *Von Restorff* or isolation effect (Von Restorff, 1933). Available evidence shows that additional cues such as semantic elaboration (Bradshaw & Anderson, 1982), syntactic complexity (McDaniel, 1981) and distinctive superficial features such as typography (Kolers, 1973) facilitate memory recall of items that participants were instructed to remember. It has been argued that additional features increase the cognitive cost of encoding these items in memory but, in turn, memory traces last longer, as in the *levels of processing* model (Craik & Lockhart, 1972; Lockhart, Craik, & Jacoby, 1976).

In sentence comprehension, memory retrieval is an implicit operation necessary to complete dependencies in real time. For example, in the sentence *I saw the woman the journalist interviewed last night*, *the woman* must be retrieved at the verb *interviewed* for a full understanding of the sentence. Speakers have the option to produce antecedent NPs referring to the same entity with

different degrees of semantic and/or syntactic complexity. For example, both *the woman* and *the businesswoman* may refer to the same person, but the latter is semantically more complex, whereas *the world-renowned businesswoman* is both semantically and syntactically more complex. If elaboration or distinctiveness facilitate the retrieval of memory targets also in real time language processing comprehension, the representational complexity of antecedents should impact retrieval times.

Sentence processing models based on cue-based retrieval mechanisms predict that additional cues contribute to the uniqueness of memory targets, distinguishing them from potential competitors and reducing the probability of similarity-based interference. However, only the cues that are relevant for retrieval are predicted to reduce interference, since non-relevant cues are ignored. In terms of activation (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006), it is possible that the higher cognitive cost associated with processing additional cues may increase the activation level of an item, speeding up its subsequent retrieval from memory, in line with Craik and Lockhart (1972) and Lockhart et al. (1976). This idea has been formulated as the *Memory Facilitation Hypothesis* (Hofmeister, 2007, 2011), which predicts longer encoding times for complex antecedents, since more cognitive resources are needed to process them, but faster retrieval times due to a higher level of activation.

Previous experimental studies show that additional semantic information facilitate memory retrieval. In self-paced reading studies focusing on *wh*-dependencies and relative clauses, adding an extra adjective to the antecedent (e.g. *the communist* vs. *the alleged communist*) yields faster reading times at the verb, suggesting faster memory retrieval (Hofmeister, 2007, 2011; Hofmeister & Vasishth, 2014). In addition, superiority condition violations (Kuno & Robinson, 1972; Chomsky, 1973) and island violations (Ross, 1967) are given higher acceptability ratings when the antecedent is semantically richer relative to a baseline with a simpler antecedent (Hofmeister, 2007; Hofmeister & Sag, 2010). Complexity effects are not restricted to relative clauses and *wh*-dependencies, where *that* or a *wh*-phrase signals clearly identifies the beginning of an antecedent. For example, processing a third person singular pronoun such as *she* or *he* requires retrieving an NP referent. Experimental data shows that semantically more complex NPs are more likely to be interpreted as the referent of a pronoun (Karimi & Ferreira, 2016) and are retrieved more easily (Karimi, Swaab, & Ferreira, 2018).

Overall, there is a general consensus that the semantic complexity of memory targets boosts memory activation and facilitates memory retrieval. Distinctive extra-linguistic features, however, do not seem to play a role. For example, Hofmeister and Vasishth (2014) failed to find evidence of processing facilitation when manipulating the color of the antecedent. Regarding syntactic complexity, evidence is less clear. In most studies, additional semantic information entails increasing the complexity of the NP syntactic internal structure. To the best of my knowledge, there is only one study where semantic and syntactic complexity were assessed separately. Hofmeister (2011) ran a self-paced reading study comparing two antecedents with different semantic complexity, while keeping syntactic complexity constant (*which soldier* vs. *which person*). Results show faster reading times at the verb for the semantically richer

antecedent. However, this only indicates a facilitation effect driven by semantic complexity and does not entail syntactic complexity does not contribute to the overall activation level of the antecedent.

2 Dependency length

The linear distance between an antecedent and its head —commonly referred to as *dependency length*— is one of the most studied phenomena in psycholinguistics, both in sentence production and comprehension. Different measures have been used to talk about dependency length in the literature, such as syntactic complexity, measured as the number of intervening phrasal nodes (Ferreira, 1991), or the number of intervening novel discourse referents (Gibson, 1998, 2000). In most recent studies, dependency length is operationalized as simply the number of words between an antecedent and its head (Gildea & Temperley, 2010; Futrell, Mahowald, & Gibson, 2015a; Hawkins, 1994, 2004). Since all these metrics have been proved to be strongly correlated (Wasow, 2002; Szmrecányi, 2004), I will use the latter throughout this dissertation due to its easier implementation.

In language production, it has been observed that speakers tend to minimize the distance between antecedents and heads, yielding shorter dependencies. This general preference is commonly known as *dependency length minimization* (DLM). DLM has been argued to be a general cognitive bias affecting syntactic choice and shaping grammars, as well as a driving force of language change (Tily, 2010). Hawkins (1994, 2004) argues that word orders that shorten the distance between co-dependents are predicted in production since they are more efficient and require fewer computations, facilitating sentence comprehension. For example, in a ditransitive sentence in English such as (6), having a relative clause attached to the head of O, as in (6a), increases the dependency length between V and IO (in bold). Speakers have a tendency to produce shorter constituents first, as in (6b), hence minimizing the distance between co-dependents. This preference for minimizing dependencies has been observed in typologically diverse languages, yielding opposite linearizations in OV languages (Choi, 2007; Ros, Santesteban, Fukumura, & Laka, 2015). This general preference has also been attested in large corpus crosslinguistic studies (Liu, 2008; Gildea & Temperley, 2010; Gildea & Jaeger, 2015; Futrell et al., 2015a). For example, Futrell et al. (2015a) measured dependency length in 37 languages and reported that average dependency length was shorter than chance level in all of them.

(6) a. The girl **gave** the ball she had gotten for Christmas to the **boy**.

b. The girl **gave** the **boy** the ball she had gotten for Christmas.

In sentence comprehension, increased dependency length has been identified as a source of comprehension difficulty. There is a wide range of experimental evidence showing that

adding intervening words increases reading times at the verb (Gibson, 1998, 2000; Grodner & Gibson, 2005; Bartek, Lewis, Vasishth, & Smith, 2011), a phenomenon commonly called *locality* effect. For example, the slowdown at the verb in object relative clauses in comparison to subject relative clauses has been argued to be the result of having intervening words between co-dependents (Gibson, 2000; Grodner & Gibson, 2005).

(7) a. The reporter **who sent** the photographer to the editor hoped for a story.

b. The reporter **who** the photographer **sent** to the editor hoped for a story.

(Grodner and Gibson, 2005, p. 266)

This preference for shorter dependencies in comprehension is predicted by different accounts based on the limitation of memory resources, which I will call *memory-based* accounts. On the one hand, DLT predicts a greater integration cost when new discourse referents intervene (Gibson, 1998, 2000). The reason is that more cognitive resources are needed in order to identify the referent relative to referents that are already part of the discourse, such as the first and second singular personal pronouns *I* and *you*. On the other hand, LV05 attributes increased difficulty to activation decay of the representation from memory as a function of time. Additionally, longer dependencies are more likely to include words which share cues with the antecedent, resulting in similarity-based interference at retrieval.

Locality effects have consistently been found in VO languages, such as English (Grodner & Gibson, 2005), Chinese (Hsiao & Gibson, 2003) and Russian (Levy, Fedorenko, & Gibson, 2013). However, studies investigating OV languages, such as German (Konieczny, 2000; Konieczny & Döring, 2003) and Hindi (Vasishth, 2003; Lewis & Vasishth, 2005), have found an *anti-locality* effect, that is, distance-based processing facilitation. LV05 propose that pre-verbal arguments re-activate representations of the antecedent in memory, resulting in an activation boost and facilitating memory retrieval at the verb. Alternatively, the difference observed between VO and OV languages could be that general word order patterns impact the predictability of upcoming linguistic elements. Expectation-based accounts of language processing predict that the probability of a word appearing in a given context impacts processing facilitation (Hale, 2001; Levy, 2008). In OV languages, preverbal arguments intervening between and antecedent and its head may sharpen the predictions with regards to upcoming words, facilitating sentence processing.

Available evidence of both locality and antilocality effects is limited in at least two ways. First, most studies finding locality effects investigate exclusively *wh*-dependencies and relative clauses, whereas those finding antilocality focus on argument-verb dependencies in declarative sentence. The former are unbounded, meaning that dependencies can span across more than one clause; the latter, however, are local dependencies that must be resolved within the same clause. This difference may have an impact on the predictability of the head of the dependency, since in local dependencies the head must be in the same clause, unlike in

unbounded dependencies. Second, current evidence of antilocality effects is limited by (i) the small number of languages tested and (ii) the fact that all OV languages investigated are strictly head-final. It has been suggested that antilocality effects arise in OV languages because verbs are predicted to appear in final position (Levy & Keller, 2013). However, some OV languages have a higher degree of word order flexibility, and the position of the verb may be less predictable. In this dissertation, I address these two gaps in order to shed light on the nature of both locality and antilocality effects in sentence comprehension and to understand the interaction between memory decay and expectation-driven facilitation.

3 Aims and outline

In what follows, I provide a brief outline of the main research questions addressed in each of the following three chapters (Chapters 2 to 4). Overall, the main goal of this dissertation is to investigate two factors that determine the activation level of memory representations and, hence, their subsequent retrieval from memory. Specifically, I study (a) whether and how the representational complexity of an antecedent impacts its subsequent retrieval from memory (Chapter 2), and (b) activation decay as a function of time and its interaction with expectation-driven facilitation by investigating (i) different types of dependencies —unbounded and local— (Chapter 3) and (ii) an OV language with flexible word order (Chapter 4).

- In **Chapter 2**, I focus on the effect of antecedent complexity in on-line dependency comprehension. I investigate the individual contribution of additional semantic and syntactic complexity to the antecedent's subsequent retrieval from memory. In general, previous evidence suggests the parser benefits from encoding additional semantic cues of the antecedent, as these cues contribute to the uniqueness of that representation in memory, both enhancing its activation level and making it more distinct from potential competitors. However, extra semantic cues often involve creating syntactically more complex constituents that are phonologically longer. I also investigate the neural correlates of the memory mechanisms involved in dependency processing. Some researchers have identified at least two event-related potential (ERP) components: a P600 at the verb, which has been thought to reflect syntactic integration difficulty, and a sustained anterior negativity throughout the dependency, which is believed to index memory maintenance costs. In Experiment 1, I used the event-related potentials (ERPs) technique to compare the ERP components elicited by sentences with no dependency in comparison to sentences containing *wh*-dependencies, where the complexity of the *wh*-phrase was manipulated. This resulted in three conditions: a bare wh-filler (*who*), a two-word syntactically more complex antecedent (*which person*), and a two-word syntactically and semantically more complex antecedent (*which waiter*). Results show a modulation of the P600 effect as a function of syntactic complexity, suggesting longer antecedents are easier to retrieve due to a higher level of activation. I found no evidence

of a sustained effect throughout the dependency, providing support to retrieval models of sentence comprehension that assume there is no need to maintain information actively in memory.

- In **Chapter 3**, I investigate the effect of increasing dependency length on different types of dependencies, namely unbounded and local dependencies. Memory-based models of sentence comprehension (DLT and LV05) predict a locality effect due to either (a) increased integration difficulty or (b) activation decay and a higher probability of interference, respectively. Locality effects have been found across the board in VO languages, such as English, but not in OV languages, where the opposite effect —antilocality— has been found. Notably, evidence of locality comes from experiments focusing exclusively on unbounded dependencies, whereas evidence of antilocality comes from experiments testing unbounded dependencies. I hypothesize that the distribution of locality and antilocality effects could be explained in terms of dependency type. This is the first time, to my knowledge, that such a hypothesis has been made. The reason behind this is that unbounded dependencies can span across more than one clause, increasing the uncertainty as to when the dependency will be completed. Local dependencies, however, must necessarily be completed within the same clause, rendering the head of the dependency more predictable. Evidence from an acceptability judgment task (Experiment 2) and two self-paced reading studies (Experiments 3 & 4) manipulating dependency length in both local and unbounded dependencies in Spanish revealed a locality effect for both dependency types, showing no difference between them. An additional cloze-completion task (Experiment 4) showed that the head of the dependency did not differ in terms of predictability, suggesting the parser prefers to complete dependencies as soon as possible. Overall, results support the idea that locality is a general phenomenon in VO languages.

- In **Chapter 4**, I investigate antilocality effects in Basque, an unstudied OV language with free word order. Previous studies provide evidence that locality effects are dominant in VO languages, whereas antilocality effects are exclusively found in OV languages. Furthermore, evidence from Chapter 2 favors the view that locality is a general phenomenon in VO languages, regardless of dependency type. These differences have been attributed to a greater exposure to verb-final clauses in OV languages, increasing the expectation for the head of the dependency and mitigating locality effects. Crucially, the available evidence is constrained in two ways: first, the number of OV languages tested is very small; second, in all these the verb is strictly final. In an acceptability judgment task (Experiment 6) and a self-paced reading (Experiment 7) task, I investigated whether antilocality also arises in Basque, a language where the verb may appear in either medial or final position. Testing declarative ditransitive sentences with canonical word order (S-IO-O-V) and where the linear distance between IO and V was manipulated, behavioral results show a clear preference for longer dependencies. In addition, I ran a cloze-completion task (Experiment 8) in order to calculate the probability of verb-medial word orders as a function of constituent length, revealing a tendency to produce verbs in

final position, probably due to written language being more formal. In general, results support the idea that antilocality can be observed in all OV languages regardless of word order flexibility.

- In **Chapter 5**, I summarize the main findings of this dissertation and sketch ideas for future research.

Chapter 2

The impact of antecedent complexity on dependency processing*

Abstract

Encoding elaborate representations facilitates memory performance in explicit memory recall tasks (Von Restorff, 1933; Bradshaw & Anderson, 1982). In sentence comprehension, results from self-paced reading studies involving *wh*-dependencies and relative clauses show that encoding a syntactically and semantically complex antecedent leads to faster reading times at the verb (Hofmeister, 2007, 2011; Hofmeister & Vasishth, 2014). This suggests that the retrieval of complex antecedents is facilitated due to a higher activation level. Even though semantic complexity has been proposed to be the key factor (Hofmeister & Vasishth, 2014; Karimi et al., 2018), semantic and syntactic complexity are often confounded. The current ERP study investigated the impact of additional semantic and syntactic complexity in *wh*-dependency processing in English by comparing dependencies with *who*, a syntactically more complex *wh*-phrase (*which person*), and a syntactically and semantically more complex *wh*-phrase (*which*-N). In line with previous studies, I report a P600 effect at the verb in comparison to a baseline for all three conditions, which has been suggested to reflect syntactic integration difficulty (Kaan, Harris, Gibson, & Holcomb, 2000). Notably, the P600 was smaller when the antecedent was syntactically more complex, but additional semantic complexity did not further decrease the amplitude of the effect. I failed to find a sustained anterior negativity (SAN) throughout the dependency, which is believed to index memory maintenance cost (Fiebach, Schlesewsky, & Friederici, 2002). Results support the idea that syntactic complexity also facilitates memory retrieval in sentence comprehension. This effect may be due to either the additional processing cost associated with a more complex syntactic structure or the additional time at encoding. The lack of a SAN suggests the content of the antecedent is not actively maintained in a memory buffer (Wanner & Maratsos, 1978) and provides evidence against a memory storage cost in dependency processing (Gibson, 1998, 2000).

*The work presented here was carried out in collaboration with Prof. Ellen Lau during a research stay at the Department of Linguistics of the University of Maryland.

1 Introduction

An NP describing an entity can be more or less informative. For example, both *the woman* and *the businesswoman* may refer to the same person in the world, but the latter is semantically richer because it is more specific. If we add more information, for example, *the world-renowned businesswoman*, this phrase is both semantically richer and syntactically more complex than *the woman*. Assuming phrases are encoded in memory as a bundle of features when comprehenders read or hear them, more informative phrases contain a higher number of features. Following Hofmeister (2011), I will use the term *complexity* to talk about phrases, where x and y are phrases describing an entitity e, and x is more complex than y when the features of y are contained in x.

There is an extensive number of studies reporting that the complexity of an encoding impacts its subsequent retrieval from memory. Specifically, both elaboration and distinctiveness of the target improve memory recall of items in isolation (Von Restorff, 1933; J. R. Anderson & Reder, 1979; Bradshaw & Anderson, 1982; McDaniel, 1981; McDaniel et al., 1988; Gallo et al., 2008), a well-known effect called the *Von Restorff* or *isolation* effect (Von Restorff, 1933). The *levels of processing* model (Craik & Lockhart, 1972) states that targets that are more costly to encode in memory are processed in more depth and, hence, they leave a long-lasting trace in memory. Lockhart et al. (1976) further specify that greater difficulty at encoding may be the result of increased semantic, syntactic or phonological processing. The underlying idea is the intuition that the unusual is remembered better.

In her seminal work, Von Restorff (1933) presented a series of experiments showing incongruent items are recalled more easily.[2]. At the sentence level, there is evidence that superficial distinct features such as unfamiliar typography enhance memory performance (Kolers, 1973). Bradshaw and Anderson (1982) investigated the role of semantic elaboration and provided evidence that participants are able to recall facts about historical personalities more successfully when they are followed by sentences elaborating on either their cause or their result, in comparison to when non-related or no elaboration follows. Other features such as syntactic structure choice also contribute to memory distinctiveness. McDaniel (1981) exposed participants to sentences with the same propositional content but different degrees of syntactic complexity. From most to least difficult, the experimental conditions were: a double center-embedding without a relative pronoun (1a), a double center-embedding with a relative pronoun (1b), and a single center-embedding (1c). Results show that comprehenders are able to recall syntactically complex sentences with a higher success rate than simpler ones.

(1) a. The car the man the dog bit drove crashed.

b. The car that the man whom the dog bit drove crashed.

c. The car driven by the man that the dog bit crashed.

[2]The original paper is published in German, but see Hunt (1995) for a summary and review in English.

In the studies reported above, participants were explicitly told beforehand that their task was to memorize the materials. In sentence comprehension, however, memory encoding and retrieval are implicit processes necessary to establish dependencies between words and phrases. For instance, in (2), the antecedent *the novel* must be retrieved at the verb *lent* to get a full understanding of the sentence. One crucial question is whether the complexity effect observed in explicit memory recall tasks is also observable in implicit memory retrieval in language comprehension.

(2) I've read **the novel** that Susan **lent** me.

Currently available experimental evidence supports the idea that the representational complexity of an antecedent facilitates its memory retrieval in online dependency processing. For example, Hofmeister (2007) measured reading times at the verb in sentences containing a *wh*-dependency where the complexity of the *wh*-phrase was manipulated: a one-word *wh*-phrase (3a), a two-word *which*-N phrase (3b), or a three word *which*-N phrase (3c).

(3) a. **What** did the reporter that Scooter avoided **discuss** during an evening news segment?

b. **Which poll** did the reporter that Scooter avoided **discuss** during an evening news segment?

c. **Which political poll** did the reporter that Scooter avoided **discuss** during an evening news segment?

Results show that reading times sped up at the verb when there was either a two-word or three-word *which*-N phrase in comparison to when there is a bare *wh*-word. In another experiment, Hofmeister (2007) increased the representational complexity of antecedents in relative clauses by attaching additional adjectives: *a communist*, *an alleged communist* and *an alleged Venezuelan communist*. Participants read the verb faster when there was either one or two adjectives compared to the bare noun condition. These results have been replicated in subsequent studies (Hofmeister, 2011; Hofmeister & Vasishth, 2014). Kaan et al. (2000) provide the only piece of counter-evidence. In an experiment using the event-related potential (ERP) technique, they found that dependencies containing a *which*-N phrase elicited a larger P600 component at the verb relative to dependencies headed by *who*, suggesting a greater integration cost for the former.

Complexity effects have also been reported in dependencies other than *wh*-questions and relative clauses (Karimi, Fukumura, Ferreira, & Pickering, 2014; Karimi & Ferreira, 2016; Karimi et al., 2018). For example, processing a third person pronoun such as *she* or *he* involves retrieving a referent from memory. Using the visual-world paradigm, Karimi and Ferreira (2016) provided evidence that whenever two NPs could be the antecedent of a pronoun, comprehenders have a tendency to pick the most complex one as the referent. In a follow-up

study, Karimi et al. (2018) investigated the ERP response at a pronoun when it referred back to either a long or a short NP, showing a larger late and sustained frontal negativity —the Nref effect— when the antecedent was short in comparison to when it was long. The authors argue longer NPs were semantically richer in their manipulation and, hence, they were more accessible in memory. These studies provide evidence in support of the *Informational Load Hypothesis* (Almor, 1999, 2004), which states that the greater the conceptual distance between an anaphor and its antecedent, the smaller the informational load at reactivation.

Cue-based retrieval parsing models (Van Dyke & Lewis, 2003; Lewis et al., 2006; Van Dyke & McElree, 2006) predict that additional features make a memory target more distinct from its competitors, reducing the probability of similarity-based interference at retrieval. However, only features matching the retrieval cues are relevant for retrieval. For example, if the retrieval cues of a verb specify that the target must be an object NP, both *the novel* and *the world-renowned historical novel* match those cues, even though the latter is semantically and syntactically more complex. Complexity effects could be accommodated by the ACT-R computational model of sentence comprehension (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006), where fluctuating activation is a major determinant of either processing facilitation through activation boosts or difficulty due to activation model. Under this view, complex representations may be more activated in memory and are therefore more accessible for retrieval operations.

The studies presented above suggest that at least semantic complexity seems to facilitate the retrieval of memory targets. Hofmeister and Vasishth (2014) investigated whether superficial extra-linguistic features of the antecedent, such as text color, could also impact retrieval in relative clause comprehension but failed to find an effect. The role of syntactic complexity is less clear, since in most cases semantic elaboration is confounded with syntactic complexity. That is, semantically richer NPs often include additional words that increase phonological length and the complexity of the syntactic internal structure. To my knowledge, only one study has addressed this confound. Hofmeister (2011) ran a self-paced reading sentences comparing a sentence containing a *wh*-dependency where semantic complexity of the antecedent was manipulated, while keeping syntactic complexity constant (*which soldier* vs. *which person*). Results show comprehenders read the verb faster when the antecedent was semantically richer. However, the lack of a baseline with a syntactically and semantically simpler antecedent (e.g. *who*) does not allow the assessment of the role of syntactic complexity alone.

In this chapter, I tease apart the impact of semantic and syntactic complexity of an antecedent on its subsequent retrieval from memory using ERPs. In what follows, I provide a review of studies showing that the complexity of *wh*-phrases also has an impact on the acceptability of sentences violating syntactic constraints on *wh*-dependencies. I also provide an overview of the neural correlates of *wh*-dependency processing in the ERP literature, focusing on the P600 as an index of syntactic integration difficulty and a sustained negative component (SAN) as a signal of memory maintenance cost. The former allowed me to investigate memory retrieval facilitation, whereas the latter is especially relevant for sentence processing accounts assuming a storage cost throughout the dependency (Wanner & Maratsos, 1978; Frazier, 1987;

Frazier & Flores d'Arcais, 1989; Frazier & Clifton Jr, 1989; Gibson, 1998, 2000).

1.1 Antecedent complexity and acceptability

It has long been observed that complex *wh*-phrases improve the acceptability of sentences that violate some syntactic constraint on *wh*-dependencies. In English, whenever there are two *wh*-phrases within the same clause, only the one which is structurally higher may be fronted to form a question, as captured by the *superiority* condition (Kuno & Robinson, 1972; Chomsky, 1973). In (4), *who* is structurally higher than *what* because it occupies the subject position. Therefore, fronting *who* yields an acceptable sentence in English (4a), but fronting *what* does not (4b). In early generative grammar studies, it was observed that complex *wh*-phrases make sentences where the superiority condition is violated sound more acceptable (Karttunen, 1977; Pesetsky, 1987). For example, (4c), where both *wh*-phrases are complex, is preferred over (4b). Hofmeister, Jaeger, Sag, Arnon, and Snider (2007) ran a controlled acceptability judgment task using sentences with multiple *wh*-words where the superiority condition was violated, manipulating the complexity of both *wh*-phrases. Results showed a higher degree of acceptability for sentences containing two *which*-N phrases over those containing only one or none.

(4) a. Who will get what?

b. *What will who get?

c. ?Which drug will which patient get?

(Karttunen, 1977)

Similarly, complex *wh*-phrases improve the acceptability of island violations too. Islands are constituents where a dependency cannot be completed (Ross, 1967), such as complex NPs (CNPs) or clauses headed by a *wh*-phrase (*wh*-islands). Attempting to complete a dependency between a dependent outside an island and a head within an island results in unacceptability, as illustrated in (5).

(5) a. ***What** did John hear [$_{\text{CNP}}$ the rumor that the man **repaired**]?

b. ***What** did John wonder [$_{\text{WH}}$ whether the man **repaired**]?

Some islands, known as *weak* islands, are permeable and allow the establishment of dependencies under certain conditions. Cinque (1990) noted that the acceptability of *wh*-island violations improved when a complex *wh*-phrase was involved in the dependency (6).

(6) ?**Which car** did John wonder [$_{\text{WH}}$ whether the man **repaired**]?

Controlled experimental work has provided further evidence that *which*-N phrases improve the acceptability of *wh*-island violations (Hofmeister & Sag, 2010; Alexopoulou & Keller, 2013; Villata, Rizzi, & Franck, 2016). In all these acceptability judgment tasks, results consistently show that *wh*-island violations are given higher acceptability ratings when a complex *wh*-phrase is involved in the dependency than when there is a bare *wh*-word.

Other islands, such as CNPs, are considered *strong* since they have been claimed to be immune to factors improving the acceptability of *weak* islands. However, controlled experimental work has provided evidence that CNPs too receive higher acceptability ratings when a *which*-N is involved in the dependency (Hofmeister & Sag, 2010; Goodall, 2015). This is shown in (7). This finding casts doubt on the classical distinction between strong and weak islands and raises the question of what properties make complex *wh*-phrases more resistant to grammatical constraints previously thought of as rigid.

(7) ?**Which car** did John hear [$_{\text{CNP}}$ the rumor that the man **repaired**]?

Pesetsky (1987) argued that this improvement in acceptability was related to the connection of the *wh*-phrase with the context: *wh*-phrases are D(iscourse)-linked, whereas bare *wh*-words are not. D-linked *wh*-phrases presuppose a set that both speaker and hearer(s) have in mind, thus constraining the range of possible felicitous answers to a question. For example, in a question such as *which car did John buy?*, the only felicitous answer would be a specific car out of a set —which may include all the cars in the world or be restricted to some cars that have been previously mentioned in the conversation. Non-D-linked *wh*-phrases, on the other hand, do not limit the range of possible answers. If discourse accounts for the differences in acceptability between bare and *which*-N phrases in superiority and island violations, we should be able to find a similar improvement if we provide a rich context for a bare *wh*-phrase. That is, we should be able to D-link *what* or *who*. Sprouse (2007) tested this hypothesis in an acceptability judgment task in which he included superiority violations in questions with both *which*-N and bare *wh*-phrases. Before the questions were displayed, participants were given a rich context, so that in both cases the *wh*-phrase was linked to the discourse (example from Sprouse 2007, p. 75):

> Last semester, Professor Smith assigned 36 books for his literature students to read, but he knows that no one read all of them. In fact, he's pretty sure that each book was read by only one student, so he wants to only order the books that the literature majors read. Before placing the book order for the next semester, he thinks to himself…

After that, participants were shown a question violating the superiority condition, with either a simple one-word *wh*-phrase (8a) or a complex *wh*-phrase (8b), and were asked to rate it using a 7-point Likert scale. Acceptability judgment data shows that the presence of a context does not erase the difference between simple and complex antecedents. In other words,

even though the bare *wh*-phrase was linked to the discourse, no improvement was observed, suggesting that the formal properties of complex phrases must play a role.

(8) a. I wish I knew what who read.

b. I wish I knew which book which student read.

In addition to an increase in acceptability, a reading speed-up at the verb when antecedent complexity is increased has also been observed in sentences violating the superiority condition (Hofmeister et al., 2007) and islands (Hofmeister, 2007; Hofmeister & Sag, 2010). Together with acceptability data, these results have been taken as evidence against the assumption that the superiority condition and islands are part of core grammatical constraints. Rather, Hofmeister and Sag (2010) argue islands effects are the result of additive extra-grammatical processing difficulty.

1.2 ERP indices of *wh*-dependency processing

Previous ERP studies focusing on *wh*-dependencies have identified two components: a P600 after reading the subcategorizing verb and a sustained anterior negativity (SAN) throughout the dependency. Each of these components have been suggested to reflect different memory processes during dependency processing: the P600 is linked to syntactic integration difficulty in general, whereas the SAN indexes the cost of memory maintenance, supporting the idea that active memory maintenance plays a role in dependency processing.

The P600 is an ERP component characterized by a positive-going wave peaking at around 500-700 ms after the onset of the stimulus with a posterior distribution. In early work, the P600 was found as a response to a wide range of syntactic violations and syntactically ambiguous sentences (*garden path* sentences) (Neville, Nicol, Barss, Forster, & Garrett, 1991; Osterhout & Holcomb, 1992; Hagoort, Brown, & Groothusen, 1993; Friederici, Pfeifer, & Hahne, 1993; Kaan & Swaab, 2003; Allen, Badecker, & Osterhout, 2003). The general assumption was that the P600 signalled a syntactic anomaly or syntactic reanalysis, as opposed to the N400, a negative-going wave peaking at around 400 ms, which was believed to be sensitive to syntactic anomalies (but see Lau, Phillips, and Poeppel, 2008 for a review). However, this view has been challenged by ERP studies finding P600 responses in other contexts, e. g. semantic anomalies (Kolk, Chwilla, Van Herten, & Oor, 2003; Kim & Osterhout, 2005; Van Herten, Kolk, & Chwilla, 2005).

Crucially, the P600 has been reported in at least three studies focusing on well-formed unambiguous *wh*-dependencies (Kaan et al., 2000; Fiebach et al., 2002; Phillips, Kazanina, & Abada, 2005), suggesting it indexes syntactic integration difficulty in general. Kaan et al. (2000) compared the electrical response elicited at the verb in sentences containing a *wh*-dependency headed by either *who* or a *which*-N phrase to a baseline with no dependency. Results show *wh*-dependencies elicited a positive response relative to the control after 500 ms after the verb

onset. This positivity was larger for *which*-N phrases than for *who*, suggesting an increased processing cost. Kaan and colleagues argue *which*-N phrases should be more costly precisely because they encode more information, which is contrary to the findings reported above. Fiebach et al. (2002) ran a similar study with indirect subject and object questions and control *whether*-sentences with no dependency in German. Results show a late positivity for object questions in comparison to both subject questions and controls right before the verb. Phillips et al. (2005) also report a P600 in an ERP study comparing *which*-N questions to *that*-sentences where dependency length was manipulated, showing a delayed effect for longer dependencies.

The sustained anterior negativity (SAN) is a negative slow wave that is observed across a long period of time with a left anterior distribution. It was first reported in studies investigating explicit memory retention. For example, Ruchkin, Johnson Jr, Canoune, and Ritter (1990) ran an ERP study in order to investigate the time-course of explicit retention of information in working memory. During the task, participants had to memorize a series of consonants and information load was manipulated depending on the number of consonants to be recalled. This task was combined with a search task where no memory retention is needed. Results show a long frontal negativity only in the retention task, suggesting this component is reflecting memory maintenance.

The SAN has also been found in sentence comprehension studies where there is no explicit memory retention. To the best of my knowledge, there are three studies that have found a SAN during sentence comprehension of dependencies (King & Kutas, 1995; Fiebach et al., 2002; Phillips et al., 2005). King and Kutas (1995) ran an experiment where participants had to read subject and object relative clauses in English. In both cases, there is an unbounded dependency between the head of the relative clause and the embedded verb, but object relatives are longer. Results show a SAN most prominent over left anterior electrodes in the object relative clause in comparison to the subject relative clause. This finding has been interpreted as indexing a higher memory cost due to memory maintenance of the antecedent in object relatives, while this cost is not present in subject relatives because they are resolved earlier in the sentence. Similarly, Fiebach et al. (2002) reported a left-anterior SAN in object indirect questions in German in comparison to both subject indirect questions and controls. Additionally, the participants' working memory capacity was measured with a reading span task (Daneman & Carpenter, 1980) and further analyses revealed the SAN was larger for individuals with low working capacity, suggesting that more resources had been used in order to process object *wh*-questions. One striking finding is the absence of a SAN in subject *wh*-questions relative to the baseline. In fact, the authors report a tendency in the contrary direction, i.e. the baseline without dependency was more negative. Phillips et al. (2005) tested *wh*-questions in English, where the complexity of the *wh*-phrase was manipulated (*who* vs *which*-N). In both conditions, Phillips and colleagues found a robust sustained negativity throughout the dependency relative to a baseline with no dependency. However, unlike previous studies, the effect was not restricted to left-anterior electrodes, but was spread all over the scalp. This sustained negativity must be interpreted with caution because the authors carried out baseline correction at different words across conditions. This may have spuriously created the illusion of a sustained effect.

The SAN has been taken as evidence of active memory maintenance in dependency processing. The *Active Filler Strategy* (henceforth, AFS) (Fodor, 1978; Clifton & Frazier, 1989) assumes that unbounded dependencies, such as *wh*-dependencies and relative clauses, are costly because the antecedent has to be held in a distinguished memory state or memory buffer until it is integrated into the syntactic structure, an idea that goes back to Wanner and Maratsos (1978). Recent versions of this approach suggest that only some features of the *wh*-phrase are actively maintained in memory. Wagers and Phillips (2013) and Ness and Meltzer-Asscher (2017) argue that syntactic category information is actively held in memory, whereas semantic information is retrieved. *Dependency Locality Theory* (hereafter, DLT) (Gibson, 1998, 2000) assumes dependency processing incurs in both a storage and an integration cost. However, it differs from the maintenance approach in the type of information that is held in memory. In this case, it is assumed that unresolved dependencies are costly because a syntactic prediction must be maintained, i.e. the parser has to keep track of how many incomplete dependencies there are. There is also reading time data compatible with both the AFS and DLT. In a number of self-paced reading studies, overall increased reading times have been reported throughout a sentence when there is a dependency in comparison to when there is not (Grodner, Gibson, & Tunstall, 2002; Chen, Gibson, & Wolf, 2005; Stepanov & Stateva, 2015). In addition, Ness and Meltzer-Asscher (2017) found a processing disruption at a similar intervening noun only in sentences involving a relative clause in Hebrew, but not in NP-ellipsis sentences where there is no *wh*-phrase.

2 Experiment 1

2.1 Overview

The present study is set out to investigate the individual contribution of additional semantic and syntactic complexity to memory retrieval facilitation. Previous evidence suggests the representational complexity of antecedents facilitates sentence comprehension at the verb. There is a general agreement that semantic complexity by either distinctiveness (*the woman* vs. *the businesswoman*) or elaboration (*the woman* vs. *the most successful businesswoman in Spain*) is a key factor (Hofmeister & Vasishth, 2014; Karimi et al., 2018). However, the role of syntactic complexity is less clear. In order to address this gap, I manipulated the complexity of *wh*-phrases in English in the present ERP study, resulting in three conditions: a bare *wh*-phrase (*who*), a syntactically more complex *wh*-phrase (*which person*), and a syntactically and semantically more complex *wh*-phrase, (*which*-N), such as *which waiter*. The electrophysiological response to these sentences was compared to a baseline with no *wh*-dependency.

The following are the predictions for the current experiment. In line with previous results (Kaan et al., 2000; Fiebach et al., 2002; Phillips et al., 2005), all the conditions containing a *wh*-dependency are predicted to yield a P600 effect at the verb relative to the baseline. This is

because memory retrieval of the antecedent should reflect a greater integration cost compared to a *that*-clause with no antecedent retrieval. If only semantic distinctiveness plays a role in memory retrieval facilitation, I predicted *which*-N to elicit a smaller P600 at the verb than *which person* and *who*. If only syntactic complexity facilitates retrieval, I expect the P600 to be reduced in both the *which*-N and *which person* conditions in comparison to *who*. Finally, if both syntactic and semantic information contribute to retrieval facilitation, I expected to find a gradient effect, where the P600 is largest for *who* and smallest for *which person*.

Available evidence of complexity effects focus exclusively on encoding and retrieval processes, assuming a retrieval model of sentence comprehension where the antecedent is not actively maintained in memory. However, both the AFH and DLT do predict a memory maintenance cost, which should be indexed by a SAN throughout the dependency. Assuming an early version of the AFH, where the antecedent is believed to be stored in a separate memory buffer (Wanner & Maratsos, 1978), the whole *wh*-phrase should be held in memory throughout the dependency and both syntactic and semantic complexity should increase memory demands, yielding the largest SAN component for *which person* and the smallest one for *who*. If only syntactic category information is held in memory, as suggested by Wagers and Phillips (2013) and Ness and Meltzer-Asscher (2017), I expect to find a SAN in all three *wh*-conditions, but no modulation of this effect in terms of *wh*-complexity. This same prediction holds for DLT, where a syntactic prediction cost is expected for all sentences containing a *wh*-dependency, with no difference between conditions.

2.2 Participants

42 undergraduate students from the Univeristy of Maryland (27 women, mean age = 20.44, SD = 2.68, range = 18–34) took part in the experiment. All participants were native speakers of English, were right-handed based on the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal vision and no history of neurological disorders. Participants gave informed consent and received either $15/hour or course credit for their participation.

Participants whose accuracy in the comprehension questions fell below 70% were excluded from the analysis. Likewise, if ocular artifact rejection affected over 40% of the trials from a participant, they were removed. This procedure affected the number of participants included in the P600 and in the SAN analyses differently because epoch length was different —1200 ms and 5000 ms, respectively (see SECTION 2.7 for a detailed description of the epochs used for the analyses). An additional participant left part-way through the experiment because they were not feeling well and was therefore removed from the analysis. Overall, 38 participants were included in the analysis of the integration cost at the verb (P600) and 35 participants were included in the analysis of the maintenance cost (SAN).

Condition	Example
That	The manager knew **that** the new owner of the coffee shop would *fire* the waiter after the scandal.
Who	The manager knew **who** the new owner of the coffee shop would *fire* after the scandal.
Which-person	The manager knew **which person** the new owner of the coffee shop would *fire* after the scandal.
Which-N	The manager knew **which waiter** the new owner of the coffee shop would *fire* after the scandal.

TABLE 2.1: Sample set of the sentences used in Experiment 1. The experimental manipulation is in bold and the critical region in italics.

2.3 Materials

The materials for the study were 120 experimental items and 120 filler items of similar length. A third of the experimental sentences and a third of the filler sentences were followed by a yes/no comprehension question. In the experimental materials, I used sentences including an indirect *wh*-question where I manipulated the complexity of the filler, resulting in three *wh*-conditions: a bare *wh*-filler (*who*), a two-word syntactically more complex *wh*-filler (*which person*) and a two-word syntactically and semantically more complex *wh*-filler (*which*-N). A *that*-clause with no filler-gap dependency was used as the baseline. All items started with an NP followed by a verb (*The manager knew...*) or with a proper name followed by and adverb and a verb (*Ben already forgot...*) or by a verb in the passive form (*Joey was told...*). I used assertive verbs that are compatible with either a declarative *that*-clause or an indirect *wh*-question. Between the *wh*-phrase and its subcategorizing verb there were eight words consisting of an NP modified by either a PP or an *-ing* clause (*...the new owner of the coffee shop would...* or *...the lovely couple enjoying their honey moon finally...*). An example set of the experimental sentences is given in TABLE 2.1.

The 120 experimental items were distributed across four experimental lists following a Latin-square design, so that each list contained only one version of each item. The same set of 120 filler items was used for each list. All the items within each list, , including both experimental and filler items, were randomized.

2.4 Procedure

Participants were comfortably seated in front of a computer screen in a testing room. Sentences were presented one word at a time in 24-point white font on a black background in the center of the screen using Psychtoolbox for Matlab (Brainard & Vision, 1997; Pelli, 1997; Kleiner et al., 2007). Each sentence was preceded by a fixation cross that appeared for 500 ms. Each word appeared on the screen for 300 ms, followed by 200 ms of blank screen (i.e. 500 ms SOA) (see FIGURE 2.1). The last word of each sentence was marked with a period. For a third of the items, after 1000 ms there was a yes/no comprehension question, and participants were asked to answer it by pressing one of two buttons. Participants were instructed to avoid eye blinks and movements during the presentation of the sentences, and they were asked to read each sentence attentively in order to answer the comprehension questions. Prior to

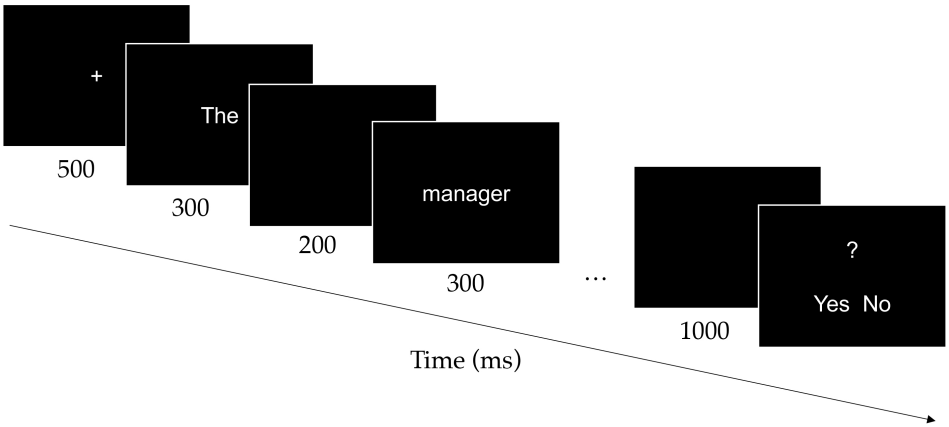


FIGURE 2.1: Experimental procedure in Experiment 1.

the experimental session, participants were presented with six practice trials with feedback to familiarize themselves with the task. The experimental session was divided into six blocks of 30 sentences each, with short pauses in between. After each block participants were given feedback about their accuracy in the comprehension questions. Including set-up time, an experimental session lasted approximately two hours on average.

2.5 EEG recording

Twenty-nine tin electrodes were held in place on the scalp by an elastic cap (Electro-Cap International, Inc., Eaton, OH) in a 10–20 configuration (O1, O2, P7, P3, Pz, P4, P8, TP7, CP3, CPz, CP4, TP8, T7, C3, Cz, C4, T8, FT7, FC3, FCz, FC4, FT8, F7, F3, Fz, F4, F8, FP1, FP2). Bipolar electrodes were placed above and below the left eye and at the outer canthus of both eyes to monitor vertical and horizontal eye movements. Additional electrodes were placed over the left and right mastoids. Scalp electrodes were referenced online to the left mastoid and re-referenced offline to the average of left and right mastoids. Impedances were maintained at less than 10 kΩ for all scalp electrode sites, and less than 5 kΩ for mastoid and ocular sites. The EEG signal was amplified by a NeuroScan SynAmps®Model 5083 (NeuroScan, Inc., Charlotte, NC) with a bandpass of 0.05–100 Hz and was continuously sampled at 500 Hz by an analog-to-digital converter.

2.6 Preprocessing

A 0.01–30 Hz digital band-pass filter including a 60 Hz notch filter was applied to the EEG data. Then, eye movements were corrected using independent component analysis (ICA) (Jung et al., 2000) on a 1-Hz high-pass filtered version of the data. After this, trials including segments containing a difference of over 120 μV in a timewindow of 150 ms or a voltage step of 50 μV/s were rejected. Participants with more than 40% of rejected trials were removed from the analysis, affecting 1 participant in the integration analysis and 4 participants in the

maintenance analysis. The data was preprocessed with the EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes in Matlab (MATLAB, 2017).

2.7 Analysis

First, in order to analyze the integration cost at the verb (henceforth, *integration* analyses), I computed event-related potentials separately for each participant and each trial for the 1200 ms after the onset of the main verb (*fire* in TABLE 2.1) relative to a 100 ms baseline preceding the critical region. I selected two time windows based on visual inspection: 500–700 ms and 700–900 ms.

Second, in order to analyze the cost of maintaining the filler throughout the dependency (henceforth, *maintenance* analyses), event-related potentials were computed for the 5000 ms after the onset of *that* or the *wh*-word (*who* or *which*) relative to a 200 ms baseline. I decided not to start the epoch in the first word after the filler to avoid any artifact resulting from baseline-correcting in different words across conditions. I extracted the mean amplitude of each word following *that* or the *wh*-word (*who* or *which*) up until the subcategorizing verb and compared the mean amplitude elicited for each word across conditions. Since the *which person* and *which*-N conditions had one more word than the *that* and *who* conditions, I only compared the ERP elicited by the words that were repeated across all four conditions (*the new owner of the coffee shop would* in the example in TABLE 2.1). This yielded 8 time windows, each corresponding to one word

I ran linear mixed models on the averaged voltage amplitude of either centro-parietal electrodes (CP3, CPz, CP4, P3, Pz, P4) in the integration analyses or left anterior electrodes (F7, F3, FT7, FC3) in the maintenance analyses using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in the R (version 3.5.3) programming environment (R Core Team, 2017), including *wh-phrase* as fixed effect. The most parsimonious model, including only varying random intercepts and slopes by subjects and by items, was fitted and complexity was added to the random-effects structure when supported by the data in order to avoid model overfitting (Bates, Kliegl, Vasishth, & Baayen, 2015). All p-values were calculated using the Satterthwaite's model of approximation using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017). A separate analysis was carried out for each time-window described above.

ERP data has traditionally been analyzed running analyses of variance (ANOVA) but there are at least two disadvantages in comparison with linear mixed models. First, unlike in linear mixed models, the random variability in the data due to participants and the experimental items cannot be simultaneously entered into the analysis. Performing two different analyses may lead to contradictory results. Second, in cases where there are experimental factors with more than two levels —such as the present experiment, where there is one experimental factor with four levels—, the source of an effect is unclear (Schad, Vasishth, Hohenstein, & Kliegl, 2020). That is, it is not clear which conditions are different from which. This is usually solved by running post-hoc analyses, which introduce additional problems such as the loss of statistical

power and multiple comparisons. In a regression model such as linear mixed models, a priori contrasts can be defined in order to make the relevant comparisons (see Schad et al. 2020 for a comprehensive review). In both the maintenance and the integration analyses, I was interested in comparing (a) each *wh*-condition (*which*-N, *which person* and *who*) with the baseline (*that*) and (b) each *wh*-condition with each other. For that purpose, two analyses were run with different sets of contrasts for the fixed factor. *Treatment* contrasts were specified for (a), whereas *sliding difference* or *repeated* contrasts were specified for (b).

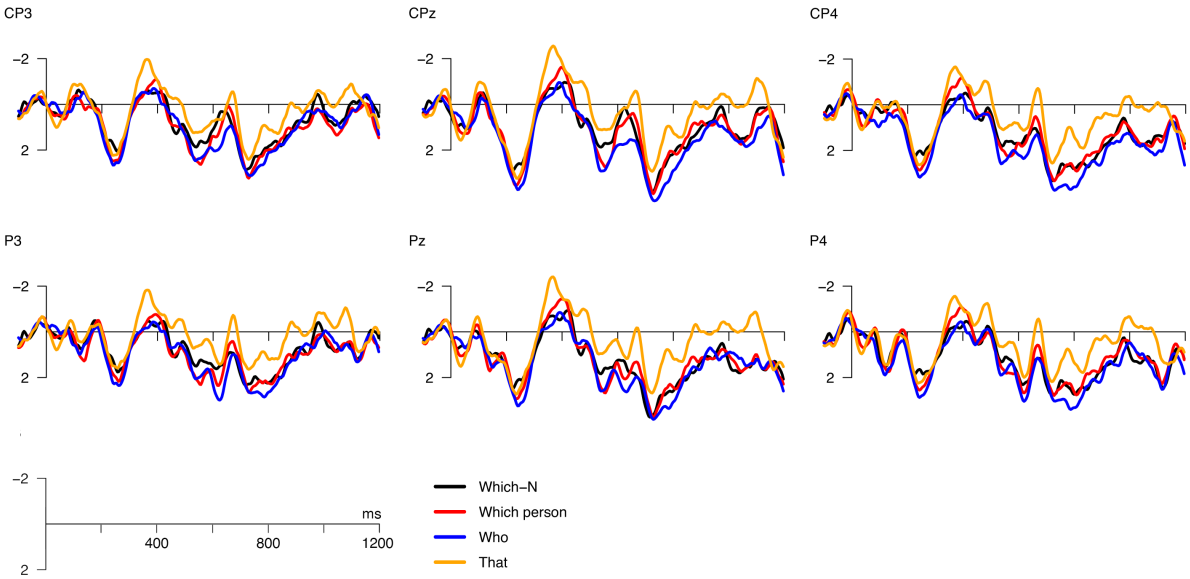


FIGURE 2.2: Grand average ERP response at the verb on centro-parietal electrodes (CP3, CPz, CP4, P3, Pz, P4) in Experiment 1.

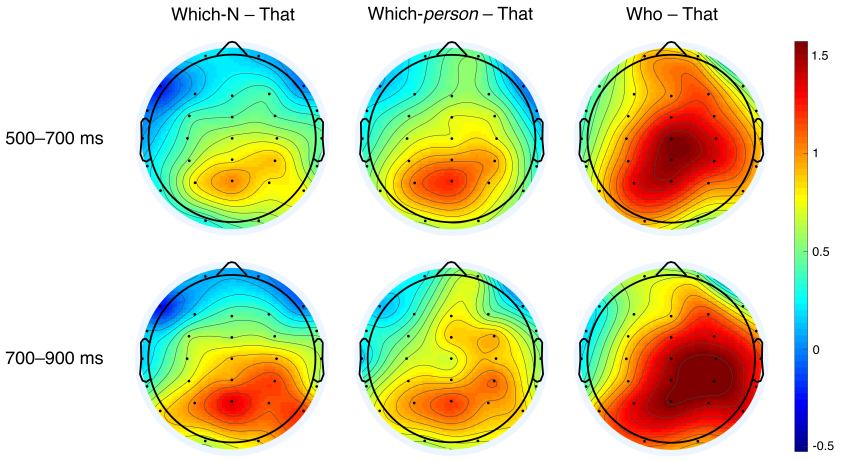


FIGURE 2.3: Topographic maps based on average voltage differences between *wh*-conditions and the baseline (*that*) between 500–700 ms and between 700-900 ms in Experiment 1.

2.8 Results

For clarity of exposition, the results section is divided into two subsections: one reporting the results from the integration analyses at the verb (P600) and the other reporting the results from the maintenance analyses (SAN).

Integration cost (P600)

FIGURE 2.2 shows the grand average event-related potential response at the subcategorizing verb. FIGURE 2.3 shows the topographic maps based on the average voltage difference between each of the *wh*-conditions (*which*-N, *which person* and *who*) and the baseline (*that*) between 500-700 ms and 700-900 ms. Visual inspection of the data reveals that the three *wh*-conditions elicited a positive-going response starting at approximately 400 ms after the onset of the verb with respect to the baseline. In addition, the positivity elicited by *who* is larger than that elicited by complex *wh*-phrases (*which*-N and *which person*). The results from the linear mixed models using treatment contrasts are reported in TABLE 2.2. Results revealed a significant difference between each of the *wh*-conditions and the baseline in both the 500-700 and the 700-900 time windows. The results from the linear mixed models using sliding difference or repeated contrasts are reported in TABLE 2.3. In both the 500–700 and the 700–900 time windows, results revealed a significant difference between *who* and *which person*. However, the difference between *which person* and *which*-N did not reach significance. Overall, all *wh*-conditions elicited a P600 response with respect to the baseline, and this component had a larger amplitude when the dependency was headed by *who* than when there was a syntactically complex *wh*-phrase (*which*-N or *which person*).

	Estimate	SE	t-value	p-value
500–700 ms				
(Intercept)	0.62	0.31	2.00	0.05
Which-N	0.76	0.16	4.79	<0.001
Which-person	0.92	0.16	5.78	<0.001
Who	1.30	0.16	8.21	<0.001
700–900 ms				
(Intercept)	1.26	0.28	4.58	<0.001
Which-N	1.05	0.17	6.33	<0.001
Which-person	0.93	0.17	5.61	<0.001
Who	1.41	0.17	8.51	<0.001

TABLE 2.2: Linear mixed model analysis at the final verb on 500-700 ms and 700-900 ms time windows using treatment contrasts.

Maintenance cost

FIGURE 2.4 shows the grand average event-related potential response at the words following the *wh*-phrase or *that* until the subcategorizing verb. Note that conditions including a complex *w*h-phrase are one word longer and therefore intervening words are not aligned in the plot.

	Estimate	SE	t-value	p-value
500–700 ms				
(Intercept)	1.36	0.29	4.62	<0.001
Which-N vs. That	0.76	0.16	4.79	<0.001
Which-person vs. Which N	0.16	0.16	1.01	0.31
Who vs. Which-person	0.38	0.16	2.42	0.01
700–900 ms				
(Intercept)	2.11	0.26	8.23	<0.001
Which-N vs. That	1.05	0.17	6.33	<0.001
Which-person vs. Which N	-0.12	0.17	-0.71	0.48
Who vs. Which-person	0.48	0.17	2.89	0.004

TABLE 2.3: Linear mixed model analysis at the final verb on 500-700 ms and 700-900 ms time windows using sliding difference or repeated contrasts.

This plot was only used for visual inspection of the data; analyses were carried out aligning the mean voltage response elicited at each specific word. FIGURE 2.5 shows the average voltage elicited at each word of the dependency on left anterior electrodes (F7, F3, FT7, FC3) at each word after either the *wh*-phrase or *that* (*the new owner of the coffee shop*). Visual inspection of the data revealed that at W1, *who* and *that* patterned together and elicited a more negative response than *which*-N and *which person*. At W2, all four conditions elicited a similar electrical response. From W3 to W8, there seems to be no difference between *which*-N, *which person* and *who*, which were all more positive than *that*. Overall, there is no evidence of a sustained negativity of *wh*-conditions with respect to the baseline. In fact, there seems to be an effect in the opposite direction.

Results from the linear mixed models using treatment contrasts are reported in TABLE 2.4. At W1 and W2, *which*-N and *which person* were more positive than *that*, but *who* did not differ from the baseline. At W3, the pattern reversed and *who* was more positive than the baseline, but *which*-N and *which person* did not differ from *that*. At W4, there was no difference between any of the *wh*-conditions and the baseline. At W5 and W6, *which*-N and *who* were elicited a more positive response than *that*, but *which person* was not significantly different from the baseline. At W7, *who* patterned with *that*, but *which*-N and *which person* were more positive. Finally, at W8 all three *wh*-conditions differed from the baseline. In sum, I failed to find a sustained negativity on left anterior electrodes at any of the regions of interest. In fact, some or all of the *wh*-conditions were more positive than *that* in a number of words.

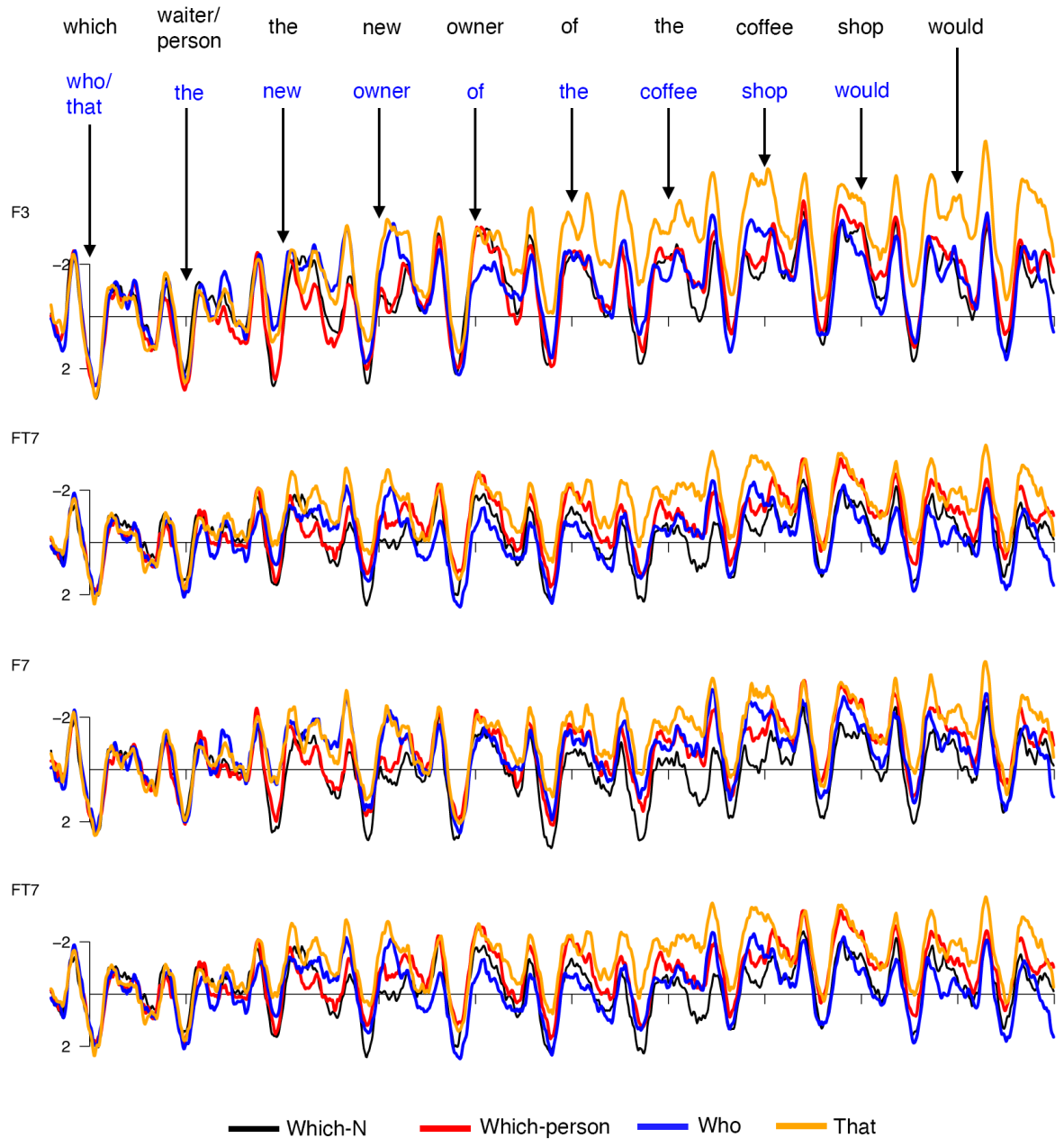


FIGURE 2.4: Grand average ERP response between the *wh*-phrase or *that* and the subcategorizing verb (*which waiter/which person/who/that the new owner of the coffee shop would*) on left anterior electrodes (F3, FC3, F7, FT7).

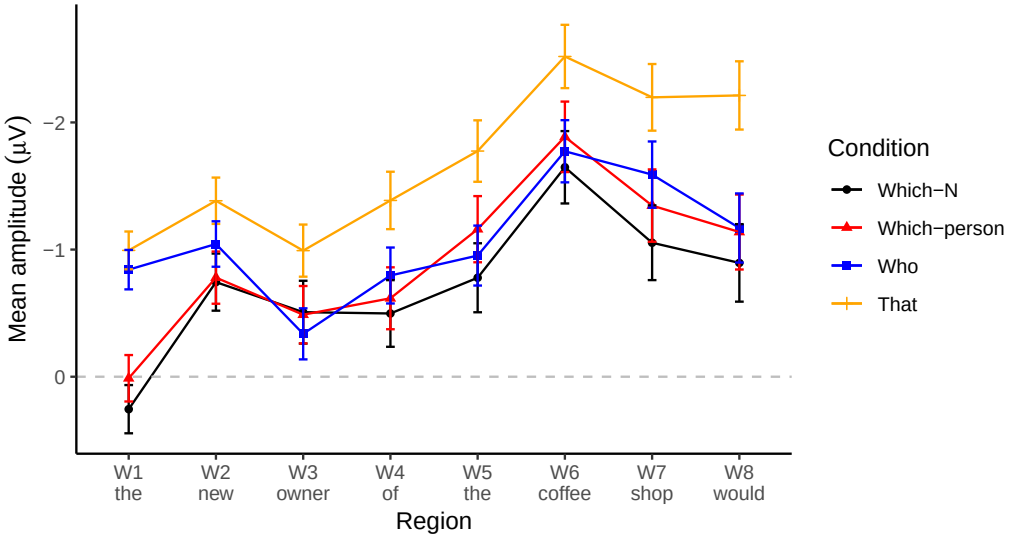


FIGURE 2.5: Mean voltage amplitude elicited at each of the words between the *wh*-phrase or *that* and the subcategorizing verb (*the new owner of the coffee shop would*) at left anterior electrodes (F3, FC3, F7, FT7). Error bars show standard error.

Results from the linear mixed models using sliding difference or repeated contrasts are reported in TABLE 2.5. The relevant results here are the difference between *which person* and *which*-N and between *who* and *which person*, as the difference between *which*-N and *that* was already reported in the analyses using treatment contrasts. At W1, there was a difference between *who* and *which person*, but not between *which person* and *which*-N. From W2 to W8, there was no difference between *which person* and which-N, nor between *who* and *which person*. Overall, all three *wh*-conditions elicited a similar response with no significant differences across the dependency.

3 Discussion

Previous studies have provided evidence showing faster reading times at the verb for dependencies where complex antecedents were involved (Hofmeister, 2007, 2011; Hofmeister & Vasishth, 2014), suggesting that additional semantic cues increase the activation of memory targets (Hofmeister & Vasishth, 2014; Karimi et al., 2018). However, the role of syntactic complexity was not clear due to the fact that semantic complexity was confounded with syntactic complexity in previous experiments. In the ERP experiment presented in this chapter, I investigated the role of semantic and syntactic complexity separately by examining the neural correlates of *wh*-dependency processing: the P600 as an index of syntactic integration difficulty, and the SAN as an index of memory maintenance cost.

Consistent with previous work, I reported a P600 effect at the verb for all three conditions containing a *wh*-dependency in comparison to the baseline, showing a greater integration cost

	Estimate	SE	t-value	p-value
Word 1 (the)				
(Intercept)	-1.01	0.29	-3.44	<0.001
Which-N	1.26	0.24	5.32	<0.001
Which-person	0.99	0.24	4.18	<0.001
Who	0.16	0.24	0.66	0.51
Word 2 (new)				
(Intercept)	-1.37	0.38	-3.58	<0.001
Which-N	0.65	0.28	2.35	0.02
Which-person	0.59	0.28	2.15	0.03
Who	0.33	0.28	1.19	0.24
Word 3 (owner)				
(Intercept)	-0.97	0.45	-2.14	0.04
Which-N	0.48	0.31	1.57	0.12
Which-person	0.48	0.31	1.58	0.11
Who	0.65	0.31	2.12	0.03
Word 4 (of)				
(Intercept)	-1.36	0.57	-2.40	0.02
Which-N	0.91	0.60	1.51	0.13
Which-person	0.82	0.57	1.44	0.15
Who	0.59	0.52	1.13	0.26
Word 5 (the)				
(Intercept)	-1.76	0.45	-3.90	<0.001
Which-N	1.01	0.35	2.88	0.004
Which-person	0.61	0.35	1.74	0.08
Who	0.85	0.35	2.42	0.02
Word 6 (coffee)				
(Intercept)	-2.49	0.51	-4.89	<0.001
Which-N	0.91	0.37	2.48	0.01
Which-person	0.63	0.37	1.72	0.09
Who	0.78	0.37	2.13	0.03
Word 7 (shop)				
(Intercept)	-2.19	0.55	-3.99	<0.001
Which-N	1.19	0.38	3.11	0.002
Which-person	0.85	0.38	2.21	0.03
Who	0.65	0.38	1.69	0.09
Word 8 (would)				
(Intercept)	-2.20	0.56	-3.89	<0.001
Which-N	1.36	0.40	3.43	<0.001
Which-person	1.09	0.40	2.75	0.01
Who	1.06	0.40	2.68	0.01

TABLE 2.4: Summary of the coefficients from the linear mixed models at each intervening word in Experiment 1 using treatment coding.

	Estimate	SE	t-value	p-value
Word 1 (the)				
(Intercept)	-0.41	0.26	-1.60	0.12
Which-N – That	1.26	0.24	5.32	<0.001
Which-person – Which-N	-0.27	0.24	-1.14	0.26
Who – Which-person	-0.83	0.24	-3.53	<0.001
Word 2 (new)				
(Intercept)	-0.98	0.34	-2.85	0.01
Which-N – That	0.65	0.28	2.35	0.02
Which-person – Which-N	-0.06	0.28	-0.20	0.84
Who – Which-person	-0.27	0.28	-0.96	0.34
Word 3 (owner)				
(Intercept)	-0.56	0.41	-1.37	0.18
Which-N – That	0.48	0.31	1.57	0.12
Which-person – Which-N	0.00	0.30	0.00	1.00
Who – Which-person	0.17	0.30	0.55	0.58
Word 4 (of)				
(Intercept)	-0.79	0.42	-1.89	0.07
Which-N – That	0.90	0.33	2.73	0.01
Which-person – Which-N	-0.14	0.33	-0.44	0.66
Who – Which-person	-0.15	0.33	-0.46	0.65
Word 5 (the)				
(Intercept)	-1.15	0.40	-2.89	0.01
Which-N – That	1.01	0.35	2.88	0.004
Which-person – Which-N	-0.40	0.35	-1.14	0.25
Who – Which-person	0.24	0.35	0.68	0.50
Word 6 (coffee)				
(Intercept)	-1.91	0.46	-4.18	<0.001
Which-N – That	0.91	0.37	2.48	0.01
Which-person – Which-N	-0.28	0.37	-0.77	0.44
Who – Which-person	0.15	0.37	0.42	0.68
Word 7 (shop)				
(Intercept)	-1.52	0.49	-3.06	0.003
Which-N – That	1.19	0.38	3.11	0.002
Which-person – Which-N	-0.35	0.38	-0.90	0.37
Who – Which-person	-0.20	0.38	-0.52	0.60
Word 8 (would)				
(Intercept)	-1.32	0.51	-2.59	0.01
Which-N – That	1.36	0.40	3.43	<0.001
Which-person – Which-N	-0.27	0.40	-0.68	0.50
Who – Which-person	-0.03	0.40	-0.07	0.94

TABLE 2.5: Summary of the coefficients from the linear mixed models at each intervening word in Experiment 1 using sliding difference or repeated contrasts.

at the verb when a dependency is involved due to retrieval from memory. I interpret this as an indicator of greater integration difficulty in comparison to the baseline due to the retrieval of the antecedent from memory. In addition, results show both *which*-N and *which person* elicited a smaller P600 effect than *who*. That is, syntactically more complex *wh*-phrases facilitated memory retrieval at the verb. However, there was no difference between *which*-N and *which person*, suggesting extra semantic features did not further ease memory retrieval. It could be the case that the effect size of semantic complexity is smaller than the effect of syntactic complexity and more statistical power is needed to detect a difference between *which person* and *which*-N. These possibilities should be further addressed in future work.

It is possible that syntactically more complex *wh*-phrases overall had a greater impact on retrieval because they were displayed for a longer time than bare *wh*-phrases. Previous self-paced reading studies report increased processing times when reading complex antecedents relative to simpler antecedents (Hofmeister, 2007, 2011; Hofmeister & Vasishth, 2014). This has been taken as evidence that more cognitive resources are needed in order to compute additional information, which, in turn, enhances memory retrieval. It should be noted that complexity in these experiments was generally incremented by adding extra words, increasing both semantic and syntactic complexity. As a result, complex antecedents were phonologically longer and took more time to read. In fact, Karimi and Ferreira (2016) and Karimi et al. (2018) discuss that phonological longer NPs are more accessible in memory than shorter NPs. A plausible hypothesis is that longer encoding times also impact memory activation. In Experiment 1, participants did not have control over the amount of time they spent reading the antecedents. Sentences were presented one word at a time with a 500 ms SOA, so two-word complex *wh*-phrases were displayed for twice the time as *who* —1000 ms vs. 500 ms, respectively. Extra encoding time may have contributed to a higher level of activation in memory, impacting memory retrieval.

Even though semantic complexity alone did not yield an effect in Experiment 1, other experimental data suggests syntactic complexity alone cannot explain complexity effects. For example, Hofmeister (2011) compared reading times of sentences containing *wh*-dependencies where semantic complexity was manipulated without adding syntactic complexity (e.g. *which person* vs. *which soldier*). This manipulation did not increase the length of the antecedent. Yet, results show a marginal effect at the antecedent and a strong effect in the following word, showing slower reading times for the semantically complex condition. That is, additional semantic cues alone increase reading times due to additional processing, even though the antecedents were syntactically equally complex.

Contrary to previous studies, I failed to find a SAN throughout the dependency. In fact, results point in the opposite direction: all *wh*-conditions elicited a more positive response in comparison to the control *that*-sentence. It is possible that previous studies reporting a left anterior SAN in *wh*-processing may have been affected by the choices made in either the experimental design or the pre-processing of the EEG data. For example, King and Kutas (1995) directly compared the ERP elicited over the course of subject and object relative clauses where

words did not align, i.e. different words were compared. In (9), I have numbered the words that were directly compared: an adverb against a determiner, a verb against a noun, etc. The neurophysiological response elicited at these words may have been influenced by comparing different word classes (e.g. function and content words), categories (e.g. verbs and nouns), and/or words with different frequencies.

(9) a. The **reporter** who harshly$_1$ **attacked**$_2$ the$_3$ senator$_4$ admitted$_5$ the error.

b. The **reporter** who the$_1$ senator$_2$ harshly$_3$ **attacked**$_4$ admitted$_5$ the error.

(King and Kutas, 1995, p. 379)

Phillips et al. (2005) compared *which*-N questions to *that*-sentences and, as in Experiment 1, sentences containing *which* necessarily included an extra word. In order to avoid comparing different words, the authors decided to apply baseline correction on different words: *which* in the case of the *wh*-questions, and the verb in the case of *that*-sentences. Overall, these results suggest the parser is sensitive to the amount of information in memory targets, increasing their level of activation in memory and their subsequent retrieval from memory.

The absence of a sustained effect over the course of a dependency can be interpreted as a lack of an additional processing cost associated with memory maintenance. A storage cost was predicted by both the AFS, which states the parser keeps the antecedent active in memory, and DLT, which states the parser must keep track of syntactic predictions. Results support the idea that *wh*-phrases are encoded in memory and are subsequently retrieved from memory instead of remaining active. One crucial advantage of retrieval models in comparison to a maintenance account is that encoding and retrieval are general memory mechanisms that do not depend on the type of dependency involved. That is, if one assumes a representation must be kept active in memory or that incomplete dependencies must be tracked incurring in an increased cognitive cost, the parser must clearly identify antecedents. However, this is only possible when the antecedent contains a clear cue, such as a *wh*-word in *wh*-dependencies and relative clauses. Therefore, a retrieval account unifies dependency processing and predicts similar complexity effects for all types of dependencies, such as in pronoun processing (Karimi & Ferreira, 2016; Karimi et al., 2018). In line with my results, Foraker and McElree (2011) reviewed studies using speed-accuracy tradeoff methodology and discussed that the comprehension of a wide range of non-adjacent dependencies (e.g. subject-verb, pronoun co-reference and relative clauses) is mediated by direct-access retrieval, and not by active maintenance.

In addition, the lack of a SAN suggests that it may not be necessary to assume a separate memory state where information is kept actively maintained (Cowan, 1995, 2001). In contrast, bipartite models assume that information is either in focal attention or in a passive memory state, from which it can be brought back to focal attention via a retrieval operation (McElree, 2000, 2006). It must be noted that discussing the existence of a separate memory state goes beyond the scope of this dissertation and no clear conclusion should be drawn from my results.

4 Concluding remarks

In this chapter, I investigated the role of syntactic and semantic complexity of antecedents in real-time dependency comprehension. Previous research on this topic has provided evidence that semantic complexity increases the level of activation of memory representations, facilitating their subsquent retrieval. However, the role of syntactic complexity was not clear. In an ERP experiment using *wh*-questions in English, I provided novel evidence that the syntactic complexity of the *wh*-phrase modulates the amplitude of the P600 component at the verb, suggesting retrieval facilitation. I argued that additional features may have contributed to increase the initial level of activation in memory of the *wh*-phrase, increasing the probability of retrieval success. Contrary to previous studies, I failed to find a left anterior sustained negativity (SAN) over the course of the dependency, a component that has been argued to reflect memory maintenance (Fiebach et al., 2002; Phillips et al., 2005). The lack of a sustained effect may indicate there is no need to assume antecedents have to be actively maintained in memory, providing strong evidence in favor of retrieval models. In addition, assuming a retrieval model provides a unified account of dependency processing, since maintenance requires that the antecedent be clearly identifiable, such as in *wh*-dependencies or relative clauses.

Chapter 3

Locality effects in local and unbounded dependencies

Abstract

There is accumulating evidence that increasing dependency length hinders processing in VO languages (locality effects) (Gibson, 2000; Grodner & Gibson, 2005) but leads to facilitation in OV languages (antilocality effects) (Konieczny, 2000; Vasishth & Lewis, 2006). However, evidence for locality comes mostly from unbounded dependencies (e.g. relative clauses or *wh*-questions), whereas evidence for antilocality comes from local dependencies (e.g. argument-verb dependencies). In this chapter, I hypothesized that the reason why local and unbounded dependencies behave differently lies in the predictability of the dependency head. Crucially, local dependencies must be resolved within the same clause, whereas unbounded dependencies can span across clauses. Expectation-based accounts of sentence comprehension (Levy, 2008) predict that increasing dependency length will yield an antilocality effect in local dependencies, but a locality effect in unbounded dependencies. Memory-based accounts (Gibson, 2000; Lewis & Vasishth, 2005), however, predict a locality effect in both local and unbounded dependencies. I ran an acceptability judgment task and two self-paced reading experiments in Spanish using local and unbounded dependencies where dependency length was manipulated. Reading time data showed a locality effect in both types of dependencies. Additionally, data from a cloze completion task suggested that the verb is highly predictable in both local and unbounded dependencies. Overall, results support memory-based accounts predicting processing difficulty as dependency length increases. The evidence presented in this chapter favors the view of locality as a general phenomenon in VO languages —regardless of dependency type.

1 Introduction

The linear distance between two elements in a dependency relation is a major determinant of processing difficulty. Accumulating evidence shows that increasing the amount of intervening material makes dependency completion more costly, as observed in self-paced reading tasks (Grodner & Gibson, 2005), eye-tracking experiments (Bartek et al., 2011), and electrophysiological experiments (Phillips et al., 2005). For example, the verb *sent* in (1) is read faster in a subject relative clause (1a) than in an object relative clause (1b). This is because in the latter, *the photographer* intervenes between *who* and the verb.

(1) a. The reporter **who sent** the photographer to the editor hoped for a story.

b. The reporter **who** the photographer **sent** to the editor hoped for a story.

(Grodner and Gibson, 2005, p. 266)

This so-called *locality effect* is captured by memory-based sentence processing models (Gibson, 1998, 2000; Lewis & Vasishth, 2005; Lewis et al., 2006; Vasishth & Lewis, 2006). The common assumption is that memory can make use of limited cognitive resources to compute all the relations necessary to build a syntactic representation. This limitation makes comprehenders tend to resolve dependencies at the first available site before bottom-up evidence is available. In (2a), for instance, the parser encounters the *wh*-pronoun *who* and attempts to complete de dependency at the verb *bring*, interpreting *who* as the object of the sentence. However, the next word (*us*) is the true object of the sentence, yielding a slowdown in reading times known as the *filled-gap effect* (Stowe, 1986) in comparison to the same word in (2b), where there is no dependency.

(2) a. My brother wanted to know **who** Ruth will bring us home **to** at Christmas.

b. My brother wanted to know if Ruth will bring us home to Mom at Chrtistmas.

Contrary to the predictions made by memory-based models, there is evidence that increasing dependency length can also facilitate language processing (Konieczny, 2000; Konieczny & Döring, 2003; Vasishth, 2003; Vasishth & Lewis, 2006), i.e. a speedup in reading times at the completion point when words are interposed between two co-dependents. Konieczny (2000) conducted a self-paced reading study in German manipulating the distance between the object and the verb. There were three possibilities: no intervening PP (3a), a three-word PP (3b) or a five-word PP (3c). Reading times at the verb *(an)gelegt* decreased, i.e. verbs were read faster, as intervening material increased. In other words, the verb in (3a) was read more slowly than in (3b) and (3c). This so-called *antilocality* effect has been explained in terms of reactivation (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006) or expectation-driven facilitation (Hale, 2001; Levy, 2008) as a result of strong predictions of upcoming syntactic heads.

(3) a. Er hat die **Rose hingelegt**, und...
he AUX-3SG the rose lay-PTCP and
"He laid down the rose and..."

b. Er hat die **Rose** auf den Tisch **gelegt**, und...
he AUX-3SG the rose on the table lay-PTCP and
"He laid the rose on the table and..."

c. Er hat die **Rose** auf den kleinen runden Tisch **gelegt**, und...
he AUX-3SG the rose on the small round table lay-PTCP and
"He laid the rose on the small round table and..."

(Konieczny, 2000, p. 632)

Locality effects have mainly been found in VO languages, such as English (Grodner & Gibson, 2005), Chinese (Hsiao & Gibson, 2003) and Russian (Levy et al., 2013); whereas antilocality effects have been reported in OV languages, such as German (Konieczny, 2000; Konieczny & Döring, 2003) and Hindi (Vasishth, 2003; Lewis & Vasishth, 2005). This distributional pattern of (anti)locality effects may respond to the parsing strategies speakers of different word order languages resort to. In OV languages, the position of the verb at the end of the clause makes the position of the verb more predictable. Moreover, preverbal arguments give away information about the verb, both lexical and syntactic, allowing the parser to update its predictions and facilitation processing, as predicted by expectation-based accounts (Levy, 2008).

Even though preliminary data suggests that basic word order in a language seems to be correlated with whether a language shows locality or antilocality effects, there are exceptions. Jaeger, Fedorenko, Hofmeister, and Gibson (2008) manipulated the distance between a subject and the main verb in English by attaching a relative clause and adding adjuncts. Unexpectedly, longer dependencies elicited faster reading times. Similarly, Levy and Keller (2013) conducted an eye-tracking experiment in German where they manipulated dependency length in a verb-final embedded clause, finding an antilocality effect. In a second experiment, they added more embeddings and they found that increased memory demands resulted in a locality effect, providing evidence that complex sentences exceeding memory resources override distance-based facilitation. However, Vasishth, Mertzen, Jäger, and Gelman (2018) failed to replicate this effect in seven different self-paced reading and eye-tracking experiments using the exact same materials as Levy and Keller (2013). This mixed evidence suggests that we need to integrate both memory-based and expectation-based models in order to provide a full account of sentence comprehension. Also, these results call into question the claim that (anti)locality effects are dependent on general word order preferences within a language (VO vs. OV). Rather, difficulty or facilitation when completing a dependency may depend on the complexity of the syntactic structure and whether the head of the dependency can be predicted more or less easily.

Paper	Language	Dependency type
Locality effect		
Gibson (2000)	English	Unbounded
Gibson and Warren (2004)	Russian	Unbounded
Grodner and Gibson (2005), Exp. 1	English	Unbounded
Vasishth and Drenhaus (2011)	German	Unbounded
Bartek, Lewis, Vasishth, and Smith (2011)	English	Unbounded & local
Levy, Fedorenko, and Gibson (2013)	Russian	Unbounded
Antilocality effect		
Konieczny (2000)	German	Local
Vasishth and Lewis (2006)	Hindi	Local
Nakatani and Gibson (2008)	Japanese	Local
Nakatani and Gibson (2010)	Japanese	Local

TABLE 3.1: Summary of the studies that have found either locality or antilocality effects, including the language and the dependency type (local or unbounded) they tested.

Crucially, there is a bias in the type of dependencies involved in studies testing (anti)locality effects. Current available evidence of locality effects is almost exclusively reduced to *unbounded* dependencies, whereas evidence of antilocality effects comes from studies testing *local* dependencies (see TABLE 3.1 for a summary of some of the studies that find either locality or antilocality effects). Unbounded dependencies (also known as *long-distance* dependencies or *filler-gap* dependencies), e.g. *wh*-questions or relative clauses, may be resolved within the same clause, as in (4a), or they may span over one (4b), two (4c) or more clauses. Local dependencies, e.g. argument–verb dependencies, however, must be completed within the same clause obligatorily and cannot cross clause boundaries. This difference may have implications for online dependency formation because the position of the verb is more predictable in local dependencies than in unbounded dependencies.

(4) a. **Who** did John **see**?

b. **Who** did John say [that Mary **saw**]?

c. **Who** did John mention [that Tim said [that Mary **saw**]]?

(5) The **neighbor** who lives next door **saw** the child.

In this chapter, I investigate whether the distribution of locality and antilocality effects can be best explained in terms of dependency type. This question has been addressed in previous studies (Grodner & Gibson, 2005; Bartek et al., 2011) but results are not consistent, and materials may have included additional complexity that may have affected the results. This will be discussed more in depth in the following sections. In what follows, I will give a more detailed

account of memory-based models and expectation-based models and how they account for locality and antilocality effects. I will then explain the implications for local and unbounded dependencies and present the experimental series I carried out.

1.1 The source of *locality* and *antilocality* effects

Memory-based accounts

Memory-based models of sentence processing assuming memory resources are limited predict locality effects. *Dependency Locality Theory* (Gibson, 1998, 2000) (henceforth, DLT) is a model of sentence comprehension that assumes there are two components that make use of memory resources when processing dependencies: a *storage* cost associated with keeping information about the structure built so far in memory and keeping track of incomplete dependencies; and an *integration* cost which results from incorporating new input into the syntactic structure built incrementally. DLT assumes that both storage and integration processes use resources from the same pool (Just & Carpenter, 1992). Therefore, an increased storage cost will also affect integration because fewer resources are available.

It has long been observed that double center embeddings (6b) are more difficult to process than single center embeddings (6a) probably because keeping track of a higher number of incomplete dependencies incurs in higher memory demands (Chomsky & Miller, 1963; Bever, 1970). In DLT, the storage cost depends on how many syntactic heads are necessary to complete a dependency. The assumption is that every predicted head uses one memory unit (MU). For example, when seeing the word *the* at the beginning of a sentence, the storage cost is minimally 2 MUs because at least a noun and a verb are needed to form a grammatical sentence in English (e.g. *The reporter slept*).

(6) a. The reporter [who the senator **attacked**] disliked the author.

b. The reporter [who the senator [that John met] **attacked**] disliked the author.

c. The reporter [who the senator [that I met] **attacked**] disliked the author.

(Gibson, 2000, p. 105, 107)

However, the number of predicted dependencies alone cannot account for the increase of difficulty in (6b) with respect to (6c), as revealed by lower ratings in acceptability judgment tasks and slower reading times in self-paced reading tasks (Warren & Gibson, 2002), because the same number of incomplete dependencies are involved in both cases. DLT gives a more prominent role to the integration cost in order to account for this increased difficulty. In this model, it is assumed that elements introducing new discourse referents (e.g. NPs, proper nouns and verbs) require additional processing resources than elements that are already part of the

discourse (e.g. second and first person pronouns). Hence, there is a relationship between the integration cost and the number of new discourse referents intervening. This processing cost is measured in terms of energy units (EU), each new discourse referent consuming 1 EU. Therefore, the integration cost at the verb *attacked* in (6b) is 2 EUs when integrating *the senator* as the subject (new discourse referents: *John* and *met*) and 4 EUs in order to coindex *who* with the empty category in object position (new discourse referents: *the senator*, *John*, *met* and *attacked*). In (6b), however, the integration cost is 1 EU less because *I* does not introduce a new referent in the discourse.

DLT has the advantage of providing a simple but precise metric to measure processing difficulty in sentence comprehension, but antilocality effects constitute a challenge as they remain unexplained. Other memory-based models, however, predict both locality and antilocality effects. One such model is the activation-based model of sentence comprehension (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006; Lewis et al., 2006) (henceforth, LV05) based on the Adaptive Character of Thought–Rational cognitive architecture (J. R. Anderson, 2005; J. R. Anderson & Lebiere, 1998; J. R. Anderson et al., 2004). Unlike DLT, LV05 is based on general cognitive mechanisms that are independent of language. The basic assumptions are that there is a highly limited focus of attention and, consequently, information has to be constantly encoded and retrieved from memory through a content-addressable cue-based retrieval mechanism. In this model, sentence comprehension is shaped by fluctuating activation in memory and similarity-based interference. For example, in (7), *toy* is encoded in memory as a bundle of features. At that point, the parser expects *toy* to be the subject of the sentence, so they generate a representation of the expectation of the predicate it is part of as a bundle of features and maintain it in memory —but out of the focus of attention. The activation of this representation decays from memory with time in a nonlinear manner. At *arrived*, a set of retrieval cues trigger a search for all possible elements in memory that match them. These retrieval cues are matched in parallel against the cues of all the possible candidates in memory.

(7) Melissa knew that the **toy** from her uncle in Bogotá **arrived** yesterday.

(Lewis et al., 2006, p. 448)

At this point, we can draw two predictions from this model. First, the larger the distance between the target and the retrieval point, the slower the retrieval will be due to activation decay. Second, when there are other items also matching the retrieval cues —called *distractors*—, they interfere at retrieval and, as a result, activation for all items is reduced. Both proactive (Van Dyke & McElree, 2006, 2011) and retroactive (Gordon et al., 2001; Van Dyke & Lewis, 2003; Vasishth & Lewis, 2006; Van Dyke & McElree, 2011) inhibitory similarity-based interference have been attested in experimental work, as well as facilitatory interference leading to grammatical illusions commonly known as *attraction* effects (Bock & Miller, 1991; Wagers, Lau, & Phillips, 2009), i.e. sentences with agreement mismatches where an intervening item is retrieved instead of the correct target (e.g. **The key to the cabinets are...*). Inhibitory similarity-based interference is connected to dependency length in the sense that

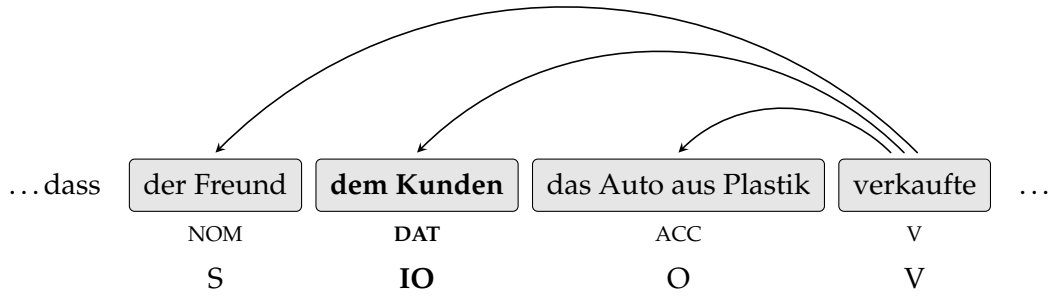
there is a higher probability of having distractors when the distance between the target and the retrieval site is larger. Crucially, both activation decay and similarity-based interference play an important role in explaining locality effects in language comprehension.

Activation decay, however, can be attenuated by subsequent retrievals of the item from memory. In other words, every time there is a retrieval event, the activation of the items matching the retrieval cues is boosted. This increase in activation prevents items from decaying from memory and, consequently, speeds up their retrieval at the dependency completion point, leading to an antilocality effect. Vasishth and Lewis (2006) illustrate this with an O-Adv-V sequence in Hindi, an OV language, where the activation level of a predicted VP fluctuates as the sentence unfolds. At O, the parser makes a prediction of a VP, and its activation peaks and then decays as a function of time. At the adverb, there is a retrieval event that boosts its activation again, followed by an activation decay. Finally, the retrieval event at the verb boosts the activation of the predicted VP a second time. As a result of multiple activation boosts, retrieval speed is enhanced. However, it must be noted that activation boosts are also affected by similarity-based interference; that is, when more than one item matches some of the retrieval cues, activation is shared by them.

Expectation-based accounts

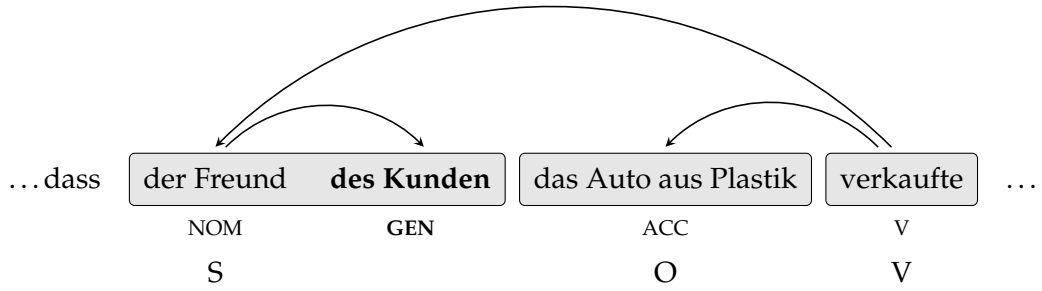
Expectation-based models of sentence comprehension do not consider memory limitations as the key constraining factor, but linguistic experience. Sentence comprehension is incremental (Altmann & Kamide, 1999; Kaiser & Trueswell, 2004) and comprehenders use previous input to rapidly generate expectations on the fly with regards to upcoming words, as revealed by numerous ERP studies showing that the N400 component is sensitive to the predictability of a word in a given context (Kutas & Hillyard, 1980, 1984; Lau et al., 2008; Lau, Holcomb, & Kuperberg, 2013). In the case of dependencies, constituents intervening between a dependent and its head help narrow down the possible continuations and reduce the uncertainty with respect to upcoming words. In early work finding antilocality effects, this was formulated as the *anticipation hypothesis* (Konieczny, 2000; Konieczny & Döring, 2003). As an example, let us focus on the sentences in (8), which are verb-final embedded clauses in German. In (8a), there are three preverbal arguments: S (in nominative case), IO (in dative) and O (in accusative). In (8b), however, there are only two arguments: S (in nominative case) and O (in accusative). Changing one letter in the determiner of the second constituent (*den/des Kunden*) makes a crucial difference: in (8b) it is a genitive NP modifying the preceding NP, and thus it is not part of the core verbal argument structure. DLT would predict (8a) to be harder to process because more integrations take place at the verb —three— in comparison to (8b) —two. However, Konieczny and Döring (2003) found faster reading times at the verb in (8a). The explanation is that each new preverbal argument adds information about the verbal argument structure, constraining the possible continuations, sharpening predictions and facilitating processing. The dative in (8a) allows the parser to expect a ditransitive verb, whereas in (8b) both a transitive and a ditransitive verb are possible.

(8) a. *Three preverbal arguments: S-IO-O-V*



"...that the friend sold the car made from plastic to the client..."

b. *Two preverbal arguments: S-O-V*



"...that the friend of the client sold the car made from plastic..."

(Konieczny and Döring, 2003, p. 3)

In order to to measure predictability, Levy (2008) proposed a metric which is equivalent to surprisal theory (Hale, 2001). In this model, after listening or reading a partial input, the parser assigns a probability to all the syntactic constructions that are plausible continuations. These probabilities are updated as comprehenders are exposed to new words. Sentence comprehension, then, involves allocating resources along the possible continuations of a sentence depending on how probable they are according to the comprehender's language experience. A sudden change in this probability distribution due to an exposure to a less expected continuation will thus trigger a relocation of resources, resulting in processing difficulty. However, if incoming words are consistent with the syntactic construction which had been assigned a higher probability, processing is facilitated. The difficulty associated with this resource reallocation update can be quantified using the *surprisal value* of a word (Hale, 2001), i.e. the negative log-probability of a word appearing after a given context. Expectations are based on the comprehender's language experience and there is evidence that they can be shaped easily. In two self-paced reading studies, Fine, Jaeger, Farmer, and Qian (2013) found that processing difficulty of ambiguous sentences, i.e. garden-path sentences such as *The experienced soldiers warned about the dangers conducted the midnight raid*, decreased when participants were exposed to a higher proportion of this type of constructions than they would normally be. Results support the idea that expectations can be rapidly adapted depending on the input received.

Regarding dependency resolution, Husain, Vasishth, and Srinivasan (2014) provided evidence that strong expectations facilitate language processing. In a self-paced reading experiment, they tested the effect of increasing dependency length in light verb constructions in Hindi (e.g. *khayaal rakhnaa*, "to take care of"), a complex predicate where the identity of the verb is highly expected after seeing the noun, and simple predicates (e.g. *gitaar rakhnaa*, "to put down or keep a guitar") made up of a noun and a verb whose meaning is compositional, where the identity of the verb is less predictable. Whereas longer dependencies produced a locality effect in simple predicates, an antilocality effect arose in light verb constructions, suggesting that strong expectations can override locality effects. However, results should be interpreted with care, as Safavi, Husain, and Vasishth (2016) failed to replicate this effect using the same type of constructions in Persian.

1.2 Local and unbounded dependencies

In psycholinguistics, much research has focused on investigating the mechanisms by which unbounded dependencies are resolved (Aoshima, Phillips, & Weinberg, 2004). Some proposals, such as the AFS (Frazier, 1987; Clifton & Frazier, 1989), which proposes that *wh*-words are actively maintained in a memory buffer until they can be integrated into the syntactic representation, do not make any claims about local dependencies. Electrophysiological work studying the neural correlates of active dependency formation also only focus on *wh*-dependencies (Kaan et al., 2000; Felser, Clahsen, & Münte, 2003; Phillips et al., 2005; Ueno & Garnsey, 2008; Gouvea, Phillips, Kazanina, & Poeppel, 2010). A crucial question is whether unbounded and local dependencies make use of the same memory mechanisms in order to integrate a dependent with its head. Both DLT and LV05 make no distinction between unbounded and local dependencies and, therefore, processing difficulty should affect both dependency types equally. However, expectations regarding incoming words may differ and this may impact on online dependency completion.

Take for example the Spanish sentences in (9). In (9a), there is a PP (*a la abogada*) that marks the beginning of a local dependency and comprehenders expect a verb to appear within the same clause obligatorily. In (9b), however, the initial PP includes a *wh*-operator that signals the opening of an unbounded dependency. In this case, different continuations are possible: the verb that completes the dependency can either appear in the same clause (9b)i or in a different clause (9b)ii.

(9) a. *Local dependency*
A la abogada...
To the lawyer

i. ... le **han** **entregado** la documentación.
CL-3SG AUX-3PL hand.over-PTCP the document
"They have handed over the documents to the lawyer."

b. *Unbounded dependency*

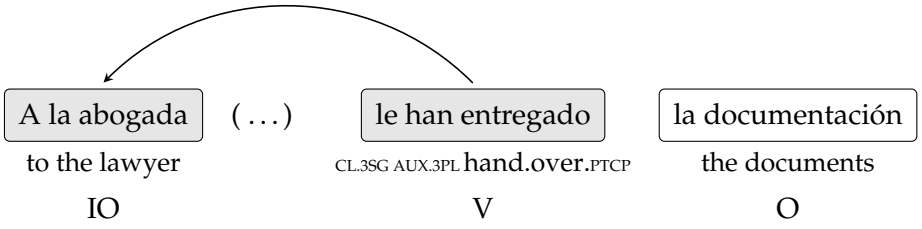
¿ **A qué abogada**...
To which lawyer

i. ... le **han** **entregado** la documentación?
CL-3SG AUX-3PL hand.over-PTCP the document
"To which lawyer have they handed the documents?"

ii. ... ha dicho el juez [que le **han** **entregado** la documentación]?
AUX-3SG say-PTCP the judge that CL-3SG AUX-3PL hand.over-PTCP the document
"To which lawyer has the judge said [that they have handed over the documents]?"

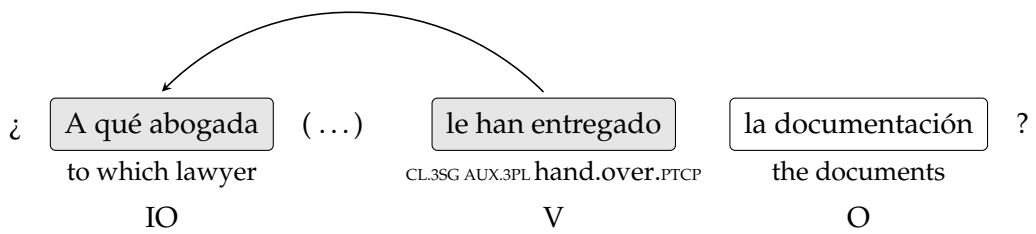
This difference regarding the predictability of the position of the verb constitutes a crucial difference with respect to the predictions made by expectation-based and memory-based models for sentences like (10). On the one hand, according to expectation-based accounts, a strong expectation of a verb appearing within the same clause in local dependencies (10a) should facilitate language processing when intervening words appear, yielding an antilocality effect. A higher uncertainty in unbounded dependencies (10b), however, should increase processing difficulty at the verb when dependency length is increased, resulting in a locality effect. On the other hand, according to memory-based accounts, increasing dependency length will hinder sentence comprehension regardless of dependency type, either due to new discourse referents being introduced (DLT) or due to activation decay and inhibitory similarity-based interference (LV05). Therefore, a locality effect is predicted to arise in both local and unbounded dependencies. In this chapter, I test the predictions made by expectation-based and memory-based accounts.

(10) a. *Local dependency*



"They have handed over the documents to the lawyer."

b. *Unbounded dependency*



"To which lawyer have they handed over the documents?"

To the best of my knowledge, only two previous studies have included both local and unbounded dependencies in their design in order to investigate locality effects. Grodner and Gibson (2005) conducted a self-paced reading experiment where dependency length was manipulated in sentences including either local dependencies between a subject and a verb in a main clause (11a) or unbounded dependencies between the head of relative clause and an embedded verb (11b).[1]

(11) a. **The nurse** (. . .) **supervised** the administrator while. . .

b. **The administrator who** the nurse (. . .) **supervised** scolded the medic while. . .

(Grodner and Gibson, 2005, p. 273)

Results show increased reading times at the embedded verb when dependency length was increased, i.e. a locality effect, only in unbounded dependencies. In local dependencies, however, there was no difference between the short and the long condition. Bartek et al. (2011) ran a self-paced reading experiment with the exact same materials and replicated those results. In a second experiment using an eye tracker to measure reading difficulty, Bartek and colleagues reported a locality effect in both local and unbounded dependencies, revealed as an increase in total fixation times at the verb in long dependencies in comparison to short dependencies. In two subsequent experiments —a self-paced reading task and an eye-tracking experiment— using a modified version of the materials maintaining its structural configuration, Bartek et al. reported a locality effect in both local and unbounded dependencies at the critical verb. Overall, results from Bartek et al. (2011) support the hypothesis that locality affects both local and unbounded dependencies.

It should be noted that the experimental design used by Grodner and Gibson (2005) and Bartek et al. (2011) is problematic. In their materials, local dependencies were always resolved in a matrix clause, whereas unbounded dependencies were resolved in an embedded clause, where additional complexity was introduced. For instance, in (11a), only the subject *the nurse* has to be integrated with the verb *supervised*, whereas in (11b), both *the administrator* and *the nurse* have to be integrated with the verb. For this reason, results must be interpreted with care. Comparing local and unbounded dependencies in the same type of clause is more appropriate as it would allow one to assess the effect of increasing dependency length while controlling for any other factor that may affect the results.

I carried out four experiments using sentences such as (10). I conducted an acceptability judgment task (Experiment 2) and two self-paced readings studies (Experiments 3 & 4) in order to investigate whether increasing dependency length facilitated or hindered dependency completion at the verb in local and unbounded dependencies. Additionally, I ran a cloze

[1]Grodner and Gibson (2005) refer to local dependencies as *matrix* in their design and to unbounded dependencies as *embedded*. I stick to the terms *local* and *unbounded* for homogeneity and because they reflect more accurately the type of dependencies the authors used: a subject-verb dependency (*local*), and a relative clause involving a *wh*-operator (*unbounded*).

completion task (Experiment 5) in order to estimate the probability of completing a dependency locally after seeing the beginning of either a local or an unbounded dependency.

2 Experiment 2

2.1 Overview

The goal of the current experiment was to investigate the effect of increasing dependency length on the acceptability of both local and unbounded dependencies. I presented Spanish native speakers with ditransitive sentences with a fronted IO where I manipulated the distance between IO and V by either attaching a postnominal relative clause to IO or not. In an acceptability judgment task, participants were asked to rate these sentences using a 7-point Likert scale. The linking hypothesis was that increased processing difficulty should yield lower acceptability ratings.

Let me briefly summarize the predictions drawn from the hypotheses I put forth in the previous section. Local and unbounded dependencies differ with regards to when the dependency must be completed: local dependencies must be resolved within the same clause, whereas unbounded dependencies can span over more than one clause. In principle, the position of V is more predictable in local dependencies than in unbounded dependencies. According to expectation-based accounts of sentence comprehension, if the position of V is highly predictable in local dependencies but not in unbounded dependencies, there should be an antilocality effect in local dependencies, i.e. long dependencies should receive higher ratings than short dependencies, but a locality effect in unbounded dependencies due to a greater degree of uncertainty. According to memory-based accounts, dependency length should affect both local and unbounded dependencies equally, yielding a locality effect, manifested here as an overall lower rating for long dependencies than for short dependencies.

2.2 Participants

47 native speakers of Spanish (35 women, mean age = 38.3, SD = 12.9, range = 21–64) took part in the experiment through the online platform Ibex Farm. Participants were recruited using social media (Twitter and Facebook). They were given informed consent and received no compensation for their participation.

In order to make sure participants were native speakers of Spanish, they were asked whether they had been born and raised in Spain and whether they always spoke Spanish at home. Participants who failed to answer positively to these two questions were removed from the analysis, affecting 4.3% of the participants (2 out of 47). Overall, data from 45 participants was entered into the analysis.

2.3 Materials

24 experimental items and 72 filler items were used in the experiment. The experimental items were ditransitive sentences in Spanish with a fronted IO (IO-V-O). Four experimental conditions were created by manipulating two factors (2 x 2): *dependency length*, i.e. the distance between IO and V, by attaching a prenominal relative clause to either IO (long) or O (short), and *dependency type*, by either including a *wh*-operator in the IO (unbounded dependency) or not (local dependency). IO was always animate, whereas O was always inanimate, and all relative clauses were made up of five words. An example of the materials is provided in TABLE 3.2. All experimental items are listed in Appendix C.

Filler items were carefully designed in order to mask the true purpose of the study. 48 filler items were ungrammatical in order to avoid any equalizing bias (Poulton, 1979), i.e. participants trying to balance the number of times they give high and low ratings, even after being exposed to fully grammatical sentences. 24 filler items included a relative clause, so that half of the items in the experiment contained one.

2.4 Procedure

Participants were asked to read some sentences and rate them according to how acceptable they seemed. Sentences were presented on a white background. Acceptability was explained in terms of plausibility and naturalness. Participants were told to follow their intuitions and not to think about prescriptive grammatical correctness. A 7-point Likert scale was provided to rate the sentences (1 = totally unacceptable, 7 = totally acceptable). Before starting the experiment, six practice sentences with gradient acceptability were displayed, although participants were not aware these were not part of the study. The experiment could be completed using a computer or a mobile device. Including the initial questionnaire, the study did not last longer than 20 minutes.

Cond	Example
Short–Local	**A la abogada** le **han entregado** la documentación [que buscaba el secretario judicial]… to the lawyer CL-3SG AUX-3PL hand-PTCP the document that look-for the secretary judicial "They handed the documents that the clerk of court was looking for to the lawyer."
Long–Local	**A la abogada** [que buscaba el secretario judicial] le **han entregado** la documentación… to the lawyer that look-for the secretary judicial CL-3SG AUX-3PL hand-PTCP the document "They handed the documents to the lawyer that the clerk of court was looking for."
Short–Unbounded	¿**A qué abogada** le **han entregado** la documentación [que buscaba el secretario judicial]…? to the lawyer CL-3SG AUX-3PL hand-PTCP the document that look-for the secretary judicial "To what lawyer did they hand the documents that the clerk of court was looking for?"
Long–Unbounded	¿**A qué abogada** [que buscaba el secretario judicial] le **han entregado** la documentación…? to the lawyer that look-for the secretary judicial CL-3SG AUX-3PL hand-PTCP the document "To what lawyer that the clerk of court was looking for did they hand the documents?"

TABLE 3.2: Sample set of the sentences used in Experiment 2.

2.5 Analysis

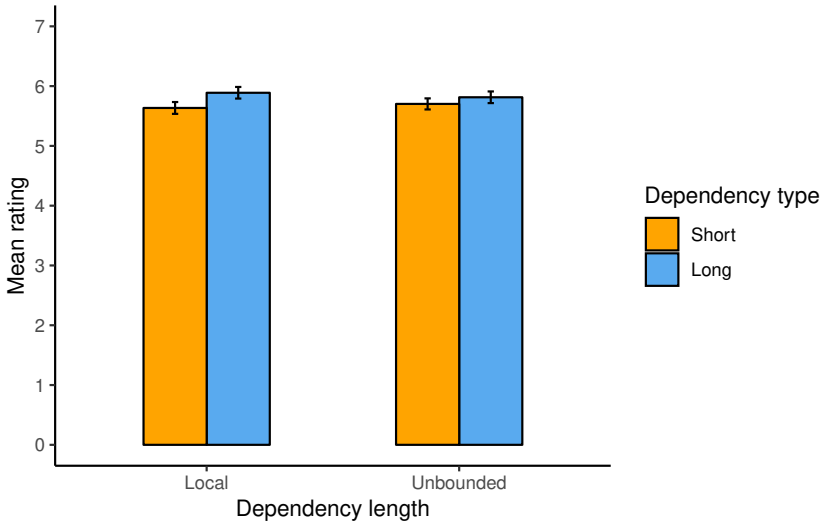
Raw ratings were transformed into *z*-scores taking into account each participant's mean and standard deviation for two reasons. First, ratings from Likert scales are ordinal and cannot be entered into a regression model, whereas *z*-score ratings are continuously distributed. Second, *z*-score transformation standardizes ratings and gets rid of scale biases as a result of participants using only some points of the scale.

Linear mixed models were run on *z*-score ratings including *dependency length*, *dependency type* and their interaction as fixed factors using the *lme4* package (Bates, Mächler, et al., 2015) in the R (version 3.5.3) programming environment (R Core Team, 2017). Sum contrasts were defined for the fixed factors: *long* was coded as 1 and *short* as -1; and *local* as 1 and *unbounded* as -1. The most parsimonious model was fitted, including varying intercepts by subject and by item, and complexity was added to the random-effects structure only when supported by the data in order to avoid model overfitting (Bates, Kliegl, et al., 2015). After carrying out model comparisons by means of likelihood ratio tests, the following model was fitted: *z.score.rating* $\sim$ *length*dependency + (1 | subject) + (1+length | item)*. All *p*-values were calculated using the Satterthwaite's model of approximation using the *lmerTest* package (Kuznetsova et al., 2017).

2.6 Results

Results are summarized in FIGURE 3.1. Mean raw ratings are reported in FIGURE 3.1a (top) and mean *z*-score ratings in FIGURE 3.1b (bottom). All graphs were generated using the *ggplot2* package (Wickham, 2016). Note that I have also provided raw ratings because they can be interpreted more intuitively, but no analysis was run on them. The summary of the coefficients of the model is provided in TABLE 3.3. Results from the linear mixed model show no significant main effect of *dependency length* or *dependency type* and no interaction. Overall, all conditions received similar ratings, which were well above the average.

(A) Mean raw ratings



(B) Mean z-score ratings

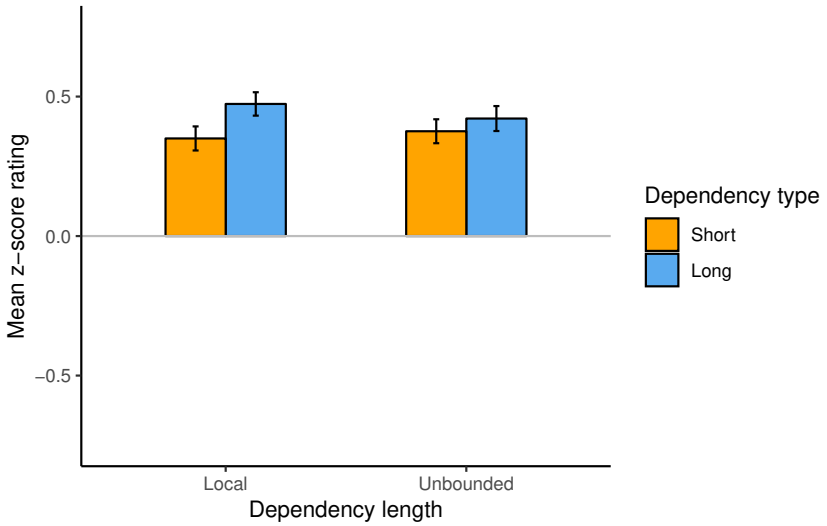


FIGURE 3.1: Mean acceptability ratings obtained in Experiment 2. Error bars show standard error.

	Estimate	SE	t-value	p-value
(Intercept)	0.41	0.05	8.85	<0.001
Length	0.04	0.03	1.34	0.19
Type	0.01	0.02	0.45	0.65
Length x Type	0.02	0.02	0.94	0.34

TABLE 3.3: Summary of the coefficients from the linear mixed model in Experiment 2.

2.7 Discussion

The present experiment investigated (anti)locality effects in both local and unbounded dependencies by gathering offline acceptability judgments from native speakers of Spanish. If predictability of V is higher in local dependencies than in unbounded dependencies, expectation-based accounts predicted a locality effect in unbounded dependencies and an antilocality effect in local dependencies; memory-based accounts, however, predicted a locality effect in both local and unbounded dependencies. Contrary to these predictions, results do not provide evidence in favor of either hypothesis, since there was no difference in acceptability between short and long dependencies or between local and unbounded dependencies. In other words, all conditions receive relatively high ratings.

It should be noted that null results should be interpreted with care in frequentist analyses because the null hypothesis may not be rejected due to high Type II error in underpowered studies.[2] However, it can be the case that no difference was found because acceptability judgments are not sensitive enough to capture effects of dependency length in grammatical sentences. The ditransitive sentences used in Experiment 2 are relatively simple sentences in Spanish which can be interpreted by native speakers with virtually no problem. Even if dependency length is affecting processing facilitation/difficulty, offline tasks may not be the best tool to capture these subtleties. In Experiment 3, I tested the same materials using an online task (self-paced reading).

3 Experiment 3

3.1 Overview

In this experiment, I investigated the impact of increasing dependency length on both local and unbounded dependencies during online sentence processing. In Experiment 2, no differences were found between conditions probably because offline acceptability judgments cannot possibly capture the increased difficulty incurred by long dependencies in a somewhat simple ditransitive sentence. In this occasion, I tested the sentences from Experiment 2 using the self-paced reading technique, where participants had to read sentences word-by-word and reading times for each word were recorded.

The predictions were similar to those from Experiment 2. If V is more predictable in local dependencies than in unbounded dependencies, expectation-based accounts predict that, in local dependencies, increasing dependency length will yield an antilocality effect —faster reading times at V for longer dependencies— due to facilitation, but a locality effect —slower reading times at V for longer dependencies— in unbounded dependencies. This pattern

[2]Type II error (β) is the probability of failing to reject the null hypothesis (H_0) when it is false. The lower the statistical power (1-β), the higher Type II error is.

would be revealed by an interaction between dependency type and dependency length (see FIGURE 3.2, left). Memory-based accounts, on the other hand, predict a locality effect when dependencies are longer for both local and unbounded dependencies, yielding a main effect of dependency length but no interaction (see FIGURE 3.2, right).

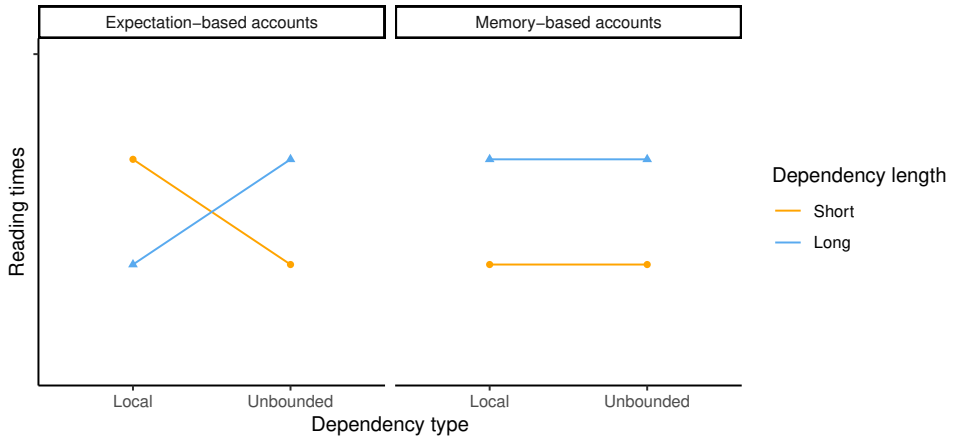


FIGURE 3.2: Informal predictions of expectation-based accounts (left) and memory-based accounts (right) for Experiments 3 & 4.

3.2 Participants

61 students from the University of the Basque Country (UPV/EHU) (46 women, mean age = 21.94, SD = 3.97, range = 18–38) took part in the experiment. All participants self-reported being native speakers of Spanish through a questionnaire filled out prior to the experiment. Participants gave informed consent and received €6 for their participation.

The criterion for exclusion was that participants whose accuracy in the comprehension questions fell below 70% were removed from the analysis. No participants were affected by this. Likewise, participants whose mean reading times were 2.5 SD above or below the general mean were excluded. Reading times from 1 participant were taken out of the analysis. Overall, data from 60 participants was analyzed.

3.3 Materials

24 experimental items and 72 filler items were used in the experiment. The experimental sentences were the same as in Experiment 2 (see TABLE 3.2 for an example of the materials, and Appendix C, Section 1 for a full list). The experimental items were ditransitive sentences in Spanish with a fronted IO (IO-V-O). Four experimental conditions were created by manipulating two factors (2 x 2 design): *dependency length*, i.e. the distance between IO and V, by attaching a prenominal relative clause to either IO (long) or O (short); and *dependency type*, by either including a *wh*-operator in the IO (unbounded dependency) or not (local dependency).

IO was always animate, whereas O was always inanimate; and all relative clauses were made up of five words.

Filler items were carefully designed in order to mask the true purpose of the study. 24 filler items included a relative clause, so that half of the items in the experiment contained one. In order to make sure participants were actually paying attention to the sentences, all the experimental items and 40 of the filler items were followed by a yes/no comprehension question. Overall, two thirds of the trials were followed by a comprehension question.

3.4 Procedure

The experiment was a word-by-word moving window self-paced reading task. Participants were comfortably seated in front of a computer in an experimental booth. At the beginning of each trial, one dash per word was displayed on the screen. Every time the participant pressed the space bar, a new word appeared and the previous one disappeared, so that they could only see one word at a time. Participants were instructed to read at a natural pace and to understand the sentences they were reading. After two thirds of the trials, they had to answer a yes/no comprehension question by pressing one of two buttons. Feedback was provided after every answer. Each experimental list was divided into four blocks of 24 trials, so that there were three breaks. Including completing the initial questionnaire, the experiment took around 30 minutes.

3.5 Analysis

Three words made up the critical region in the current experiment: the clitic (Clitic), the auxiliary verb (Aux) and the main verb (V). (Anti)locality effects in online experiments are measured at the verb. The materials for this experiment included verbs in the perfect form in Spanish, made up by an auxiliary verb carrying the agreement and tense features, a main verb (past participle) carrying lexical information. In addition, a clitic agreeing with IO is adjacent to the verb. Since it is not clear where exactly retrieval takes place, I decided to include Clitic, Aux and V as critical regions.

Reading times below 200 ms. and above 4000 ms. were removed because they cannot possibly be reflecting language-related processes. Raw reading times were log-transformed in order to normalize the data. The position of the critical region was different depending on the experimental condition. In the *short* condition, it was on words 4–6, whereas on the *long* condition, it is in words 9–11. Comparing critical regions on different positions with the sentence is not ideal because reading times speed up as comprehenders read further into the sentence (Gernsbacher, 1990). In other words, positions towards the end of the sentence are generally read faster. In order to control for this confound, I fitted a linear mixed model on logarithmic reading times including *word position* within the sentence as fixed effect and random intercepts by subject using the *lme4* package (Bates, Mächler, et al., 2015) in the R

(version 3.5.3) programming environment (R Core Team, 2017). The residuals from this model were used as the dependent variable in the final analysis. A linear mixed model was fitted on residual logartihmic reading times with *dependency length*, *dependency type* and their interaction as fixed factors. Sum contrasts were defined for the fixed factors: *long* was coded as 1 and *short* as -1; and *local* as 1 and *unbounded* as -1. The most parsimonious model was fitted —including only varying intercepts by subject and by item— and complexity was added to the random-effects structure only when supported by the data in order to avoid model overfitting (Bates, Kliegl, et al., 2015). After carrying out model comparisons by means of likelihood ratio tests, the following model was fitted when analyzing logarithmic reading times at the clitic (Clitic): *logRTs* $\sim$ *length*dependency + (1+length | subject) + (1+length | item)*. For the analyses of logarithmic reading times at the verb (V) and the two postcritical regions (POST 1 and POST 2), a model including only varying intercepts by subjects and by items was fitted: *logRTs* $\sim$ *length*dependency + (1 | subject) + (1 | item)*. All p-values were calculated using the Satterthwaite's model of approximation using the *lmerTest* package (Kuznetsova et al., 2017).

3.6 Results

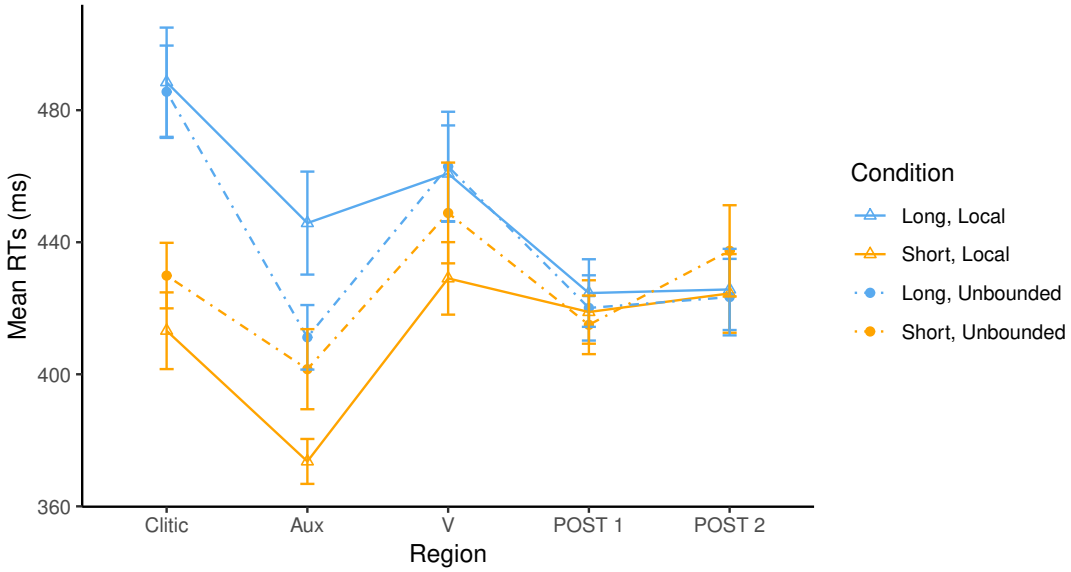
Results are summarized in FIGURE 3.3. Mean reading times in miliseconds are reported in FIGURE 3.3a (top) and mean residual logarithmic reading times in FIGURE 3.3b (bottom). Note that I have also included raw reading times because they can be interpreted more intuitively, but no analysis was run on them. Interaction plots using residual logarithmic reading times are provided in FIGURE 3.4: the critical regions in FIGURE 3.4a (top) and the postcritical regions in FIGURE 3.4b (bottom). All graphs were generated using the *ggplot2* package (Wickham, 2016). After carrying out model comparisons through likelihood ratio tests, The summary of the coefficients of the linear mixed models are provided in TABLE 3.4.

Visual inspection of the data shows faster reading times for short dependencies than for long dependencies in all critical and postcritical words. Results from linear mixed models confirmed a main effect of *dependency length* in all critical and postcritical regions. No effect of *depdency type* was found. An interaction was reported at Aux due to the fact that the difference between short and long dependencies was reduced in unbounded dependencies, although the reading time pattern was not reversed with respect to local dependencies, i.e. long dependencies were not read faster than short dependencies in unbounded dependencies, as predicted by expectation-based accounts. There was no interaction in any other critical or postcritical region. Overall, participants sped up more when dependency length was short than when it was long.

3.7 Discussion

The goal of Experiment 3 was to investigate (anti)locality effects in the comprehension of local and unbounded dependencies using an online method. The most important finding of this

(A) Mean raw reading times



(B) Mean residual logarithmic reading times

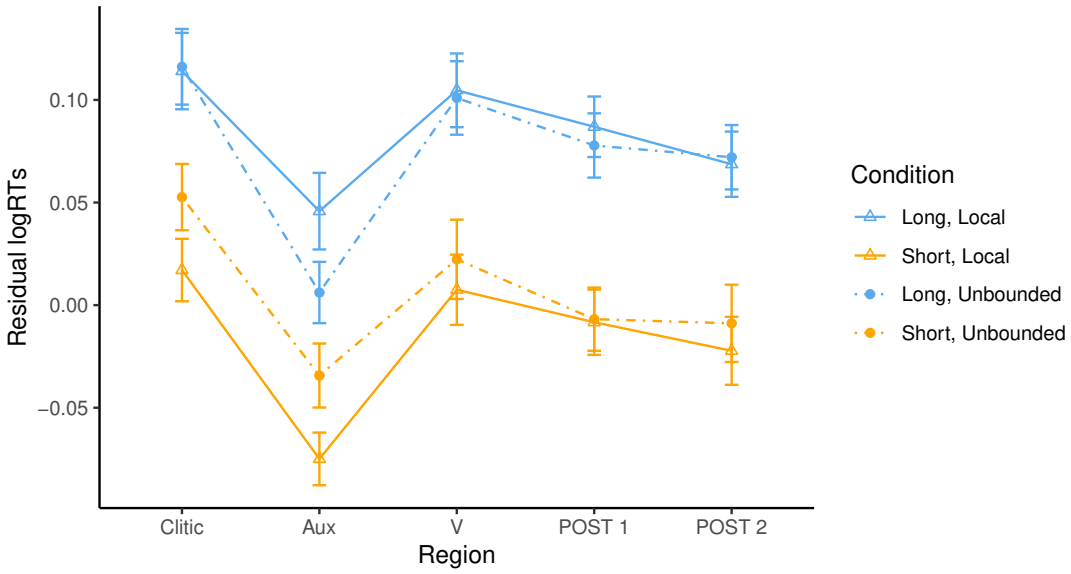
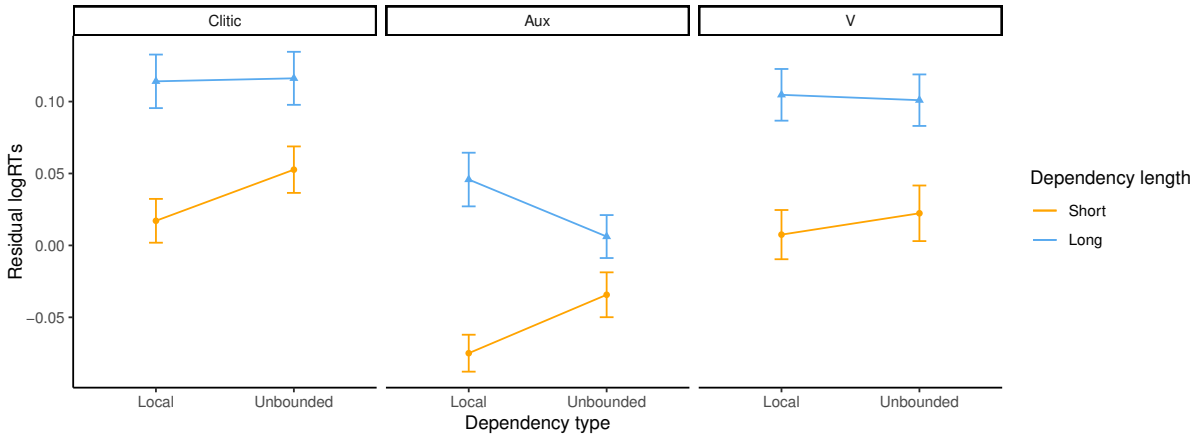


FIGURE 3.3: Raw reading times and residual logarithmic reading times (logRTs) at critical and postcritical regions in Experiment 3. Error bars show standard error.

(A) Interaction plot of residual logRTs times at critical regions.



(B) Interaction plots of residual logRTs at postcritical regions.

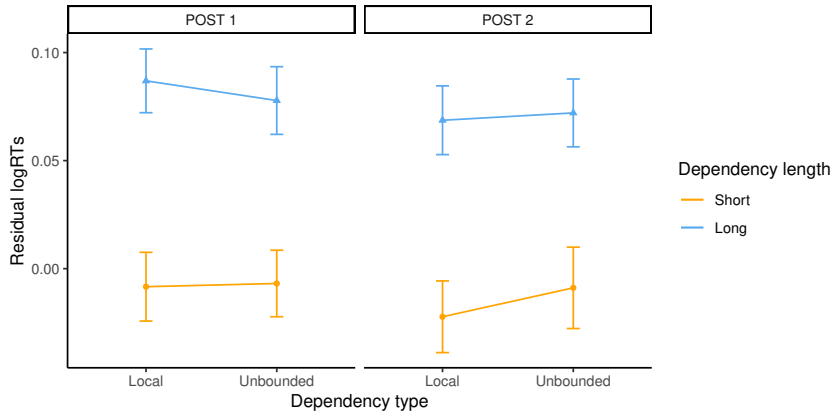


FIGURE 3.4: Interaction plots of the residual logarithmic reading times (logRTs) at critical (top) and postcritical (bottom) regions in Experiment 3. Error bars show standard error.

	Estimate	SE	*t*-value	*p*-value
Clitic				
(Intercept)	0.08	0.01	5.74	<0.001
Length	0.04	0.01	3.95	<0.001
Type	-0.01	0.01	-1.14	0.25
Length x Type	0.01	0.01	1.04	0.30
Auxiliary verb (Aux)				
(Intercept)	-0.01	0.01	-1.46	0.16
Length	0.04	0.01	5.18	<0.001
Type	<0.001	0.01	-0.04	0.97
Length x Type	0.02	0.01	2.58	0.01
Main verb (V)				
(Intercept)	0.06	0.01	4.14	<0.001
Length	0.04	0.01	5.02	<0.001
Type	0.003	0.01	-0.34	0.74
Length x Type	0.01	0.01	0.54	0.59
Postcritical 1 (POST 1)				
(Intercept)	0.04	0.02	2.54	0.02
Length	0.04	0.01	6.04	<0.001
Type	0.002	0.01	0.32	0.75
Length x Type	0.002	0.01	0.29	0.77
Postcritical 2 (POST 2)				
(Intercept)	0.03	0.02	1.80	0.08
Length	0.04	0.01	5.25	<0.001
Type	-0.004	0.01	-0.51	0.61
Length x Type	0.002	0.01	0.28	0.78

TABLE 3.4: Summary of the coefficients from the linear mixed models run on residual logartithmic reading times in Experiment 3.

study is that there was a clear locality effect at the verbal area (Clitic, Aux and V) and following words. That is, short dependencies elicited faster reading times than long dependencies. This difference does not seem to be modulated by dependency type, even though it was noticeably smaller at the auxiliary verb. Overall, results support the idea that dependencies in VO languages are subject to locality effects, most likely due to memory decay and interference derived from memory limitations.

In the introduction, I had outlined that current evidence of locality and antilocality effects was correlated with basic word order: locality was mainly found in VO languages and OV languages. The current experiment provides evidence that processing difficulty in a VO language like Spanish is not dependency dependent. However, conclusions must be drawn with care. One potential confound in the present experiment is that the critical region appeared on different positions within the sentence depending on the experimental condition: on words 4–6 in the *short* condition and on words 9–11 in the *long* condition. Reading times tend to be faster towards the end of the sentence, so this may have had an impact on the results obtained. Even though position within the sentence was taken into account in the residualization step of the analysis, a cleaner design where the critical region is stable across conditions would be

more appropriate. This was addressed in Experiment 4.

4 Experiment 4

4.1 Overview

In Experiment 3, I investigated the impact of dependency length on local and unbounded dependencies using the self-paced reading technique. In this experiment, the critical region appeared on different positions within the sentence across conditions. This is problematic because reading times tend to increase towards the end of the sentence. Even though this confound was corrected in the analysis through residualization, an experimental design where the critical region remains on the same position is more appropriate to study this phenomenon. In the current experiment, I present sentences where dependency length and dependency type was manipulated without affecting the position within the sentence of the critical region. The predictions for the current experiment are the same as for Experiment 3 (see FIGURE 3.2 for a summary).

4.2 Participants

65 students from the University of the Basque Country (UPV/EHU) (53 women, mean age = 19.46, SD = 2.05, range = 18–26) took part in the experiment. All participants self-reported being native speakers of Spanish through a questionnaire filled out prior to the experiment. Participants gave informed consent and received €8 for their participation.

Participants whose accuracy in the comprehension questions fell below 70% were excluded from the analysis. Likewise, participants whose mean reading times were 2.5 SD above or below the general mean were excluded. This procedure affected 1 participant. Overall, data from 64 participants was analyzed.

4.3 Materials

36 experimental items and 108 filler items were used in the experiment. The experimental sentences were embedded transitive sentences in Spanish with canonical word order (S-V-O) introduced by the verb *preguntar* ("to ask"). Four experimental conditions were created by manipulating two factors (2 x 2 design): *dependency length*, i.e. the distance between S and V, by attaching a postnominal relative clause to either S in the embedded clause (long) or S in the main clause (short); and *dependency type*, by either including a *wh*-operator in the embedded S (unbounded dependency) or not (local dependency). S was always animate, whereas O was always inanimate. All relative clauses were made up of five words. An example of the materials is provided in TABLE 3.5. All experimental items are listed in Appendix C, Section 3.

Cond	Example
Short–Local	El funcionario [que buscaba el secretario judicial] ha preguntado si **la abogada ha entregado**
	the civil-servant that look-for the secretary judicial AUX ask-PTCP if the lawyer AUX-3SG hand-PTCP
	la documentación…
	the document
	"The civil servant that the clerk of court was looking for has asked if the lawyer has handed the documents…"
Long–Local	El funcionario ha preguntado si **la abogada** [que buscaba el secretario judicial] **ha entregado**
	the civil-servant AUX ask-PTCP if the lawyer that look-for the secretary judicial AUX-3SG hand-PTCP
	la documentación…
	the document
	"The civil servant has asked if the lawyer that the clerk of court was looking for has handed the documents…"
Short–Unbound	El funcionario [que buscaba el secretario judicial] ha preguntado que **qué abogada ha entregado**
	the civil-servant that look-for the secretary judicial AUX ask-PTCP that what lawyer AUX-3SG hand-PTCP
	la documentación…
	the document
	"The civil servant that the clerk of court was looking for has asked which lawyer has handed the documents…"
Long–Unbound	El funcionario ha preguntado que **qué abogada** [que buscaba el secretario judicial] **ha entregado**
	the civil-servant AUX ask-PTCP that what lawyer that look-for the secretary judicial AUX-3SG hand-PTCP
	la documentación…
	the document
	"The civil servant has asked which lawyer that the clerk of court was looking for has handed the documents…"

TABLE 3.5: Sample set of the sentences used in Experiment 4.

Filler items were carefully designed in order to mask the true purpose of the study. 36 filler items contained a relative clause so that, overall, half of the trials included one. Likewise, 36 filler sentences contained the verb *preguntar* ("to ask"). In order to make sure participants were actually paying attention to the sentences, all experimental items were followed by a yes/no comprehension question, as well as 36 of the filler items. Overall, half of the trials included a comprehension question.

4.4 Procedure

This experiment was a word-by-word moving window self-paced reading task. Participants were comfortably seated in front of a computer in an experimental booth. At the beginning of each trial, one dash per word was displayed on the screen. Every time the participant pressed the space bar, a new word appeared and the previous one disappeared, so that they could only see one word at a time. Participants were instructed to read at a natural pace and to understand the sentences they were reading. After two thirds of the trials, they had to answer a yes/no comprehension question by pressing one of two buttons. Feedback was provided after every answer. Each experimental list was divided into four blocks of 24 trials, so that there were three breaks. Including completing the initial questionnaire, the experiment took around 45 minutes.

4.5 Analysis

Two words made up the critical region in the current experiment: the auxiliary verb (Aux) and the main verb (V). As in Experiment 3, the experimental sentences included verbs in the perfect

form in Spanish, made up by an auxiliary verb carrying the agreement and tense features and a main verb (past participle) carrying lexical information. Since it is not clear where exactly retrieval takes place, I decided to include both Aux and V as critical regions.

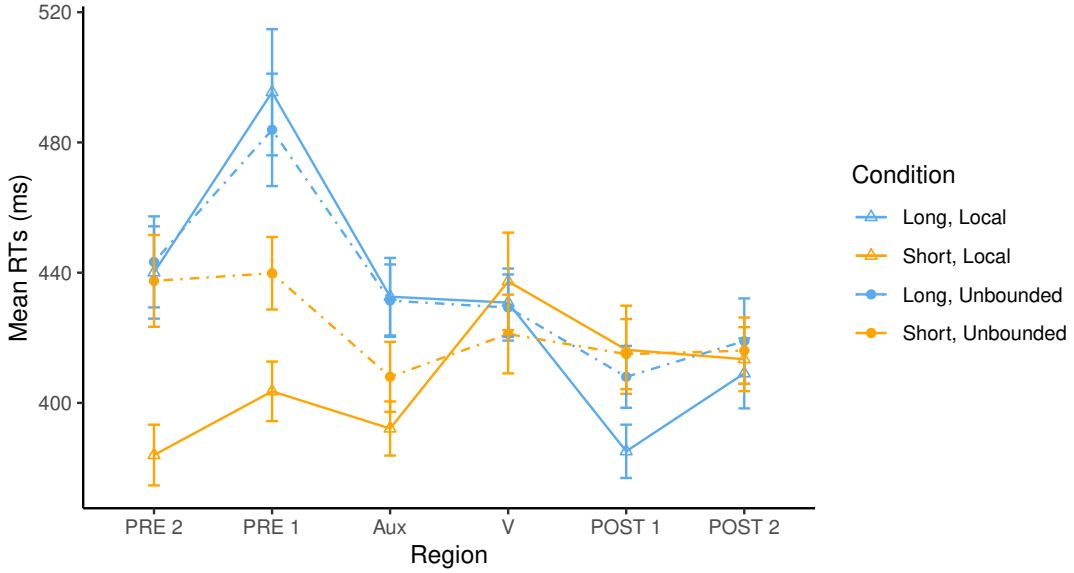
Reading times below 200 ms. and above 4000 ms. were removed because they cannot possibly be reflecting language related processes. Raw reading times were log-transformed in order to normalize the data. A linear mixed model was fitted on logartihmic reading times with dependency length and dependency type as fixed factors using the *lme4* package (Bates, Mächler, et al., 2015) in the R (version 3.5.3) programming environment (R Core Team, 2017). Sum contrasts were defined for the fixed factors: *long* was coded as 1 and *short* as -1; and *local* as 1 and *unbounded* as -1. The most parsimonious model was fitted, including varying random intercepts and slopes by subjects and by items, and complexity was added to the random-effects structure when supported by the data in order to avoid model overfitting (Bates, Kliegl, et al., 2015). After carrying model comparisons by means of likelihood ratio tests, the following model was fitted when analyzing logarithmic reading times at the auxiliary verb (Aux): *logRTs* $\sim$ *length + (1+length | subject) + (1 | item)*. For the analyses of logarithmic reading times at the verb (V) and the two postcritical regions (POST 1 and POST 2), a model including only varying intercepts by subjects and items was fitted: *logRTs* $\sim$ *length + (1 | subject) + (1 | item)*. All *p*-values were calculated using the Satterthwaite's model of approximation using the *lmerTest* package (Kuznetsova et al., 2017). A separate analysis was carried out for each critical and postcritical word.

4.6 Results

Results are summarized in FIGURE 3.5. Mean reading times in miliseconds are reported in FIGURE 3.5a (top) and mean residual logarithmic reading times in FIGURE 3.5b (bottom). Note that I have also included raw reading times because they can be interpreted more intuitively, but no analysis was run on them. Interaction plots using logarithmic reading times are provided in FIGURE 3.6: the critical regions in FIGURE 3.6a (top) and the postcritical regions in FIGURE 3.6b (bottom). All graphs were generated using the package *ggplot2* (Wickham, 2016). The summary of the coefficients of the models are provided in TABLE 3.6.

Visual inspection of the data shows faster reading times at the critical regions (Aux and V) in long dependencies than in short dependencies. This pattern appears to be reversed at the postcritical regions. Results from the linear mixed models revealed a main effect of *dependency length* at the auxiliary verb, but this effect did not reach significance at the main verb or the postcritical regions. There was no main effect of *dependency type* or interaction in any of the critical nor the postcritical regions. In sum, reading times were faster for shorter dependencies at the auxiliary, and no differences are reported in any of the other regions of interest. It should be noted that visual inspection of the data reveals a drop in reading times in the local–long condition at the postcritical region (POST 1).

(A) Mean raw reading times



(B) Mean residual logarithmic reading times

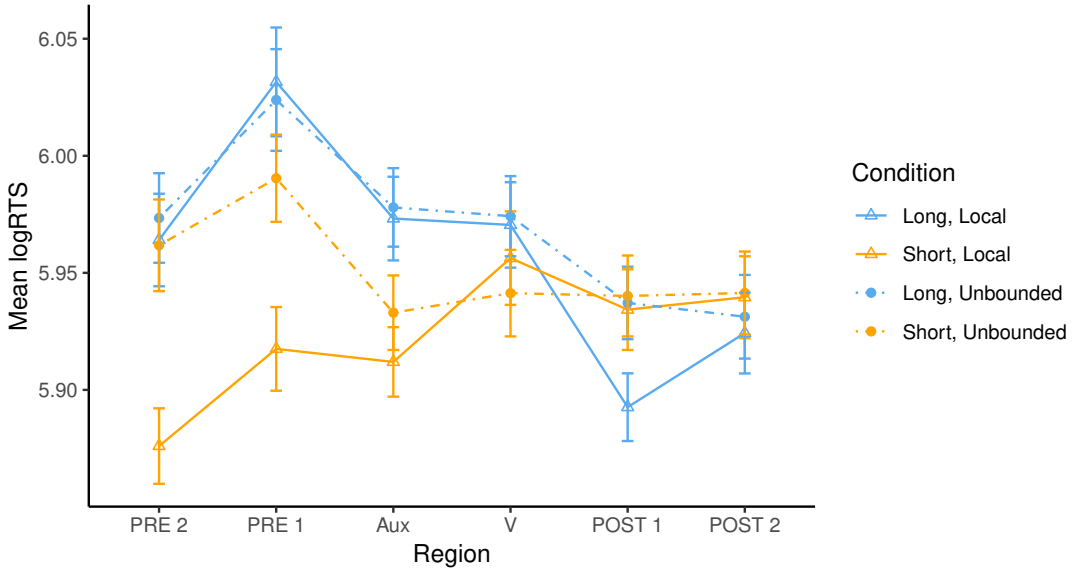
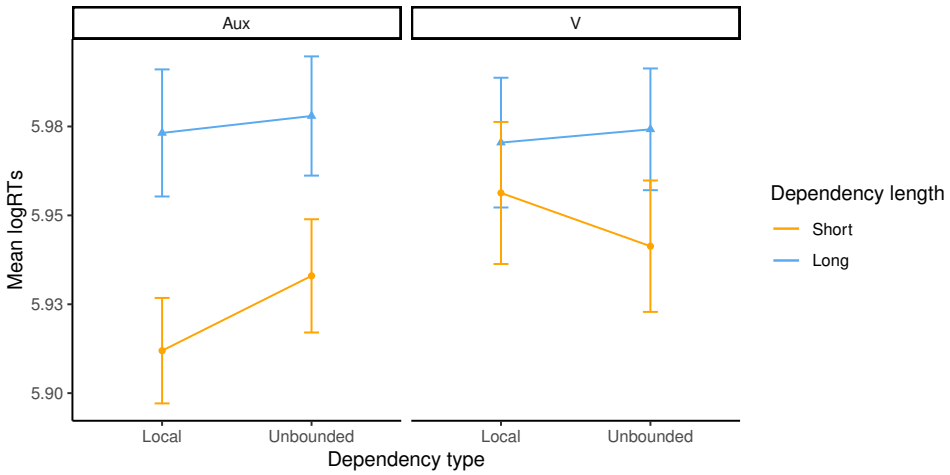


FIGURE 3.5: Raw reading times and residual logarithmic reading times (logRTs) at critical and postcritical regions in Experiment 4. Error bars show standard error.

4.7 Discussion

The aim of this experiment was to investigate (anti)locality effects in local and unbounded dependencies. The most important result of Experiment 4 is the finding of a locality effect in both local and unbounded dependencies at the auxiliary verb. The critical region was stable across conditions, unlike in Experiment 3, and yet results partially replicate. The main difference is that, in Experiment 3, the locality effect was observed throughout all critical and postcritical regions. In Experiment 4, however, the effect was only noticeable at the auxiliary

(A) Interaction plot of logRTs times at critical regions.



(B) Interaction plots of logRTs at postcritical regions.

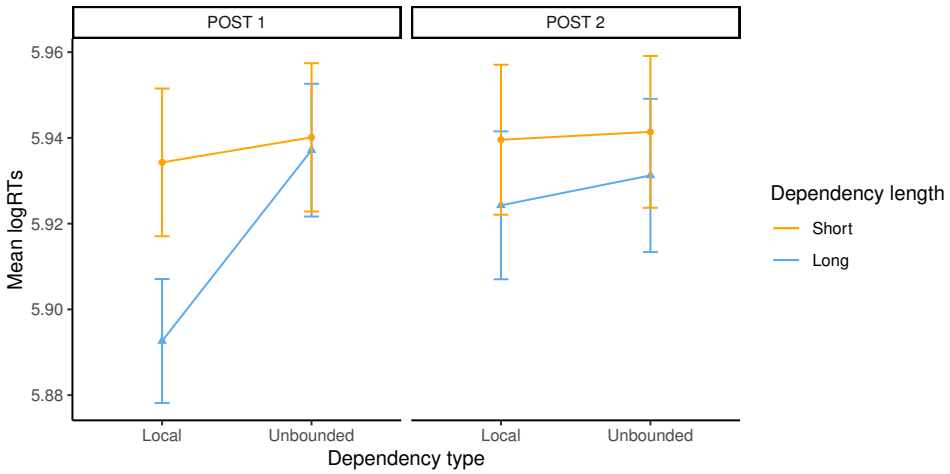


FIGURE 3.6: Interaction plots of logarithmic reading times (logRTs) at critical (top) and postcritical (bottom) regions in Experiment 4. Error bars show standard error.

	Estimate	SE	*t*-value	*p*-value
Auxiliary verb (Aux)				
(Intercept)	5.94	0.03	227.41	<0.001
Length	0.03	0.01	3.08	0.003
Type	-0.01	0.01	-1.41	0.16
Length x Type	0.002	0.01	0.25	0.80
Main verb (V)				
(Intercept)	5.96	0.03	204.35	<0.001
Length	0.01	0.01	1.85	0.07
Type	0.001	0.01	0.19	0.85
Length x Type	-0.01	0.01	-0.86	0.39
Postcritical 1 (POST 1)				
(Intercept)	5.92	0.02	243.23	<0.001
Length	-0.01	0.01	-1.68	0.09
Type	-0.01	0.01	-1.92	0.06
Length x Type	-0.01	0.01	-1.55	0.12
Postcritical 2 (POST 2)				
(Intercept)	5.93	0.03	204.85	<0.001
Length	-0.01	0.01	-0.89	0.37
Type	-0.004	0.01	-0.52	0.61
Length x Type	-0.01	0.01	-0.61	0.54

TABLE 3.6: Summary of the coefficients from the linear mixed models run on residual logartithmic reading times in Experiment 4.

verb, and the reading time pattern is even reversed at the first postcritical word (POST 1), with faster reading times reported in long unbounded dependencies in comparison to the other three conditions —although this difference did not yield any significant effect. Results seem to be less robust in Experiment 4 than in Experiment 3.

Overall, results provide further evidence in support of the hypothesis that locality effects are a general phenomenon in VO languages due to memory limitations and can be observed in all types of dependencies. That is, locality effects can be found outside unbounded dependencies and can be detected using behavioral measures.

5 Experiment 5

5.1 Overview

So far in this chapter, I have focused on the effects of increasing dependency length on sentence comprehension in both local and unbounded dependencies. The predictions made by expectation-based accounts are based on the assumption that the position of the verb is more predictable in local dependencies because they must be completed locally, in contrast with unbounded dependencies, where the dependency can span over more than one clause. However, we do not yet have estimates of how often speakers complete unbounded

dependencies locally. It could be the case that unbounded dependencies too are completed locally with a high frequency. In that scenario, there would be no difference in terms of expectations between local and unbounded dependencies. The aim of the present experiment is to obtain a quantitative measure of how probable it is that speakers complete a dependency depending on whether they have been exposed to the beginning of either a local or an unbounded dependency. In a written cloze completion task (Taylor, 1953), participants were asked to complete sentences containing local and unbounded dependencies. I measured the proportion of times the first constituent was a verb, i.e. the *cloze probability* of participants completing dependencies locally. If the predictability of V depends on dependency type, I predicted a higher proportion of verb responses in local dependencies in comparison to unbounded dependencies.

5.2 Participants

57 native speakers of Spanish (42 women, mean age = 32.8, SD = 8.8, range = 17–60) took part in the experiment through the online platform Ibex Farm. Participants were recruited using social media (Twitter and Facebook). They were given informed consent and received no compensation for their participation.

In order to make sure participants were native speakers of Spanish, they were asked whether they had been born and raised in Spain and whether they have always spoken Spanish at home. Participants who failed to answer positively to these two questions were removed from the analysis, affecting 15.8% of the participants (9 out of 57). In total, data from 48 participants was analyzed.

5.3 Materials

24 experimental items and 24 filler items were used in the experiment. The experimental sentences were based on the sentences used in Experiments 2 and 3. Sentences began with a fronted IO that needed to be integrated into a full sentence. Four experimental conditions were created by manipulating two factors (2 x 2 design): *OI length*, by either attaching a postnominal relative clause to the IO (long) or not (short); and *dependency type,* by either including a *wh*-operator in the IO (unbounded dependency) or not (local dependency).

5.4 Procedure

The experiment was an online written cloze completion task ran on Ibex Farm. Participants were asked to read the beginning of some sentences and complete them using their keyboard. They were instructed not to overthink and to finish the sentence with the first words that came to their mind. Sentences were presented using a black font on a white background. Before

Condition	Example
Short–Local	**A la abogada** _____ to the lawyer "To the lawyer…"
Long–Local	**A la abogada** [que buscaba el secretario judicial] _____ to the lawyer that look-for the secretary judicial "To the lawyer that the clerk of court was looking for…"
Short–Unbounded	¿**A qué abogada** _____? to what lawyer "To what lawyer…?"
Long–Unbounded	¿**A qué abogada** [que buscaba el secretario judicial] _____? to what lawyer that look-for the secretary judicial "To what lawyer that the clerk of court was looking for…?"

TABLE 3.7: Example of the experimental items used in Experiment 5.

starting the experiment, three practice sentences were displayed, although participants were not aware these were not part of the study. The experiment could be completed using a computer or a mobile device and there was no time limit. Including the initial questionnaire, the study did not last longer than 20 minutes.

5.5 Analysis

Each response was hand-coded as either *V* or *Other* and proportions were calculated. I coded a response as V if the first constituent in a participant's response was the main verb of the sentence (with or without an auxiliary verb and/or a clitic). The proportion of V responses was used as the dependent variable in the analysis. The data was analyzed using logit mixed effects models rather than traditional ANOVAs or linear mixed models because these are not appropriate for binomial data (Jaeger, 2008). Logit mixed models were fitted on the proportion of V responses including *IO length* and *dependency type* as fixed effects using the *lme4* package (Bates, Mächler, et al., 2015) in the R (version 3.5.3) programming environment (R Core Team, 2017). Since I did not have any predictions regarding an interaction, I did not include it in the model. Sum contrasts were defined for the fixed factors: *long IO* was coded as 1 and *short IO* as -1; and *local dependency* as 1 and *unbounded dependency* as -1. The most parsimonious model was fitted and complexity was added to the random-effects structure (by-subject and by-item random intercepts and slopes) only when supported by the data in order to avoid model overfitting (Bates, Kliegl, et al., 2015). After carrying out model comparisons by means of likelihood ratio tests, the following model was fitted: *proportion of V responses* $\sim$ *IO length*dependency + (1|subject) + (1+IO length|item)*. All *p*-values were calculated using the Satterthwaite's model of approximation with the *lmerTest* package (Kuznetsova et al., 2017).

5.6 Results

Mean proportions of V responses are reported in FIGURE 3.7. All graphs were generated using the *ggplot2* package (Wickham, 2016). The summary of the coefficients of the logit mixed model is provided in TABLE 3.8. Visual inspection of the data shows that the proportion of V responses was remarkably high in all four conditions, ranging from 0.89 [95% CI = 0.86–0.93] in short local dependencies to 1 in long local dependencies. That is, participants tended to complete dependencies as soon as soon as possible in most of the trials. Results from the logit mixed model show both a significant main effect of *IO length*, such that there was a greater probability of V appearing after a long IO (0.99 [95% CI = 0.99–1]) than after a short IO (0.94 [95% CI = 0.92–0.96]). There was also a main effect of *dependency type*: the probability of V appearing after IO was greater in unbounded dependencies (0.98 [95% CI = 0.97–0.99]) than in local dependencies (0.95 [95% CI = 0.93–0.97]). Both effects seem to be the produced by the lower proportion of V responses in the local short dependency condition in comparison to the other three conditions. Crucially in these results, proportions are in the high probability side of the scale, so differences between conditions may not be relevant.

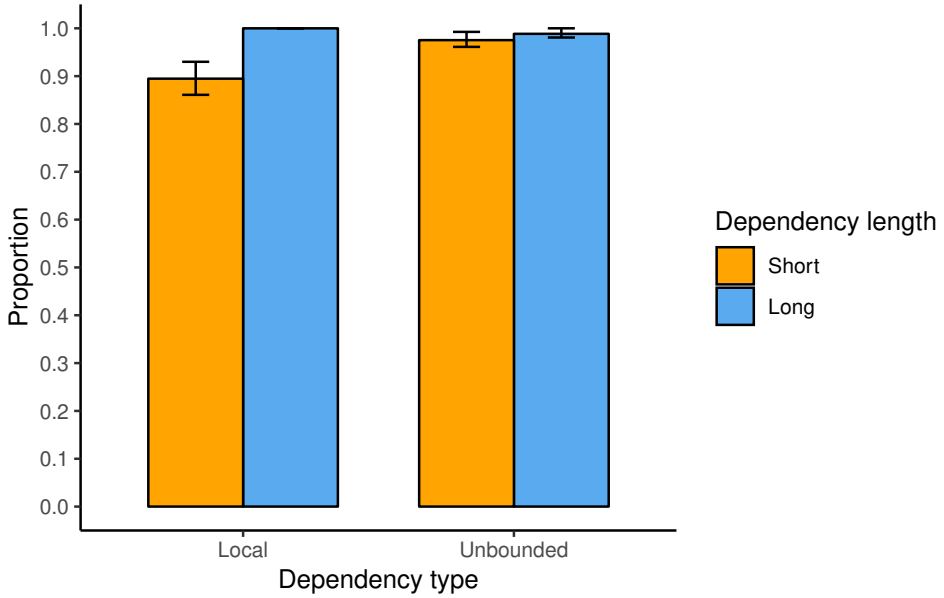


FIGURE 3.7: Proportion of V responses in Experiment 5. Error bars show 95% confidence intervals.

	Estimate	SE	*z*-value	*p*-value
(Intercept)	5.01	0.48	10.44	<0.001
Length	1.43	0.31	4.64	<0.001
Type	-0.68	0.19	-3.57	<0.001

TABLE 3.8: Summary of the coefficients from the logit mixed model in Experiment 5.

5.7 Discussion

In the present experiment, I obtained estimates of the probability of V appearing after a fronted IO in order to measure the proportion of times both local and unbounded dependencies are completed locally. I had hypothesized that local dependencies would yield a higher proportion of V responses than unbounded dependencies. Results show that speakers strongly tend to complete dependencies locally, producing the shortest dependency possible. This pattern is coherent with previous results from both experimental and corpus studies showing a preference towards the minimization of dependencies in language production (Hawkins, 1994; Futrell et al., 2015a; Ros et al., 2015; Ros, 2018).

Even though there was a difference between long and short dependencies and between local and unbounded dependencies, there is a high proportion of V responses —over roughly 90%— in all four conditions. This means that V was highly predictable in all experimental conditions and differences between conditions with such a high cloze probability may not have an impact on experimental results. This would explain why no difference across dependency types was observed in behavioral measures (Experiments 3 & 4). Surprisingly, according to expectation-based accounts of language processing, a high probability of V as the next constituent after seeing IO should increase expectations and result in an antilocality effect when dependency length is increased. However, results from Experiments 3 & 4 show a locality effect in both local and unbounded dependencies, supporting the hypothesis that increasing dependency length is a major determinant of processing difficulty in VO languages.

6 General discussion

The goal of this chapter was to investigate whether locality effects, i.e. increased processing difficulty when dependencies are longer, affected unbounded dependencies exclusively or both local and unbounded dependencies. It has been noted that basic word order seems to be correlated with the presence or absence of locality effects: VO languages show locality effects, but OV languages show antilocality, i.e. processing facilitation when dependency length is increased. However, most studies finding locality effects had only tested unbounded dependencies, whereas studies finding antilocality only tested local dependencies. According to expectation-based accounts of language comprehension, local dependencies may show a processing advantage because they must be completed locally and, therefore, the position of the verb is more predictable. In unbounded dependencies, however, a higher uncertainty due the possibility of additional clauses intervening before the verb should hinder processing. Memory-based accounts, however, predict processing difficulty in both local and unbounded dependencies when additional material is added between a dependent and its head due to either a higher integration cost or activation decay and similarity-based interference.

In Experiment 2, I carried out an online acceptability judgment task where participants

had to rate local and unbounded dependencies where dependency length was manipulated. I found no difference in ratings between conditions, suggesting that offline acceptability judgments may not be sensitive enough to capture subtle processing effects. In Experiments 3 and 4, I conducted two self-paced reading studies to measure the effect of dependency length on both local and unbounded dependencies during online dependency formation. These two studies provide evidence that longer dependencies yield processing difficulty at the verbal area where dependency completion takes place. Results suggest that locality is a general phenomenon in VO languages like Spanish, affecting all types of dependencies regardless of whether the head is more or less easily predicted. For the first time to my knowledge, results from Experiments 4 and 5 provide reading time data from a direct comparison between local and unbounded dependency with no structural differences, unlike in previous studies (Grodner & Gibson, 2005; Bartek et al., 2011). In Experiment 5, an online cloze completion task was run in order to obtain estimates of the probability of completing dependencies locally in both local and unbounded dependencies. Results show that participants complete dependencies locally in most cases, regardless of dependency length and dependency type.

Overall, this chapter provides further evidence in favor of memory-based accounts that predict locality effects in language comprehension when dependent-head distance is increased. The current experimental series adds a new study in Spanish to the pool of VO languages tested where locality effects have been reported. Crucially, locality effects were found in both local and unbounded dependencies, providing evidence that distance-based processing difficulty is a general phenomenon in verb-medial languages, and not particular to unbounded dependencies. Available evidence from this and previous studies suggests that locality effects are dominant in VO languages, whereas antilocality effects are more likely to occur in OV languages, an idea that has also been suggested by Levy et al. (2013). One plausible explanation could be that expectation-based facilitation can only be observed in OV languages, where there is a high proportion of verb-final clauses. As a consequence, speakers of OV languages may be using specific language statistics in order to make and maintain predictions of upcoming syntactic heads more efficiently than speakers of VO languages. A crucial piece of evidence comes from the so-called *VP-missing effect* (Gibson & Thomas, 1999). Double center embeddings such as (12a) are known to increase processing difficulty (Chomsky & Miller, 1963; Bever, 1970). In English, deleting the second V, as in (12b), makes the sentence unequivocally ungrammatical but creates a grammatical illusion, increasing reading times (Gibson & Thomas, 1999; Christiansen & MacDonald, 2009; Vasishth, Suckow, Lewis, & Kern, 2010). In German and Dutch, both with verb-final embedded clauses, deleting the second VP increases processing difficulty, as revealed by slower reading times, and decreases acceptability (Vasishth et al., 2010; Frank, Trompenaars, & Vasishth, 2016). Vasishth et al. (2010) argue that speakers of OV languages are able to make stronger predictions of the verbs and these are not forgotten. Taking into account evidence from (anti)locality and VP-missing effects, I conclude that specific distributional patterns in a language play a key role in shaping sentence processing and the strategies employed by speakers of those languages.

(12) a. The patient [the nurse [the clinic had hired] admitted] met Jack.

b. *The patient [the nurse [the clinic had hired] ~~admitted~~] met Jack.

In Experiment 5, an online cloze completion task was run in order to obtain estimates of the probability of completing dependencies locally in both local and unbounded dependencies. Results show that participants complete dependencies as soon as possible in most trials, regardless of dependency length and dependency type. This tendency to produce shorter dependencies has been attested crosslinguistically in both experimental and corpus studies (Hawkins, 1994; Futrell et al., 2015a; Ros et al., 2015; Ros, 2018). In other words, results show that the position of the verb is highly predictable in both local and unbounded dependencies. Contrary to the predictions made by expectation-based accounts, a sharper prediction of a verb did not speed up reading times when increasing dependency length in Experiments 3 and 4. As argued before, a plausible explanation is that predictions play a more prominent role in OV languages due to a higher exposure to verb-final clauses where dependency length is longer, overriding locality effects, unlike in VO languages.

7 Concluding remarks

In this chapter I investigated the impact of increasing dependency length in both local and unbounded dependencies. Available evidence of locality and antilocality suggested that locality effects, i.e. increased processing difficulty in long dependencies, was only present in unbounded dependencies, whereas antilocality effects, i.e. distance-based facilitation, was only observed in local dependencies. For the first time, I hypothesized that this pattern may result from differences in expectations: local dependencies must be resolved locally, and therefore the position of the verb is highly predictable; whereas unbounded dependencies may span over more than one clause and the position of the verb is then more uncertain. Expectation-based accounts of sentence comprehension predict that local dependencies should yield an antilocality effect, whereas unbounded dependencies should show locality. Memory-based accounts, however, predict that both types of dependencies should be subject to locality due to memory limitations. I provided evidence that locality effects are a general phenomenon in VO languages affecting all dependency types. In addition, even though the verb was highly predictable in both types of dependencies, no expectation-based facilitation was observed, probably due to the fact that expectations in VO languages are not maintained as strongly as in OV languages. Overall, the data presented in this chapter provides a strong validation of memory-based accounts of sentence comprehension.

Chapter 4

Antilocality effects in a free word order OV language: Evidence from Basque

Abstract

Memory-based accounts of sentence comprehension predict that increasing dependency length should hinder sentence comprehension (locality effect) due to the limitations of the human memory system (Gibson, 1998, 2000; Lewis & Vasishth, 2005). However, there is counterevidence from OV languages showing that longer dependencies facilitate language processing (antilocality effect). This facilitation is predicted by expectation-based accounts (Hale, 2001; Levy, 2008), as the position of the verb in OV languages is highly predictable. In fact, it has been suggested that head directionality plays a crucial role in online dependency resolution (Levy & Keller, 2013): VO languages show locality effects, whereas OV languages display antilocality effects. Currently available evidence from three OV languages (German, Hindi and Japanese) supporting this hypothesis is highly limited by the small amount of data and the fact that these languages are strictly verb-final. In the present chapter, I investigated whether Basque, an OV language allowing verb-medial word orders, shows antilocality effects. I ran an acceptability judgment task and a self-paced reading study using ditransitive main clauses where dependency length was manipulated. Reading times showed a clear processing facilitation in long dependencies already at a preverbal region, providing evidence in support of expectation-based accounts. An additional cloze completion task revealed that expectation effects do not seem to be driven by a preference for verb-medial word orders. The present study adds a new language to the pool and broadens the cross-linguistic evidence available for current theories regarding antilocality. Overall, results suggest that antilocality is a general phenomenon in OV languages, which correlates with a weaker tendency to minimize dependencies in language production.

1 Introduction

Increasing dependency length has been shown to hinder sentence comprehension (*locality* effect) in VO languages (Grodner & Gibson, 2005; Bartek et al., 2011). Memory-based accounts of sentence processing, such as *Dependency Locality Theory* (DLT) (Gibson, 1998, 2000) and the activation-based model (LV05) (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006; Lewis et al., 2006), predict locality effects due to the limited capacity of the cognitive resources devoted to build syntactic representations in real time comprehension. DLT predicts higher memory costs as a consequence of a greater storage and integration cost, whereas LV05 puts the focus on activation decay and similarity-based interference at retrieval. In Chapter 3, I provided evidence that in Spanish, a canonical VO language, increasing the distance between a dependent and its head hinders sentence comprehension in both local and unbounded dependencies. Those results, together with accumulating evidence from languages such as English (Gibson, 2000; Grodner & Gibson, 2005) and Russian (Levy et al., 2013) suggest that locality effects are a general phenomenon in VO languages.

Contrary to the predictions made by memory-based accounts, there is also evidence that increasing dependency length facilitates language processing, yielding an antilocality effect. For example, Levy and Keller (2013) found that adding intervening words in a dependency between a subject and a verb in German such as (1) facilitates processing at the verb in an eye-tracking experiment.

(1) a. . . . , hat **Hans Gerstner** den Fußball **versteckt**. . .
have.3SG Hans Gerstner the.ACC football hide.PTCP
". . . , Hans Gerstner hid the football. . . "

b. . . . , hat **Hans Gerstner** dem ungezogenen Sohn des fleißigen
have.3SG Hans Gerstner the.DAT naughty son the.GEN industrious
Hausmeisters den Fußball **versteckt**. . .
janitor the.ACC football hide.PTCP
". . . , Hans Gerstner hid the football from the naughty son of the industrious janitor. . . "

(Levy and Keller, 2013, p. 204)

This distance-based facilitation is predicted by expectation-based accounts of language comprehension (Hale, 2001; Levy, 2008). Levy (2008) argues that comprehenders use prior linguistic experience in order to make predictions regarding upcoming words. Cognitive resources are then allocated according to the probabilities of all possible continuations based on language statistics. These probabilities are updated whenever the parser is exposed to a new word. When incoming input is consistent with the initial predictions, comprehension is facilitated. However, if the input disconfirms those predictions, cognitive resources need to be reallocated, resulting in increased processing difficulty.

General distributional patterns may favor expectation-driven facilitation in OV languages in comparison to VO languages. In fact, evidence of antilocality effects comes almost exclusively from three OV languages: German (Konieczny, 2000; Konieczny & Döring, 2003), Hindi (Vasishth, 2003; Lewis & Vasishth, 2005) and Japanese (Nakatani & Gibson, 2008, 2010). In these languages, the verb is mostly at the end of the clause and, therefore, its position is highly predictable. If the length of a dependency between an argument and the verb is increased by adding extra words, intervening material increases the expectations for the verb. In cases where the intervening material is a verbal argument, the parser has more cues available in order to constrain its predictions regarding the verb. For this reason, it has been suggested that locality effects dominate in VO languages, whereas antilocality effects are a general phenomenon in OV languages (Levy & Keller, 2013). There are, however, some exceptions. Vasishth and Drenhaus (2011) ran a self-paced reading, an eye-tracking and an ERP experiment in order to investigate the effect of increasing dependency length on online sentence processing in German relative clauses. Results suggest that longer dependencies are more difficult for the parser. Similarly, Levy and Keller (2013) ran a second eye-tracking experiment using the same manipulation as in (1) within a relative clause and found that increasing dependency length yielded a locality effect. They argued that an increase in memory load may override expectation-driven facilitation and result in greater processing difficulty. Nonetheless, results should be taken with care as Vasishth et al. (2018) failed to replicate Levy and Keller's results in seven direct replication attempts.

Overall, currently available evidence suggests the empirical basis of antilocality effects is highly limited by the small amount of data and the reduced number of languages tested. Crucially, in all the OV languages where antilocality has been attested, namely German, Hindi and Japanese, the verb must obligatorily occupy the last sentential position, making the position of the verb highly predictable. In German main clauses, the finite verb must occupy the second position of the sentence (2a). In perfect forms, the tensed auxiliary occupies the second position and the lexical verb must obligatorily appear at the end of the clause (2b). In embedded clauses, both the lexical and the auxiliary verb must be placed in the last position (2c).

(2) a. Sie **liest** die Zeitung.
She read.3SG the newspaper
"She reads the newspaper."

b. Sie **hat** die Zeitung **gelesen**.
She have.3SG the newspaper read.PTCP
"She has read the newspaper."

c. Georg sagt, dass sie die Zeitung **gelesen** **hat**.
Georg say.3SG that she the newspaper read .PTCP have.3SG
"Georg says that she has read the newspaper."

In Japanese, both main and embedded clauses are strictly verb final. Japanese has relatively free word order, as in the ditransitive sentences in (3) (examples from Sasano and Okumura 2016: 2236). All arguments can be ordered indistinctly but, crucially, the verb must always occupy the last sentential position.

(3) a. *S-IO-O-V*

Ken-ga	Aya-ni	camara-wo	miseta
S	IO	O	V

"Ken showed the camera to Aya."

b. *S-O-IO-V*

Ken-ga	camara-wo	Aya-ni	miseta
S	O	IO	V

c. *IO-O-S-V*

Aya-ni	camara-wo	Ken-ga	miseta
IO	O	S	V

d. *O-S-IO-V*

Camara-wo	Ken-ga	Aya-ni	miseta
O	S	IO	V

e. Other orders: *IO-O-S-V* and *O-IO-S-V*.

Hindi has also relatively free word order and all possible combinations are allowed in both transitive and ditransitive sentences (Kachru, 2006). However, non-canonical word orders, i.e. word orders other than S-(IO)-O-V, are rare. Husain, Bhatt, and Vasishth (2013) provided statistics of word order variation from a dependency treebank in Hindi (Bhatt et al., 2009), showing that non-canonical word orders amount to only 11% of occurrences. Crucially, only 0.35% of the sentences included a verb in medial position, showing that Hindi has a strong preference for verb-final clauses.[1]

Besides word order rigidity, most previous studies have generally focused on complex constructions where processing difficulty is increased. For instance, evidence of antilocality effects in Hindi (Vasishth, 2003; Vasishth & Lewis, 2006) and Japanese (Nakatani & Gibson,

[1]Husain et al. (2013) note that a limitation of the sentences included in the dependency treebank is that they were extracted from written texts. Verb-medial sentences might be more common in colloquial spoken Hindi.

2008, 2010) comes exclusively from single or double center-embeddings, which are known to increase memory demands. In German, Levy and Keller (2013) used long sentences introducing different NPs that increased difficulty in reading comprehension. To the best of my knowledge, only one study has found antilocality effects in main clauses in German (Konieczny, 2000). Finding distance-based facilitation in a structurally simple main clause would constitute a sound piece of evidence in favor of expectation-based accounts.

1.1 An overview of Basque

Basque is a language isolate spoken by over 750,000 in the Basque area in northeastern Spain and southwestern France (Eusko Jaurlaritza, 2016). It has a number of characteristics that makes it different from the Romance languages spoken in the same territory —Spanish and French—, such as agglutinative morphology and ergative case marking (Laka, 1996; Hualde & Ortiz de Urbina, 2003). Importantly for the present chapter, Basque is an OV language with free word order. For example, a ditransitive sentence in Basque canonically follows the order S-IO-O-V, as shown in (4a), but all other word order combinations are possible without modifying the propositional content of the sentence (4).

(4) a. *S-IO-O-V*

Amaiak	etorkinari	dokumentua	eman dio
S	IO	O	V

"Amaia gave the document to the immigrant."

b. *S-O-IO-V*

Amaiak	dokumentua	etorkinari	eman dio
S	O	IO	V

c. *S-O-V-IO*

Amaiak	dokumentua	eman dio	etorkinari
S	O	V	IO

d. *O-IO-V-S*

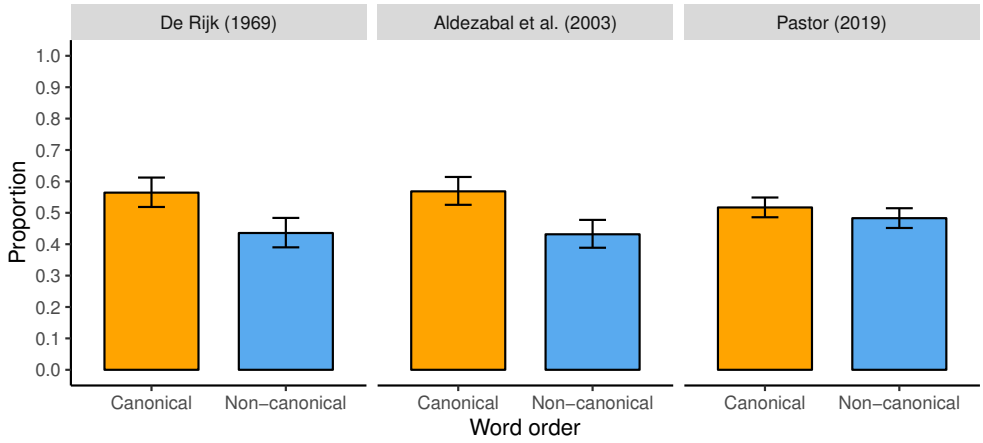
Dokumentua	etorkinari	eman dio	Amaiak
O	IO	V	S

e. Other orders: *O-S-V-IO*, *S-IO-V-O*, etc.

The availability of different word orders in Basque has been investigated in different corpus studies analyzing transitive declarative sentences (De Rijk, 1969; Aldezabal et al., 2003; Pastor, 2019). De Rijk (1969) looked at word order frequencies in 459 sentences in Basque extracted from literary work, Aldezabal et al. (2003) analyzed 512 sentences from the newspaper *Euskaldunon Egunkaria*, and Pastor (2019) analyzed 1054 sentences from the *Ereduzko Prosa Gaur* corpus, including literary work, newspapers, magazines and television scripts. Despite the limited number of sentences in comparison to dependency treebanks —for example, the Spanish treebank from Universal Dependencies v1 (Nivre et al., 2016) contains over 16,000 sentences— , the data from these three studies includes a variety of different genres that allows us to shed light on word order flexibility in Basque. As seen in FIGURE 4.1a, a proportion between 0.52 and 0.57 of the sentences followed canonical word order (SOV), whereas a proportion from 0.43 to 0.48 had some other word order (SVO, OVS, OSV, VSO or VOS). This pattern was observed across all three corpus studies and suggests that word order is highly flexible in Basque. In addition, Futrell, Mahowald, and Gibson (2015b) quantified word order freedom by measuring S-O order entropy using data from dependency treebanks in 37 languages. In general, in languages with rich case marking the entropy value was higher, meaning that that the order of S and O is more variable than in languages with poorer or no case marking. Importantly, results suggest that word order in Basque is much more flexible (entropy value: $\sim$0.80) than in Japanese ($\sim$0.50), German ($\sim$0.50) or Hindi ($\sim$0.30).

Crucially for the purposes of the present chapter, Basque allows verb-medial clauses, whereas German, Hindi and Japanese are strictly verb-final languages. FIGURE 4.1b shows the proportion of verb-final, -medial and -initial clauses found in De Rijk (1969), Aldezabal et al. (2003) and Pastor (2019). Even though verb-final clauses are the most common ones (0.61–0.69), verb-medial clauses are not infrequent (0.28–0.36), proving that Basque is not rigid regarding the position of the verb, unlike German, Hindi and Japanese.

(A) Proportion of canonical (SOV) and non-canonical word orders in transitive sentences.



(B) Proportion of verb-final, -medial and -initial word orders in transitive sentences.

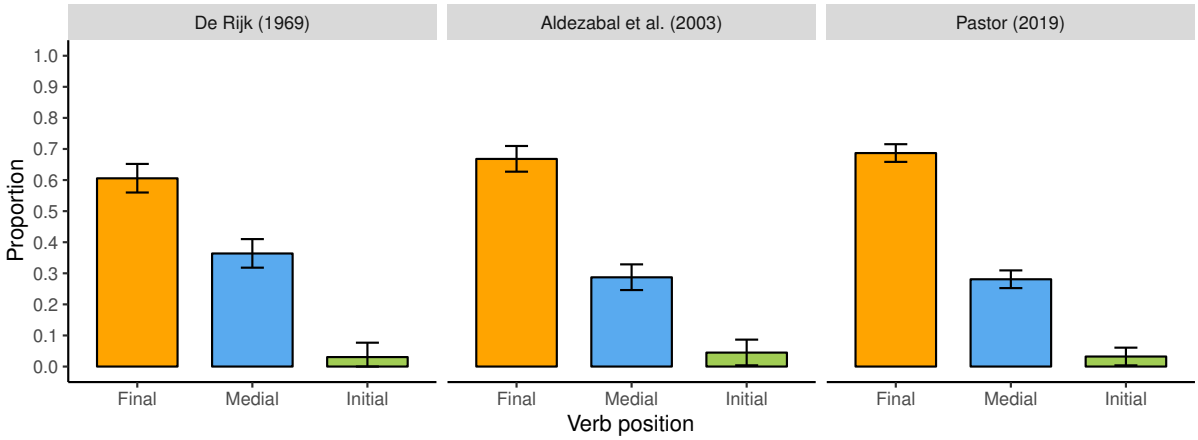


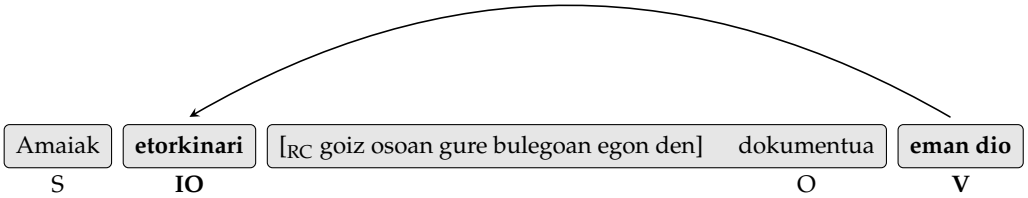
FIGURE 4.1: Summary of the results from three corpus studies targeting transitive sentences in Basque (De Rijk, 1969; Aldezabal et al., 2003; Pastor, 2019): proportion of canonical and non-canonical word orders (top), and proportion of verb-final, -medial and -initial word orders (bottom). Error bars show 95% confidence intervals.

There is also evidence from controlled production studies showing that Basque speakers produce verb-medial clauses rather frequently. Ros et al. (2015) ran a cued recall production task to investigate to what extent speakers of Basque resort to non-canonical word orders in ditransitive sentences in order to minimize dependency length.[2] In their experiment, participants were asked to look at a computer screen where four boxes were displayed, each one containing a sentential constituent —S, IO, O or V. After solving a simple arithmetic operation, the verb was displayed again as a cue for participants to produce a sentence using the constituents they saw at the beginning. In the design, the length of either IO or O was

[2]Ros et al. (2015) tested both transitive and ditransitive sentences. Here I am focusing exclusively on ditransitive sentences for the sake of simplicity and because the experimental materials in this chapter are ditransitive sentences in Basque.

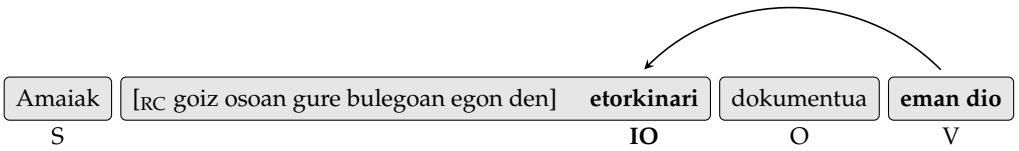
manipulated by attaching a prenominal relative clause. Since having a long O in a ditransitve sentence with canonical word order (S-IO-O-V) would increase dependency length between IO and V (5a), Ros and colleagues hypothesized that participants would produce shifted word orders where O precedes IO. No IO-O shifting was predicted when IO was long (5b). Results show that participants produced shifted (non-canonical) orders in 60.1% of the trials. Crucially, in 55.4% of those non-canonical productions participants placed the verb in medial position (e.g. S-O-V-IO). In summary, data from both corpus and experimental studies provides evidence that Basque is not a strictly verb-final language.

(5) a. *Long O – Long dependency*



"Amaia gave the document that has been in our office all morning to the immigrant."

b. *Long IO – Short dependency*



"Amaia gave the document to the immigrant that has been in our office all morning."

1.2 The present study

As outlined above, it has been hypothesized that (anti)locality effects are correlated with head-directionality: VO languages show locality effects when dependency length is increased, but OV languages show antilocality effects. However, evidence of antilocality effects so far comes exclusively from strictly verb-final languages (German, Hindi and Japanese). The fact that Basque has word order flexibility and the verb does not have to occupy the final position of the clause obligatorily makes it the perfect case study to test this hypothesis. According to expectation-based accounts of sentence comprehension, antilocality effects are the result of expectation-driven facilitation. In other words, highly predictable continuations facilitate language processing. In strictly verb-final languages, the parser highly expects the verb to

be the last element of the clause. Therefore, intervening material between an argument and the verb increases the expectations for the verb. However, it can be the case that this effect can only be observed in strictly verb-final languages and not in flexible OV languages such as Basque, where the verb can also appear in medial position relatively frequently. In that case, one could conclude that locality effects can only be overridden when word order rigidity allows the parser to highly constrain possible continuations. Another concern regarding the currently available evidence of antilocality is that effects have been tested in complex and/or infrequent structures that increase processing demands. If expectation-based accounts are on the right track, antilocality should also be found in relatively simple main clauses.

In this chapter, I report on three experiments using ditransitive main clauses such as (3) in Basque, a flexible OV language, in order to investigate whether antilocality (a) is a general phenomenon in OV languages or is restricted to strictly verb-final languages and (b) is a general phenomenon in both main and embedded clauses. I ran an acceptability judgment task (Experiment 6) and a self-paced reading experiment (Experiment 7) in order to investigate whether increasing dependency length in a main ditransitive clause hindered or facilitated dependency completion at the verb. In addition, I ran a cloze completion production task (Experiment 8) in order to estimate the proportion of times speakers produce canonical verb-final orders when dependency length is manipulated.

2 Experiment 6

2.1 Overview

The current acceptability judgment task investigated whether increasing dependency length facilitates or hinders sentence comprehension in Basque, a free word order OV language. I presented Basque native speakers with ditransitive sentences where I manipulated the distance between the IO and the V and asked them to rate them using a 7-point Likert scale. The linking hypothesis was that increased processing difficulty should yield lower acceptability ratings.

Our predictions were the following. Antilocality effects, i.e. processing facilitation when dependency length is increased, had previously been reported in rigid OV languages which are strictly verb-final. If antilocality effects are a general phenomenon in OV languages due to a greater exposure to verb-final clauses, as it has been hypothesized, I predicted long dependencies to be more acceptable than short dependencies in Basque. However, if antilocality effects are restricted to rigid OV languages, I predicted long dependencies to be less acceptable than short dependencies in Basque, i.e. a locality effect.

Cond	Example
Short	Amaiak [goiz osoan gure bulegoan egon den] **etorkinari** dokumentua **eman dio** …
	Amaia-ERG.3SG [morning all-LOC our office-LOC be-PTPC AUX-that] inmigrant-DAT.3SG document-ABS.3SG give AUX
	"Amaia gave the document that has been in our office all morning to the immigrant."
Long	Amaiak **etorkinari** [goiz osoan gure bulegoan egon den] dokumentua **eman dio** …
	Amaia-ERG.3SG inmigrant-DAT.3SG morning all-LOC our office-LOC be-PTPC AUX-that document-ABS.3SG give AUX
	"Amaia gave the document that has been in our office all morning to the immigrant."

TABLE 4.1: Sample set of the sentences used in Experiments 6 & 7. IO and V are highlighted in bold.

2.2 Participants

106 native speakers of Basque (47 women and 1 non-binary, mean age = 42.2, SD = 12.3, range = 17–70) took part in the experiment through the online platform Ibex Farm. Participants were recruited using social media (Twitter and Facebook). They were given informed consent and received no compensation for their participation.

In order to make sure participants were native speakers of Basque, they were asked whether they had been born and raised in the Basque Country and whether they had always spoken Basque at home. Participants who failed to answer positively to these two questions were removed from the analysis, affecting 29.25% of the participants (31 out of 106). Overall, data from 75 participants was analyzed.

2.3 Materials

24 experimental items adapted from Ros et al. (2015) and 72 filler items were used in the experiment. The experimental items were ditransitive sentences in Basque with canonical word order (S-OI-O-V). Two experimental conditions were created by manipulating the distance between IO and V by attaching a prenominal relative clause to either OI (in the *short* condition) or O (in the *long* condition) (see TABLE 4.1 to see an example of the materials). IO was always animate, whereas O was always inanimate and all relative clauses were made up of five words. All experimental items are listed in Appendix D.

Filler items were carefully designed in order to mask the true purpose of the study. 48 filler items were ungrammatical in order to avoid any equalizing bias (Poulton, 1979), i.e. participants trying to balance the number of times they give high and low ratings, even after being exposed to fully grammatical sentences. 24 filler items included a relative clause, so that half of the items in the experiment contained one.

2.4 Procedure

Participants were asked to read some sentences and rate them according to how acceptable they seemed. Sentences were presented on a white background. Acceptability was explained in terms of plausibility and naturalness, and participants were told not to apply any normative rule they may have learned at school through explicit instruction. A 7-point Likert scale was provided to rate the sentences (1 = totally unacceptable, 7 = totally acceptable). Before starting the experiment, six practice sentences with gradient acceptability were displayed, although participants were not aware these were not part of the study. The experiment could be completed using a computer or a mobile device. Including the initial questionnaire, the study did not last longer than 20 minutes.

2.5 Analysis

Raw ratings were transformed into *z*-scores taking into account each participant's mean and standard deviation for two reasons. First, ratings from Likert scales are categorical and cannot be entered into a regression model, whereas *z*-score ratings are continuously distributed. Second, *z*-score transformation standardizes ratings and gets rid of scale biases, i.e. participants using only some points of the scale.

I fitted a linear mixed model on *z*-score ratings including *dependency length* as fixed factor in the R (version 3.5.3) programming environment (R Core Team, 2017). Sum contrasts were defined for the fixed factor: *long* was coded as 1 and *short* as -1. The most parsimonious model was fitted, including only varying intercepts by subject and by item, and complexity was added to the random-effects structure only when supported by the data in order to avoid model overfitting (Bates, Kliegl, et al., 2015). After carrying out model comparisons by means of likelihood ratio tests, the following model was fitted: *z.score.rating* $\sim$ *length*dependency* + *(1 | subject)* + *(1 | item)*. All *p*-values were calculated using the Satterthwaite's model of approximation using the *lmerTest* package (Kuznetsova et al., 2017).

2.6 Results

Results are summarized in FIGURE 4.2. Mean raw ratings are reported in FIGURE 4.2a (left) and mean *z*-score ratings in FIGURE 4.2b (right). All graphs were generated using the *ggplot2* package (Wickham, 2016). Note that I have included raw ratings because they can be interpreted more intuitively, but no analysis was run on them. The summary of the coefficients of the model is provided in TABLE 4.2. Results revealed no difference between the short and the long condition. Participants gave similar ratings to the experimental sentences, regardless of dependency length.

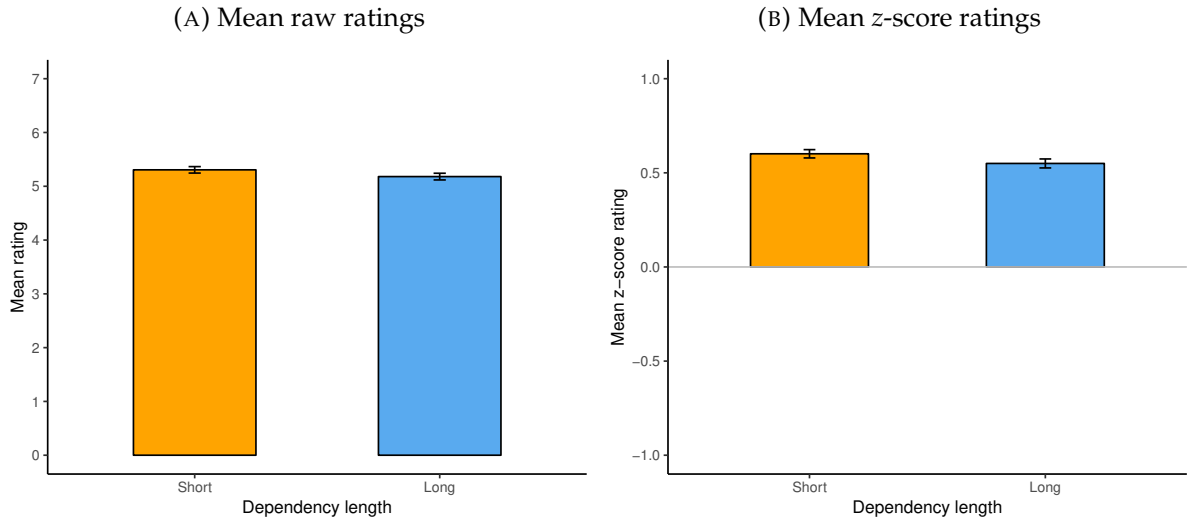


FIGURE 4.2: Mean acceptability ratings obtained in Experiment 6. Error bars show standard error.

	Estimate	SE	*t*-value	*p*-value
(Intercept)	0.57	0.05	10.95	<0.001
Length	-0.02	0.03	-0.62	0.54

TABLE 4.2: Summary of the coefficients from the linear mixed model in Experiment 6.

2.7 Discussion

The current experiment tested the impact of dependency length on acceptability in Basque ditransitive sentences. The predictions for this experiment were that long dependencies should be more acceptable if antilocality effects are a general phenomenon in OV languages, but hinder processing (locality effects) if antilocality is restricted to rigid word order OV languages. Based on the results, I found no evidence in favor of either prediction.

As in Experiment 2 (Chapter 3), it should be noted that null results should be interpreted with care in frequentist analyses because the null hypothesis may not be rejected due to high Type II error in underpowered studies. However, it can be the case that no difference was found because acceptability judgments are not sensitive enough to capture effects of dependency length in grammatical sentences. The ditransitive sentences used in Experiment 6 follow the canonical word order in Basque (S-IO-O-V) and its frequency is high. Even if dependency length is affecting processing facilitation/difficulty, offline tasks may not be the best tool to capture these subtleties. In Experiment 7, I tested the same materials using an online task (self-paced reading).

3 Experiment 7

3.1 Overview

In this self-paced reading task, I investigated how dependency length impacts online sentence processing in Basque. In Experiment 6, no differences between short and long dependencies in terms of acceptability were observed, probably because offline acceptability judgments cannot capture subtle processing effects in grammatical sentences. A self-paced reading task was conducted using the same sentences as in Experiment 6, where participants had to read them word-by-word while reading times for each word were recorded.

The predictions were similar to those in Experiment 6. If antilocality effects are a general phenomenon in OV languages, I predicted long dependencies to show faster reading times at the critical region. If antilocality effects are specific of rigid word order OV languages, then I expected long dependencies to elicit slower reading times at the critical region.

3.2 Participants

44 students from the University of the Basque Country (UPV/EHU) (39 women, mean age = 20.14, SD = 1.42, range = 18–25) took part in the experiment. All participants self-reported being native speakers of Basque through a questionnaire filled out prior to the experiment. Participants gave informed consent and received €6 for their participation.

Participants whose accuracy in the comprehension questions fell below 70% were excluded from the analysis. Likewise, participants whose mean reading times were 2.5 SD above or below the general mean were excluded. No participants were removed from the analysis.

3.3 Materials

24 experimental items and 72 filler items were used in the experiment. The experimental sentences were the same as in Experiment 6 (see TABLE 4.1 for an example of the materials, and Appendix D for a full list). The experimental items were ditransitive sentences in Basque with canonical word order (S-IO-O-V). Dependency length between the IO and the V was manipulated by attaching a prenominal relative clause to either IO (*short*) or O (*long*).

The filler items were not the same as in Experiment 6. The reason for this is because ungrammatical fillers were included in the acceptability judgment task, and I wanted to avoid any effect ungrammatical sentences may have on reading times. 24 of the filler items included a relative clause, so that half of the sentences had one.

In order to make sure participants were actually paying attention to the sentences, all experimental items were followed by a yes/no comprehension question, as well as 64 of the filler items. Overall, two thirds of the trials included a comprehension question.

3.4 Procedure

The experiment was a word-by-word moving window self-paced reading task. Participants were comfortably seated in front of a computer in an experimental booth. At the beginning of each trial, one dash per word was displayed on the screen. Every time the participant pressed the space bar, a new word appeared, but the previous one disappeared, so that they could only see one word at a time. Participants were instructed to read at a natural pace and to understand the sentences they were reading. After two thirds of the trials, they had to answer a yes/no comprehension question by pressing one of two buttons. Feedback was provided after every answer. Each experimental list was divided into four blocks of 24 trials, so that there were three breaks.

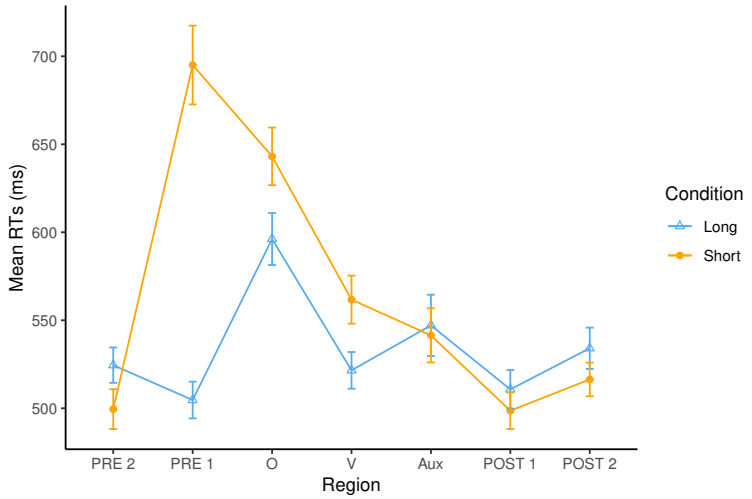
3.5 Analysis

Three words made up the critical region in the current experiment: the object (O), the main verb (V) and the auxiliary verb (Aux). Normally, in online studies testing the effect of dependency length on sentence processing, reading times are measured at the verb because it is where dependencies are resolved. In these experimental sentences in Basque, the verbal phrase is made up of two orthographical words: the main verb containing lexical and aspect information, and the auxiliary verb containing tense and agreement information (Laka, 1996; Hualde & Ortiz de Urbina, 2003). Since it is not clear where exactly retrieval takes place, I have decided to include both V and Aux as critical regions. With regards to O, the reasoning behind it being a critical region is different. Antilocality effects have been explained in terms of expectation-driven facilitation. If after seeing the sequence S-IO participants are certain that they are in a ditransitive frame, expectation for O should increase. Therefore, any facilitation as a result of high predictability should already be found at O.

Reading times below 200 ms. and above 4000 ms. were removed because they cannot possibly be reflecting language related processes. Raw reading times were then log-transformed in order to normalize the data. The word preceding O was different between conditions: IO in the *short* condition and the auxiliary verb of the relative clause in the *long*. Crucially, IO was always longer than one-syllable auxiliary verbs, and they were therefore expected to elicit longer reading times. This effect may have spilled over the critical region, leading to spurious results. In order to control for this confound, I fitted a linear mixed model on logarithmic reading times including word length (in number of characters) as fixed effect and random intercepts by subject using the *lme4* package (Bates, Mächler, et al., 2015) in the R (version 3.5.3) programming environment (R Core Team, 2017). The residuals from this model

were used as the dependent variable in the final analysis. A linear mixed model was fitted on residual logartihmic reading times with dependency length as fixed factor. Sum contrasts were defined for the fixed factor: *long* was coded as 1 and *short* as -1. The most parsimonious model was fitted, including only varying intercepts by subject and by item, and complexity was added to the random-effects structure when supported by the data in order to avoid model overfitting (Bates, Kliegl, et al., 2015). After carrying model comparisons by means of likelihood ratio tests, the following model was fitted when analyzing logarithmic reading times at both critical and postcritical regions: *residual logRTs* $\sim$ *length* + *(1 | subject)* + *(1 | item)*. All *p*-values were calculated using the Satterthwaite's model of approximation using the *lmerTest* package (Kuznetsova et al., 2017). A separate analysis was carried out for each critical and postcritical word.

(A) Mean raw reading times



(B) Mean residual logarithmic reading times

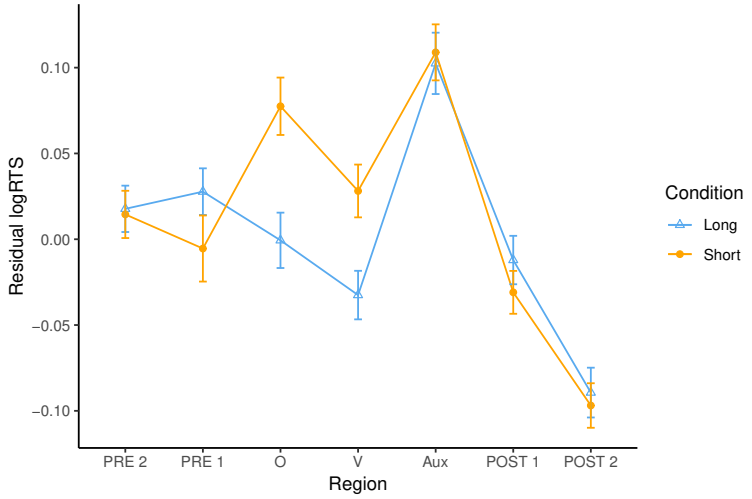


FIGURE 4.3: Raw reading times and residual logarithmic reading times (logRTs) at precritical, critical and postcritical regions in Experiment 7. Error bars show standard error.

3.6 Results

Results are summarized in FIGURE 4.3. Mean reading times in miliseconds are reported in FIGURE 4.3a (top) and mean residual logarithmic reading times in FIGURE 4.3b (bottom). All graphs were generated using the *ggplot2* package (Wickham, 2016). Note that I have included raw reading times because they can be interpreted more intuitively, but no analysis was run on them.

Visual inspection of the raw reading times revealed a large difference between the two conditions in the precritical region (PRE 1). As explained in the previous section, this difference may have been due to the fact that words of different length were being compared at the precritical region. Once word length is entered into the model as fixed factor and residual logarithmic reading times are extracted, this difference is no longer noticeable.

At the object (O), results revealed a difference between the two conditions: long dependencies elicited faster reading times than short dependencies. The same difference holds at the main verb (V). However, at the auxiliary verb (Aux) no difference was found between long and short dependencies, i.e. the auxiliary was read at a similar pace.

	Estimate	SE	*t*-value	*p*-value
Object (O)				
(Intercept)	0.04	0.02	1.76	0.09
Length	0.04	0.01	3.49	<0.001
Main verb (V)				
(Intercept)	0.001	0.02	0.05	0.96
Length	0.03	0.01	2.83	0.005
Auxiliary verb (Aux)				
(Intercept)	0.11	0.02	4.61	<0.001
Length	0.004	0.01	0.37	0.71

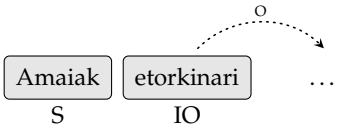
TABLE 4.3: Summary of the coefficients from the linear mixed models run on residual logartithmic reading times in Experiment 7.

3.7 Discussion

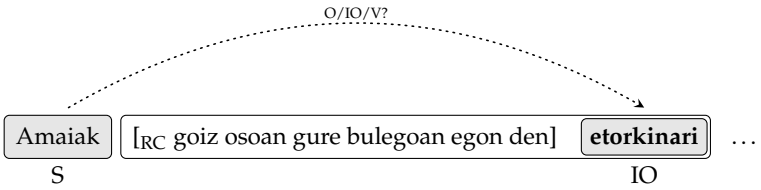
The goal of the current experiment was to investigate whether antilocality effects arise in a non-rigid OV language (Basque). I tested the hypothesis that long dependencies ease sentence comprehension in all OV languages, regardless of whether they are strictly verb-final or not. Results show faster reading times in long dependencies than in short dependencies in Basque main ditransitive sentences. In most studies, (anti)locality effects are measured at the verb because that is where arguments are integrated. Crucially, I found evidence of facilitation already at O, probably because O was more expected in the long condition than in the short condition. Finding evidence of expectation-based facilitation in the preverbal area is a strong support of expectation-based accounts.

There are at least two possible explanations why the effect was already noticeable at O. The first one is that different expectations were triggered by the beginning of the sentences used in Experiment 7. In the long condition, participants saw S and the head of IO at the beginning of the sentence, knowing immediately they were most likely in a ditransitive frame (6a). Expectation for O was thus high and increased as the prenominal relative clause unfolded. In the short condition (6b), however, participants saw S and then intervening material, increasing the uncertainty with respect to what the next sentential argument would be. The reduction of uncertainty in (6a) in comparison to (6b) may explain why an antiocality effect arose already at O.

(6) a. *Long dependency*

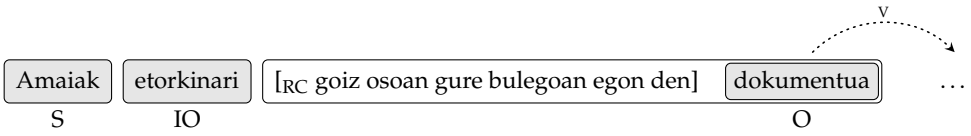


b. *Short dependency*

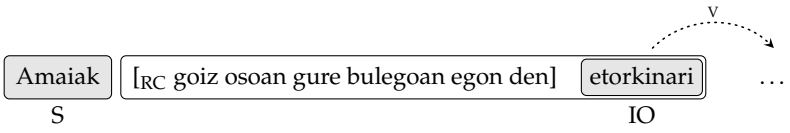


An alternative, but not excluding, explanation is that expectations are driven by informational structure. As outlined in the introduction, the basic word order of a ditransitive sentence in Basque is S-IO-O-V but all other orders are possible. Even though changing the word order does not alter the prepositional content of the sentence, different linearizations yield different informational patterns. The focus of the sentence, that is, the most relevant piece of information, immediately precedes the verb in Basque (Laka, 1996; Hualde & Ortiz de Urbina, 2003). In the design used in Experiment 7, attaching a prenominal relative clause to either IO or O makes them prominent information-wise, so participants would expect V to immediately follow. In the long condition (7a), the expected word order would therefore be the canonical one (S-IO-O-V). In the short condition (7b), on the other hand, a verb in medial position would be expected after IO (S-IO-V-O). The prediction of V after a long IO would have yielded slower reading times at O. In Experiment 8, I addressed this question by measuring the proportion of verb-medial productions after a long IO.

(7) a. *Long dependency*



b. *Short dependency*



4 Experiment 8

4.1 Overview

Reading times from Experiment 7 show that long dependencies were read faster than short dependencies. Although processing difficulty or facilitation as a function of dependency length is usually measured at the verb, an effect at the preverbal O was reported, providing evidence in favor of expectation-based accounts. However, the reason why O was less expected in the short condition than in the long condition is yet unclear.

In the present experiment, I conducted a written production task in order to investigate whether the antilocality effect reported in Experiment 7 is due to participants predicting a verb in medial position after a long salient IO, or simply because the first words in one condition trigger sharper predictions regarding the verbal argument structure from the beginning than in the other condition. In order to do so, I conducted a cloze-completion task where participants had to complete sentences after seeing S and a short or a long IO. A larger proportion of V responses after seeing a long IO in comparison to a short IO would suggest speakers shape their expectations with regards to upcoming words taking into account informational structure.

4.2 Participants

132 native speakers of Basque (89 women and 3 non-binary, mean age = 36.4, SD = 11.6, range = 20–70) took part in the experiment through the online platform Ibex Farm. Participants were recruited using social media (Twitter and Facebook). They were given informed consent and received no compensation for their participation.

In order to make sure participants were native speakers of Basque, they were asked whether they had been born and raised in the Basque Country and whether they had always spoken

Basque at home. Participants who failed to answer positively to these two questions were removed from the analysis, affecting 28.03% of the participants (37 out of 132). In total, data from 95 participants was analyzed.

4.3 Materials

24 experimental items and 24 filler items were used for the experiment. The experimental sentences were based on the materials used for Experiments 2 and 3. Experimental sentences were the beginning of a canonical ditransitive sentence (S-IO) (see TABLE 4.4). The length of the IO was manipulated by either attaching a prenominal relative clause to IO (*Long-IO*) or not (*Short-IO*). A full list of the materials is provided in Appendix D.

Cond	Example
Short IO	Amaiak **etorkinari** _____ Amaia-ERG.3SG inmigrant-DAT.3SG "Amaia to the immigrant..."
Long IO	Amaiak **[goiz osoan gure bulegoan egon den] etorkinari** _____ Amaia-ERG.3SG [morning all-LOC our office-LOC be-PTPC AUX-that] inmigrant-DAT.3SG "Amaia to the immigrant that has been in our office all morning..."

TABLE 4.4: Sample set of the sentences used in Experiment 7.

4.4 Procedure

The experiment was an online written cloze completion task ran on Ibex Farm. Participants were asked to read the beginning of some sentences and complete them using their keyboard. They were instructed not to overthink and to finish the sentence with the first words that came to their mind. Sentences were presented using a black font on a white background. Before starting the experiment, three practice sentences were displayed, although participants were not aware these were not part of the study. The experiment could be completed using a computer or a mobile device and there was no time limit. Including the initial questionnaire, the study did not last longer than 30 minutes.

4.5 Analysis

Each response was hand-coded as either V, O or Other and proportions were calculated. A response was coded as O if the first constituent in a participant's response was the object of the main sentence —regardless of whether it was a one-word NP or a complex NP, i.e. a noun modified by a relative clause. The proportion of O responses was used as the dependent variable in the analysis. The data was analyzed using logit mixed effects models rather than traditional ANOVAs because these are not appropriate for binomial data (Jaeger, 2008). Logit mixed models were fitted on the proportion of V responses including *IO length* as fixed effect

using the *lme4* package (Bates, Mächler, et al., 2015) in the R (version 3.5.3) programming environment (R Core Team, 2017). Sum contrasts were defined for the fixed factor: *long IO* was coded as 1 and *short IO* as -1. The most parsimonious model was fitted and complexity was added to the random-effects structure (by-subject and by-item random intercepts and slopes) when supported by the data in order to avoid model overfitting (Bates, Kliegl, et al., 2015). After carrying out model comparisons by means of likelihood ratio tests, the following model was fitted: *proportion of V responses* $\sim$ *IO-length + (1 | subject) + (1 | item)*. All *p*-values were calculated using the Satterthwaite's model of approximation. All *p*-values were calculated using the Satterthwaite's model of approximation with the *lmerTest* package (Kuznetsova et al., 2017).

4.6 Results

Results are summarized in FIGURE 4.4. The mean proportion of V responses (verb-medial word order) is reported in FIGURE 4.4a (left) and the mean proportion of O responses (canonical word order) in FIGURE 4.4b (right). All graphs were generated using the *ggplot2* package (Wickham, 2016). The summary of the coefficients of the model is provided in TABLE 4.5. Visual inspection of the data shows that participants completed the sentence with O much more frequently than with V, and IO length did not have any impact on the pattern of responses. Participants completed sentences containing a short IO with O (0.71 [95% CI = 0.68–0.74]) more frequently than with V in (0.23 [95% CI = 0.21 – 0.26]) of the trials. Sentences containing a long IO were also completed with O (0.70 [95% CI = 0.68 – 0.73]) in a higher proportion than with V (0.22 [95% CI = 0.20 – 0.25]). Results from the logit mixed model reveals no effect of *IO length*, i.e. there was no difference between short and long IO. Overall, participants produced verb medial orders in a low proportion of trials and regardless of whether IO was short or long.

	Estimate	SE	*z*-value	*p*-value
(Intercept)	-1.41	0.13	-10.66	<0.001
Length	-0.03	0.05	-0.48	0.63

TABLE 4.5: Summary of the coefficients from the logit mixed model in Experiment 8.

4.7 Discussion

The aim of the present written production experiment was to test whether participants expected V with a higher probability after seeing a long IO in comparison with a short IO in a ditransitive sentence. I had hypothesized that participants may be expecting V after a long IO because long constituents are more salient information-wise. The main findings of Experiment 8 are: (a) participants produced V after IO only in 20% of the trials and (b) the proportion of V responses did not vary depending on IO length. In other words, results suggest that expectations in the self-paced reading task were not driven by information structure but

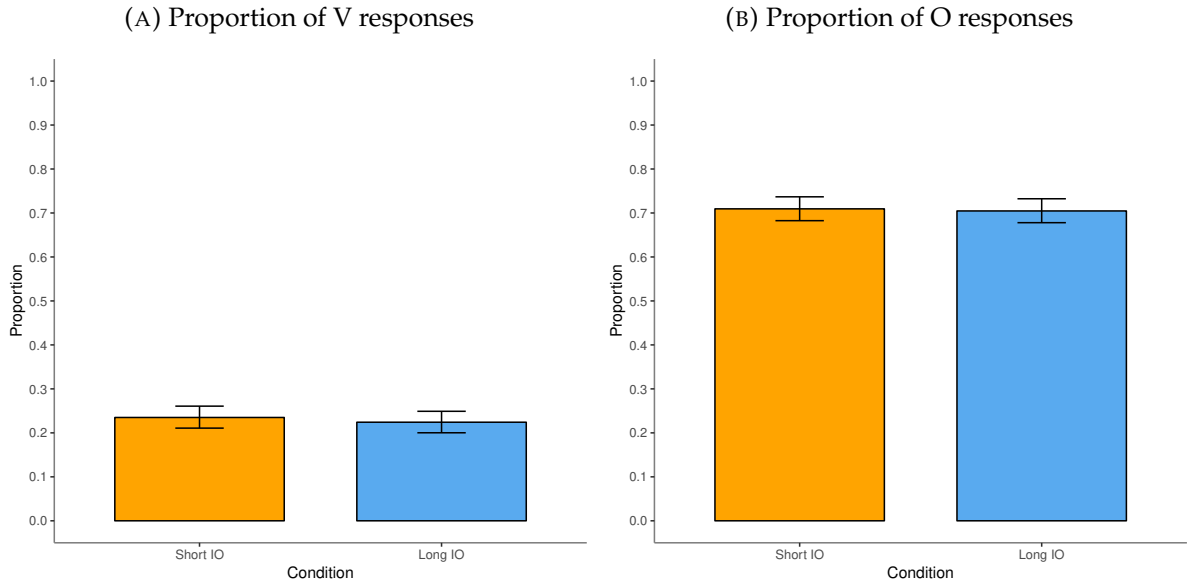


FIGURE 4.4: Proportion of V responses (left) and proportion of O responses (right) in Experiment 8. Error bars show 95% confidence intervals.

rather by the information given by the initial words of the sentence. In the long condition in Experiments 6 and 7, participants saw two sentential arguments (S and IO) at the beginning of the sentence, and they immediately knew it was a ditransitive sentence and generated a strong expectation for O that increased as extra words —the relative clause— intervened. In the short condition, however, participants only saw one sentential argument (S) at the beginning, so there is a greater level of uncertainty throughout the sentence until further verbal arguments are encountered.

Overall, results show a higher proportion of canonical word orders both when IO was long and short than in Ros et al. (2015), where the length of either IO or O (long O, long IO or all short) played an important role in the linearization of constituents. In the present study, around 70% of the productions followed canonical word order regardless of IO length. Ros and colleagues reported 67% canonical word order sentences when IO was long and O was short, 39.9% when IO was short and O was long, and 57.8% when both IO and O were short. However, it must be noted that results cannot be directly compared for a number of reasons. The present study was a written task with no time limitation. Written language is more formal and may not be directly reflecting spontaneous production patterns observed in spoken language, especially when participants could see the beginning of the sentence and think about a possible continuation for however long they wanted. Furthermore, S and IO were always the first two constituents in the sentence, so word order flexibility was highly constrained. In fact, placing the verb right after S-IO was the only option besides canonical word order. In Ros et al. (2015), participants were asked to produce sentences orally under time pressure, and memory load was increased by an arithmetic operation right before uttering the sentence. In addition, constituents could be ordered in any possible way. These characteristics may have resulted in more spontaneous productions.

5 General discussion

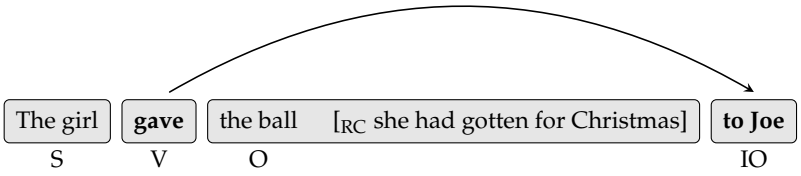
The purpose of this chapter was to test the hypothesis that antilocality effects, i.e. processing facilitation when dependency length is increased, are a general phenomenon in OV languages. Whereas distance-based processing difficulty, i.e. locality effect, is predicted by memory-based accounts of language processing (Gibson, 1998, 2000; Lewis & Vasishth, 2005), expectation-based accounts predict antilocality effects as the result of expectation-driven facilitation. That is, highly predictable continuations facilitate sentence comprehension (Hale, 2001; Levy, 2008). Locality effects have widely been reported in VO languages such as English or Spanish (see Chapter 3), but evidence of antilocality comes almost exclusively from OV languages. It has been argued that OV languages show antilocality effects because speakers are exposed to a high proportion of verb-final clauses, making the position of the verb highly predictable and facilitating language comprehension when argument-verb distance is increased. However, there are at least two issues regarding available evidence: antilocality has only been reported in strictly verb-final languages (German, Hindi and Japanese), and antilocality has mostly been found when dependency length is increased in complex structures such as double embeddings, which are known to inherently increase processing difficulty. In the present chapter, I tested the effect of increasing dependency length in a previously unstudied OV language, Basque. The main advantage of Basque is that its word order is highly flexible, allowing verb-medial clauses rather frequently.

In Experiment 6, I ran an online acceptability judgment task where participants had to rate ditransitive sentences in Basque following canonical word order (S-IO-O-V), where dependency length between IO and V was manipulated by attaching a prenominal relative clause to either IO (short dependency) or O (long dependency). I found no difference in ratings between short and long dependencies, probably because offline acceptability judgments are not sensitive enough to capture subtle processing effects in relatively easy sentences (see Experiment 2, Chapter 3). In Experiment 7, I conducted a self-paced reading task using the same manipulation as in Experiment 6 in order to investigate the impact of increasing dependency length on online sentence comprehension. Results provide evidence that long dependencies are read faster than short dependencies in main clauses in Basque, a free word order OV language. This experiment supports the hypothesis that antilocality effects are a general phenomenon in OV languages, regardless of whether they are strictly verb-final or not. In previous studies, (anti)locality effects are normally reported at V because it is the locus of argument integration. Crucially, and for the first time to the best of my knowledge, distance-based facilitation was already noticeable at a preverbal region (O), providing strong support to expectation-based accounts. In Experiment 8, an online cloze completion task was conducted in order to investigate whether the facilitation observed at O was due a higher expectation for V in medial position after a long IO (short condition) than after a short IO (long condition). Results show a low proportion of verb-medial word orders after both short and long IO, suggesting that the facilitation effect was most likely due to the increased expectations for O after seeing S and IO at the beginning of the sentence in the long condition. In summary,

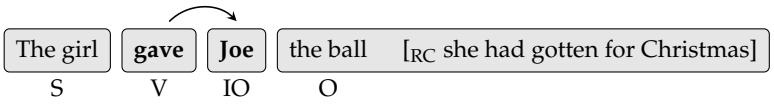
the present chapter provides evidence that antilocality effects are a general phenomenon in OV languages and are not restricted to strictly verb-final languages where the position of the verb is highly predictable.

Interestingly, a general preference for long dependencies in sentence comprehension in OV languages correlates with a weaker tendency to minimize dependencies in language production, often known as *dependency length minimization* (hereafter, DLM). It has been observed that whenever various linearizations are possible in order to express the same truth-conditional meaning, speakers tend to produce orders that minimize the distance between dependents and heads (Hawkins, 1994, 2004; Wasow, 2002; Liu, 2008; Gildea & Temperley, 2010; Futrell et al., 2015a). A seminal example of this phenomenon is the *heavy NP shift* (Ross, 1967). In English, ditransitive sentences can have two possible linearizations: S-V-O-IO or S-V-IO-O. When additional words are attached to O, as in (8a), the distance between V and IO is increased. In order to minimize dependency length, speakers tend to produce *short-before-long* orders by placing IO before O (8b).

(8) a. *Long dependency (long-before-short)*



b. *Short dependency (short-before-long)*



DLM has been attested in typologically diverse languages, regardless of head directionality. Futrell et al. (2015a) provided for the first time large-scale evidence of this preference for producing shorter dependencies in 37 languages. After parsing large dependency treebanks, they observed dependency length was generally shorter than chance, providing strong support to the idea that DLM is a universal property of human languages. Crucially, corpus studies also observe that the effect size of DLM is systematically smaller in OV languages than in VO languages (Gildea & Temperley, 2010; Futrell et al., 2015a). This difference has also been attested in cross-linguistic experimental work. Ros (2018) ran a cued recall production task in Basque, Spanish and Polish in order to investigate DLM using ditransitive sentences in typologically different languages: Basque and Polish are morphologically rich languages in comparison to Spanish, and Spanish and Polish are VO languages, whereas Basque is an OV language. Results show that the preference for DLM is stronger in Spanish (85.2%) and Polish (96.4%) than in Basque (60.3%), suggesting that head directionality plays a crucial role

in the preference for minimizing dependencies. The reason why DLM is not as strong in OV languages is not clear yet. Gildea and Temperley (2010) suggest that OV languages where the verb appears in final position cannot minimize as much because all arguments must be produced to the left of the verb, yielding longer dependencies. In VO languages, on the other hand, having the verb in the middle position allows speakers to distribute arguments both to the left and to the right of the verb, resulting in shorter dependencies. However, Basque is an OV language allowing verb-medial orders and still DLM is weaker than in VO languages, suggesting that all OV languages pattern together regardless of the degree of word order flexibility. A possible explanation is that producing non-canonical word orders is costly, so there should be a trade-off between minimizing dependencies and production difficulty.

6 Concluding remarks

Even though dependency length is a major determinant of language processing in VO languages, increasing the distance between a head and a dependent in OV languages facilitates language comprehension, most likely as a result of a high expectation for a clause-final verb. In fact, it has been hypothesized that locality effects are a general phenomenon in VO languages, whereas antilocality effects are to be found in OV languages. However, current evidence of this antilocality is restricted to strictly verb-final OV languages, such as German, Hindi and Japanese. In the present chapter, I broadened the cross-linguistic empirical basis of current psycholinguistic theories addressing both locality and antilocality effects by investigating Basque, an OV language with free word order allowing verbs in middle position. I provided evidence that longer dependencies yield faster reading times in online language comprehension in Basque. Furthermore, facilitation was already observed before the verb, providing a strong piece of evidence in favor of expectation-based accounts. Overall, I argued that antilocality is a general phenomenon in all OV languages, regardless of word order freedom, which correlates with a weaker tendency to minimize dependencies in language production.

Chapter 5

General conclusions

The present dissertation set out to investigate the memory mechanisms involved in the comprehension of syntactic dependencies in real time. The overall goal was to investigate the factors contributing to the fluctuating level of activation of memory representations, and their effect on both memory retention and retrieval processes. Specifically, I focused on (a) whether and how the representational complexity of an antecedent impacted its subsequent retrieval from memory (Chapter 2), and (b) activation decay as a function of time and its interaction with expectation-driven facilitation by investigating (i) different types of dependencies –unbounded and local– (Chapter 3) and (ii) an OV language with flexible word order (Chapter 4). For that purpose, I gathered data using a wide range of experimental techniques, comprising neurophysiological data (ERPs), reading time data from behavioral tasks, offline acceptability data, and offline production data. Furthermore, I investigated three different languages: English, Spanish and Basque. In summary, the main contributions of this dissertation are the following:

1. I provided novel evidence that encoding syntactically complex antecedents facilitates their subsequent retrieval from memory. This constitutes strong evidence that not only semantic cues contribute to increase the activation of memory representations.

2. I failed to find evidence of sustained brain activity in left anterior electrodes (i.e. the SAN component) during dependency processing, suggesting the antecedent was not actively maintained in memory. This finding is not compatible with the view that there is a memory storage cost in memory throughout the dependency. The lack of a sustained effect strongly supports retrieval models of language parsing.

3. I showed that increasing dependency length hinders language comprehension in both local and unbounded dependencies in Spanish. This is a strong validation of memory-based processing models. It also supports the hypothesis that locality is a general phenomenon in VO languages.

4. I investigated the effect of dependency length on comprehension in Basque for the first time, showing that increasing dependency length facilitates language processing. This constitutes strong evidence in support of the idea that all OV languages are subject

to expectation-driven facilitation due to a high exposure to verb-final word orders, regardless of whether they are strictly verb final or not.

1 Main findings

In Chapter 2, I investigated whether and how the representational complexity of the antecedent facilitates its subsequent retrieval in real time processing, as claimed in previous work (Hofmeister, 2007, 2011; Hofmeister & Vasishth, 2014). Researchers had argued that additional semantic processing at encoding renders memory targets more accessible in memory for retrieval operations or memory reactivation (Hofmeister & Vasishth, 2014; Karimi et al., 2014; Karimi & Ferreira, 2016). However, the role of syntactic complexity was less clear because semantic and syntactic complexity were confounded. In Experiment 1, I addressed this gap with an experimental design that manipulated semantic and syntactic complexity separately. I used neurophysiological measures (ERPs) in order to investigate the effect of representational complexity on the time course of *wh*-dependency processing. Results show that syntactic complexity, and not semantic complexity, reduced the amplitude of the P600 at the verb, an ERP component which has been argued to index syntactic integration difficulty. It is possible that the effect size of semantic complexity is smaller than that of syntactic complexity and more data is needed in order to see a difference. Assuming a cue-based retrieval mechanism, as in LV05 (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006), adding unique cues to a target could make retrieval faster because it can be more easily identified, reducing the chances of retrieving from other competitors in memory. However, only features matching the retrieval cues are relevant. I argued that the increased encoding time for syntactic complex antecedents may have led to a higher activation level in memory, making representations more accessible for retrieval operations.

An additional advantage of using ERPs in Chapter 2 is that it allowed me to investigate the time course of dependency processing. This is especially relevant for sentence processing models assuming a memory storage cost as a result of maintaining either (a) the antecedent in memory (Wanner & Maratsos, 1978; Fodor, 1978; Clifton & Frazier, 1989) or (b) a syntactic prediction. Previous research had claimed a sustained anterior negativity (SAN) on left-anterior electrodes indexes the cognitive cost of active memory maintenance (King & Kutas, 1995; Fiebach et al., 2002; Phillips et al., 2005). I failed to find a SAN throughout the dependency. The lack of a an increased memory cost during dependency processing suggests that no information has to be stored in a separate memory state, as proposed by tripartite memory models (Baddeley, 1986). I defended that this favors the view that sentence comprehension is mediated by retrieval operations that brings back encoded representations in a passive memory state to focal attention (McElree, 2000, 2006).

In Chapters 3 & 4, I focused on the effect of activation decay when the distance between co-dependents is increased and the interaction between memory decay and expectation-driven facilitation. I pointed out two problems with the current available evidence of both locality and

antilocality effects. First, most studies reporting locality focused on unbounded dependencies (Gibson, 2000; Gibson & Warren, 2004; Grodner & Gibson, 2005; Vasishth & Drenhaus, 2011), whereas studies reporting antilocality focused on local dependencies (Konieczny, 2000; Vasishth & Lewis, 2006; Nakatani & Gibson, 2008, 2010). Second, evidence of antilocality is limited to complex sentences that are inherently harder to process and to three strictly head-final languages OV languages (German, Japanese and Hindi). In Chapters 3 & 4, I broadened the empirical evidence of locality and antilocality by providing new experimental data from (a) the direct comparison of unbounded and local dependency types and (b) Basque, a previously unstudied OV language with greater word order flexibility.

In Chapter 3, I hypothesized that the difference between local and unbounded dependencies observed in previous studies may be due to differences with regards to expectations. On the one hand, local dependencies must be completed locally and, hence, the position of the verb is highly predictable. On the other, unbounded dependencies can span across more than one clause, increasing the uncertainty as to when the verb will appear. According to expectation-based accounts of sentence processing (Hale, 2001; Levy, 2008), strong expectations, as in the case of local dependencies, should cancel memory decay and facilitate processing, yielding an antilocality effect. In unbounded dependencies, however, a higher degree of uncertainty is predicted to result in a locality effect. Contrary to these predictions, I found a locality effect in both local and unbounded dependencies in Spanish in two self-paced reading tasks (Experiments 3 & 4). I argued that this is evidence that processing difficulty increases as a function of dependency length in VO languages in general, as predicted by memory-based accounts of language processing (Gibson, 1998, 2000; Lewis & Vasishth, 2005; Vasishth & Lewis, 2006).

In Chapter 4, I tested the hypothesis that antilocality is a general phenomenon in OV languages. For that purpose, I tested Basque, a previously unstudied OV language with word order flexibility, allowing verbs in medial position with a relatively high frequency. In addition, the materials consisted of simple declarative sentences in order to ensure that expectation-driven facilitation is not restricted to complex sentences. I reported the results from a self-paced reading task (Experiment 7) showing a clear facilitation effect when dependency length was increased. In addition, the facilitation effect was found already before the verb was encountered, suggesting pre-verbal arguments increased the expectation for the verb. This study adds a new language to the pool of available evidence of antilocality, showing that expectation-driven facilitation prevails in OV languages in general.

Taken together, evidence from Chapters 3 & 4 suggests that the distribution of locality and antilocality effects depends on head-directionality, as suggested in Levy and Keller (2013): VO languages are subject to locality, whereas OV languages are subject to antilocality. Interestingly, there is accumulating evidence that in OV languages (a) dependencies are longer than in VO languages, and (b) there is a weaker tendency to minimize dependency length in language production (Gildea & Temperley, 2010; Futrell et al., 2015a; Ros, 2018). In fact, in Chapter 3, I reported data from a cloze completion task (Experiment 2) in Spanish showing speakers

tend to complete both local and unbounded dependencies as soon as possible. In Chapter 4, a cloze completion task (Experiment 6) shows that Basque speakers overall tend to produce verb-final clauses, at least in written form, yielding longer dependencies. It could be the case that expectations have become more prominent in OV languages in order to cope with longer dependencies and prevent processing difficulty due to memory decay.

2 Further research

Throughout this dissertation, I have discussed that the activation level of memory representations is a key factor determining processing facilitation or difficulty (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006). In Chapter 2, I discussed that increasing the complexity of antecedents leads to a higher activation in memory which, in turn, enhances memory retrieval. This effect has been reported in previous work, as well as in this dissertation, comparing dependencies where length and intervening words were kept constant. However, it is still unclear what the strength of this effect is and how it interacts with factors constraining sentence comprehension, namely activation decay and similarity-based interference.

As reviewed in Chapter 3, increasing the linear distance between co-dependents hinder sentence comprehension, at least in VO languages such as English or Spanish. The reason underlying this effect is that the representation of the antecedent decays from memory as a function of time. Then, increasing the complexity of the antecedent should increase its activation level in memory, making the memory target more activated at retrieval overall in comparison to a less complex antecedent. That is, if the effect of representational complexity is strong enough, it should be able to mitigate locality effects in sentence comprehension. This should be addressed in future studies.

Appendix A

Resumen en español

Las palabras en una oración dependen unas de otras para su correcta interpretación sintáctica y semántica, formando así relaciones de *dependencia* entre núcleos y dependientes o antecedentes (Tesnière, 1959). La comprensión del lenguaje es incremental y las palabras nuevas tienen que ser incorporadas a la representación sintáctica sobre la marcha (Altmann & Kamide, 1999; Kaiser & Trueswell, 2004). Por tanto, la existencia de dependencias entre palabras no adyacentes supone un reto para el analizador sintáctico. Nuestro sistema de memoria debe disponer de algún mecanismo para completar dependencias no-adyacentes, bien sea mediante el mantenimiento activo de las representaciones en la memoria, o mediante la codificación de información en memoria y su posterior recuperación. En la presente tesis, investigo los mecanismos de memoria que nos permiten procesar las dependencias no-adyacentes. En concreto, investigo dos factores que afectan el nivel de activación en memoria de las representaciones y, por lo tanto, su posterior recuperación: (a) la complejidad sintáctica y semántica de las representaciones, y (b) la distancia linear entre los antecedentes y los núcleos. Para ello, he llevado a cabo una serie de experimentos, centrándome en tres lenguas tipológicamente diversas (inglés, español y euskera) y en dos tipos de dependencias (locales y no-locales). A continuación, resumo el contenido de los capítulos 2–4.

1 Resumen de los capítulos

Capítulo 2

Las expresiones referenciales pueden ser más o menos informativas. Por ejemplo, tanto *la mujer* como *la empresaria* pueden referirse a la misma persona, pero la segunda es semánticamente más rica porque es más específica. Si añadiésemos más información, como en *la empresaria de renombre*, este sintagma es más rico semánticamente y sintácticamente. Asumiendo que estos sintagmas se codifican en memoria como un conjunto de rasgos, las frases que son más informativas contienen un mayor número de rasgos. Utilizo el término *complejidad* para referirme a los sintagmas, donde x es más complejo que y cuando ambos se refieren a la misma entidad e y los rasgos de y están incluidos en x. El Capítulo 2 se centra en investigar si la

complejidad sintáctica y semántica de los antecedentes incrementa su nivel de activación en memoria, facilitando su posterior recuperación en comprensión del lenguaje.

Hay un gran número de estudios que reportan que la complejidad de una codificación en memoria influye en cómo de rápido se recupera de la memoria. En concreto, tanto la elaboración como las características distintivas han sido señaladas como factores que permiten recordar mejor ítems en aislamiento (Von Restorff, 1933; J. R. Anderson & Reder, 1979; Bradshaw & Anderson, 1982; McDaniel, 1981; McDaniel et al., 1988; Gallo et al., 2008). Este efecto es comúnmente conocido como efecto de aislamiento o *Von Restorff* (Von Restorff, 1933). El modelo de niveles de procesamiento (Craik & Lockhart, 1972) establece que los elementos que son más difíciles de codificar en memoria por contener más información, dejan una huella más duradera en la memoria. Lockhart et al. (1976) especifican que la dificultad a la hora de codificar información puede deberse a una mayor procesamiento semántico, sintáctico o fonológico. La idea que subyace a este planteamiento es que los inusual se recuerda mejor.

En los estudios que menciono en el apartado anterior, los participantes fueron advertidos de que su tarea en el experimento era recordar las palabras u oraciones que se les presentaban. Sin embargo, cuando estamos procesando una oración a tiempo real, tenemos que codificar y recuperar de forma implícita las representaciones de las palabras y frases que hemos visto u oído. Algunos estudios han encontrado que el aumentar la complejidad del antecedente en las preguntas *qu* o cláusulas de relativo hace que los tiempos de lectura en el verbo sean más rápidos (Hofmeister, 2007, 2011; Hofmeister & Vasishth, 2014). Esto se ha tomado como evidencia de que los antecedentes complejos se recuperan de la memoria más rápidamente que los simples. Se ha encontrado un efecto similar en otro tipo de dependencias lingüísticas. Por ejemplo, en el procesamiento de pronombres de tercera persona hay que reactivar el referente. En estudios utilizando diferentes metodologías se ha observado que los hablantes prefieren los referentes que son más complejos que los más simples (Karimi et al., 2014; Karimi & Ferreira, 2016; Karimi et al., 2018).

La evidencia disponible actualmente sugiere que la complejidad semántica de los antecedentes incrementa su nivel de activación en memoria. Sin embargo, el papel de la complejidad sintáctica no está claro. El principal problema en todos los estudios es que una mayor complejidad semántica va ligada a una mayor complejidad sintáctica. Por lo tanto, no es posible distinguir cuál es la contribución de los rasgos sintácticos a la activación en memoria de las representaciones lingüísticas. En el Experimento 1, llevé a cabo un estudio electrofisiológico utilizando la técnica de potenciales relacionados con el evento (ERPs, por sus siglas en inglés) con el fin de investigar la contribución individual de la complejidad semántica y sintáctica. Para ello, utilicé como materiales preguntas indirectas en inglés en las que he manipulado la complejidad de la frase *qu*: una frase *qu* simple (ej. *who*, "quién"), una frase *qu* sintácticamente más compleja (ej. *which person*, "qué persona"), y una frase *qu* sintáctica y semánticamente más compleja (*which waiter*, "qué camarero"). Las condiciones que incluyen una frase *qu* fueron comparadas con una oración de referencia sin dependencia larga.

Reporto un efecto P600 en el verbo en las tres condiciones experimentales. Este efecto ha sido encontrado en varios estudios que investigaban los correlatos neuronales de los mecanismos involucrados en el procesamiento de dependencias largas (Kaan et al., 2000; Fiebach et al., 2002; Phillips et al., 2005). En línea con estos trabajos, interpreto el P600 como un índice de dificultad de integración sintáctica, reflejando la recuperación de memoria del antecedente. Además, reporto que la amplitud de la onda del P600 es mayor para *who* que para *which person* y *which waiter*. Esto indica que la frase *qu* más simple es la que ha producido un mayor coste de integración, sugiriendo que el nivel de activación de las frases más compleja era mayor y, por tanto, se recuperan de la memoria con mayor facilidad.

También analizo la señal electrofisiológica evocada durante el transcurso de la dependencia. A diferencia de estudios anteriores (Kaan et al., 2000; Fiebach et al., 2002; Phillips et al., 2005), no encuentro una negatividad anterior sostenida (SAN, por sus siglas en inglés) durante la comprensión de dependencias *qu*. Esto puede deberse a que en estudios anteriores los materiales y/o las decisiones tomadas a la hora de procesar la señal continua. El SAN había sido relacionado con el coste cognitivo derivado de mantener activas en memoria las representaciones de los antecedentes hasta que pudieran integrarse en la representación sintáctica. Este mantenimiento en memoria es asumido en diferentes modelos de comprensión del lenguaje (Fodor, 1978; Clifton & Frazier, 1989; Gibson, 1998, 2000). La ausencia del SAN supone una validación de los modelos de comprensión del lenguaje basados en la recuperación de la memoria (Van Dyke & Lewis, 2003; Lewis et al., 2006; Van Dyke & McElree, 2006).

Capítulo 3

La distancia linear entre dos elementos en una relación de dependencia afecta a la dificultad de procesamiento. Hay un gran número de estudios que demuestran que aumentar la cantidad de material lingüístico que interviene entre un antecedente y un núcleo hace que completar una dependencia sea más costoso. Este llamado efecto de *localidad* se ha reportado principalmente en lenguas VO, como en inglés (Grodner & Gibson, 2005), chino (Hsiao & Gibson, 2003) y ruso (Levy et al., 2013). Sin embargo, en lenguas OV, como el alemán (Konieczny, 2000; Konieczny & Döring, 2003) y el hindi (Vasishth, 2003; Lewis & Vasishth, 2005), se ha encontrado un efecto de *antilocalidad*; es decir, las dependencias largas son más fáciles de procesar. En el Capítulo 3, observo que todos los estudios que han encontrado un efecto de localidad utilizan dependencias no-locales (preguntas *qu* y oraciones de relativo), mientras que los que encuentran antilocalidad utilizan dependencias locales (por ejemplo, relaciones sujeto-verbo). La diferencia fundamental entre ambas es que las no-locales pueden resolverse en una cláusula diferente, mientras que las locales deben resolver obligatoriamente en la misma cláusula. Esto afecta directamente al grado de predictabilidad del verbo: en las dependencias locales, el verbo es en principio más predecible y, según los modelos de predicción (Hale, 2001; Levy, 2008), se facilita la comprensión del lenguaje. Por primera vez, formula la hipótesis de que las diferencias en los patrones de localidad y antilocalidad se deben al tipo de dependencias utilizado.

Los modelos de procesamiento basados en memoria predicen que una mayor distancia entre dos co-dependientes dificulta la comprensión del lenguaje (Gibson, 1998, 2000; Lewis & Vasishth, 2005; Lewis et al., 2006; Vasishth & Lewis, 2006). La *Teoría de la Localidad de las Dependencias* (DLT, por sus siglas en inglés) (Gibson, 1998, 2000) establece que la comprensión de dependencias conlleva dos costes: uno de almacenamiento y otro de integración. El coste de integración varía en función del número de expresiones referenciales que intervienen, ya que el relacionar dichas expresiones con su referente consume recursos. Por lo tanto, una mayor distancia entre el antecedente y el núcleo incrementa las probabilidades de que haya expresiones referenciales que intervengan. En el modelo de procesamiento del lenguaje basado en activación (LV05) (Lewis & Vasishth, 2005; Lewis et al., 2006; Vasishth & Lewis, 2006), el nivel de activación decae en función del tiempo, por lo que una mayor distancia entre co-dependientes dificulta la comprensión. Además, si hay palabras que intervienen y que comparten rasgos con el antecedente, el nivel de activación se reparte entre ellas, generando interferencia por similitud (Gordon et al., 2001; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006; Vasishth & Lewis, 2006; Van Dyke & McElree, 2011).

Pese a las predicciones hechas por los modelos basados en memoria, algunos investigadores han sugerido que los efectos de localidad pueden ser revertidos cuando las predicciones son muy fuertes (Husain et al., 2014). Los modelos del lenguaje basados en predicciones (Hale, 2001; Levy, 2008) proponen que las transiciones entre palabras con una alta probabilidad facilitan la comprensión de las oraciones. De este modo, se ha propuesto que las lenguas OV muestran efectos de antilocalidad precisamente porque la posición del verbo es altamente predecible y, además, los argumentos preverbales contribuyen a afinar las predicciones con respecto a qué tipo de verbo es. Por esta razón, la hipótesis que testeo en el Capítulo 3 es que las dependencias locales muestran efectos de antilocalidad porque el verbo es muy predecible, ya que la dependencia debe resolverse dentro de la misma oración. Sin embargo, en las dependencias no-locales el verbo puede aparecer en otra cláusula, siendo menos predecible.

Con el objetivo de testear esta hipótesis, llevé a cabo una serie experimental comparando dependencias locales y no-locales en español en las que he manipulado la longitud (larga/corta). En el Experimento 2, recogí datos de aceptabilidad de hablantes nativos de español. Los datos no muestran ninguna diferencia entre las dependencias cortas y largas, posiblemente porque la aceptabilidad no es una medida capaz de apreciar efectos de procesamiento sutiles. En los Experimentos 3 y 4 llevé a cabo dos experimentos de lectura autoadministrada. En ambos casos encontré el mismo patrón: las dependencias largas fueron más difíciles de leer tanto en dependencias locales como no-locales. El Experimento 5 consistió en una tarea de producción escrita con el fin de obtener estimaciones de la proporción de veces que los hablantes prefieren completar las dependencias en la misma oración. Los resultados muestran que hay una fuerte preferencia por completar las dependencias dentro de la misma cláusula, tanto en dependencias locales como no-locales. Esto explica por qué no hubo diferencias entre ambos tipos de dependencias en los experimentos de lectura autoadministrada.

En resumen, los resultados de los experimentos llevados a cabo en el Capítulo 3 muestran que los efectos de localidad afectan tanto a las dependencias locales como no-locales en español. Esto supone un apoyo a los modelos basados en memoria que predicen efectos de localidad cuando aumenta la longitud de las dependencias. Además, sugiere que los efectos de localidad son un fenómeno general de las lenguas VO, y no específicos de las dependencias no-locales.

Capítulo 4

En contra de las predicciones hechas por los modelos basados en memoria, hay evidencia experimental de que las lenguas OV muestran efectos de antilocalidad (Konieczny, 2000; Konieczny & Döring, 2003; Vasishth, 2003; Lewis & Vasishth, 2005). Es decir, las dependencias largas son más fáciles de procesar que las cortas. Como se ha explicado con anterioridad, se ha sugerido que esta facilitación vienen dada por una mayor predictabilidad: en las lenguas OV, el verbo aparece al final y, por tanto, es altamente predecible. De hecho, se ha propuesto que los efectos de localidad predominan en las lenguas VO, mientras que los efectos de antilocalidad predominan en las OV (Levy & Keller, 2013). Sin embargo, la evidencia de antilocalidad disponible hoy en día está limitada por dos razones: (a) es escasa, ya que hay todavía pocos estudios; y (b) se centra exclusivamente en tres lenguas OV (alemán, hindi y japonés) donde el verbo aparece siempre o casi siempre en posición final. En el Capítulo 4, investigo si también hay efectos de antilocalidad en lenguas OV con orden libre de palabras que permiten que el verbo aparezca en posición intermedia. Para ello, me centro en el estudio del euskera.

Los estudios sobre antilocalidad se centran en el alemán (Konieczny, 2000; Konieczny & Döring, 2003), el hindi (Vasishth, 2003; Lewis & Vasishth, 2005) y el japonés (Nakatani & Gibson, 2008, 2010). En alemán, el verbo tiene que ocupar obligatoriamente la posición final en la subordinadas con *dass* ("que"). Asimismo, en oraciones declarativas, si hay un verbo compuesto, el verbo léxico aparece también en posición final. El japonés tiene un orden de palabras libre y todas las combinaciones son posibles. Sin embargo, el verbo siempre tiene que ocupar la posición final (Sasano & Okumura, 2016). En hindi, el orden de palabras también es relativamente flexible. Pese a esto, los ordenes de palabras no canónicos son raros según un estudio basado en un corpus oral de hindi (Husain et al., 2013), y las oraciones en las que el verbo no era final solo constituyen el 0,35% de los ejemplos. Además de la rigidez con respecto a la posición del verbo, los estudios que han encontrado antilocalidad en lenguas OV se centran en estructuras con una alta complejidad sintáctica.

El euskera es una lengua aislada que tiene un orden de palabras muy flexible. Esto ha sido confirmado en diferentes estudios de corpus que se centran en oraciones transitivas (De Rijk, 1969; Aldezabal et al., 2003; Pastor, 2019). En todos estos estudios, el orden más frecuente es SOV, pero SVO es también muy frecuente. Además, Futrell et al. (2015b) proporciona valores de la entropía S-O en 37 lenguas diferentes, mostrando que el euskera tiene un orden de palabras mucho más flexible (valor de entropía: ∼0,80) que el japonés (∼0,50), el alemán (∼0,50) o el hindi (∼0.30).

En el Capítulo 4, llevé a cabo una serie de experimentos utilizando oraciones declarativas ditransitivas en euskera. Se manipuló la distancia entre el IO y V, creando una dependencia corta o larga. El Experimento 6 consistió en una tarea de aceptabilidad. Los resultados no muestran ninguna diferencia entre las dependencias largas y las cortas. Al igual que en el Experimento 2, puede que esto se deba a que los juicios de aceptabilidad no puedan reflejar efectos de procesamiento sutiles. En el Experimento 7, investigué las mismas oraciones utilizando la técnica de lectura autoadministrada. Los resultados muestran que las dependencia largas produjeron tiempos de lectura más rápidos en el verbo; es decir, hubo un efecto de antilocalidad. Además, este efecto ya era significativo en el objeto preverbal. El Experimento 8 consistió en una tarea de producción escrita con el fin de investigar si la longitud del argumento preverbal influye en la preferencia por producir verbos en posición intermedia. Los resultados muestran una clara tendencia a producir oraciones con el verbo al final. Posiblemente este resultado se deba a que el lenguaje escrito es un registro más formal y no refleja los patrones de producción más espontáneos que se dan en el lenguaje oral.

En general, los experimentos del Capítulo 4 muestran que los efectos de antilocalidad son un fenómeno general en las lenguas OV, independientemente del grado de libertad en el orden de palabras. Junto con los resultados del Capítulo 3, esto sugiere que los fenómenos de localidad y antilocalidad se distribuyen en función del orden báico de palabras de una lengua, como sugieren Levy and Keller (2013). Esta preferencia por las dependencias largas en comprensión en las lenguas OV se corresponde con una tendencia más débil a minimizar dependencias en producción en comparación con las lenguas VO (Futrell et al., 2015a).

2 Principales contribuciones

En resumen, las principales contribuciones de la presente tesis son las siguientes:

1. He encontrado evidencia de que codificar antecedentes sintácticamente más complejos facilita su posterior recuperación de la memoria. Esto demuestra que no solo el añadir más carga semántica incrementa el nivel de activación en memoria de los antecedentes.

2. No he encontrado actividad cerebral sostenida durante el procesamiento de dependencias usando ERPs. Esto sugiere que el antecedente no se mantiene activamente en memoria, algo que es incompatible con los modelos que asumen un coste de almacenamiento. La ausencia de tal efecto favorece a los modelos de procesamiento basados en la recuperación de la memoria.

3. Muestro que aumentar la longitud de las dependencias dificulta la comprensión del lenguaje tanto en dependencias locales como no-locales, mostrando que los efectos de localidad son un fenómeno general en las lenguas VO. Esto constituye una validación de los modelos de procesamiento basados en la memoria.

4. He investigado el efecto de la longitud de las dependencias en euskera por primera vez. He mostrado que las dependencias largas facilitan el procesamiento. Esto muestra que en todas las lenguas OV hay una facilitación por una alta predictabilidad del verbo, independientemente de la rigidez del orden de palabras.

Appendix B

Supplementary materials to Chapter 2

1 Experimental materials used in Experiment 1

(1) a. The organizers heard which volunteer the invited speaker at the biology conference had criticized before the talk.

b. The organizers heard which person the invited speaker at the biology conference had criticized before the talk.

c. The organizers heard who the invited speaker at the biology conference had criticized before the talk.

d. The organizers heard that the invited speaker at the biology conference had criticized the volunteer before the talk.

(2) a. The spokesperson revealed which governor the presidential candidate of the liberal party had threatened after her statement.

b. The spokesperson revealed which person the presidential candidate of the liberal party had threatened after her statement.

c. The spokesperson revealed who the presidential candidate of the liberal party had threatened after her statement.

d. The spokesperson revealed that the presidential candidate of the liberal party had threatened the governor after her statement.

(3) a. The manager knew which waiter the new owner of the coffee shop would fire after the scandal.

b. The manager knew which person the new owner of the coffee shop would fire after the scandal.

c. The manager knew who the new owner of the coffee shop would fire after the scandal.

d. The manager knew that the new owner of the coffee shop would fire the waiter after the scandal.

(4) a. The commissioner confirmed which criminal the police officer at the front door had identified after the explosion.

b. The commissioner confirmed which person the police officer at the front door had identified after the explosion.

c. The commissioner confirmed who the police officer at the front door had identified after the explosion.

d. The commissioner confirmed that the police officer at the front door had identified the criminal after the explosion.

(5) a. The driver saw which teenager the old lady at the bus stop had hit with her cane.

b. The driver saw which person the old lady at the bus stop had hit with her cane.

c. The driver saw who the old lady at the bus stop had hit with her cane.

d. The driver saw that the old lady at the bus stop had hit a teenager with her cane.

(6) a. Susie was told which secretary the grant applicant with a visa issue should call for more information.

b. Susie was told which person the grant applicant with a visa issue should call for more information.

c. Susie was told who the grant applicant with a visa issue should call for more information.

d. Susie was told that the grant applicant with a visa issue should call the secretary for more information.

(7) a. The police suggested which customer the shop assistant at the liquor store should ban from buying alcohol.

b. The police suggested which person the shop assistant at the liquor store should ban from buying alcohol.

c. The police suggested who the shop assistant at the liquor store should ban from buying alcohol.

d. The police suggested that the shop assistant at the liquor store should ban a customer from buying alcohol.

(8) a. The lieutenant announced which soldier the sergeant major of the marine corps had punished for continuous disobedience.

b. The lieutenant announced which person the sergeant major of the marine corps had punished for continuous disobedience.

c. The lieutenant announced who the sergeant major of the marine corps had punished for continuous disobedience.

d. The lieutenant announced that the sergeant major of the marine corps had punished a soldier for continuous disobedience.

(9) a. The principal recommended which teacher the best student in the math class should contact for career advice.

b. The principal recommended which person the best student in the maths class should contact for career advice.

c. The principal recommended who the best student in the maths class should contact for career advice.

d. The principal recommended that the best student in the maths class should contact the teacher for career advice.

(10) a. The photographer guessed which celebrity the chief editor of the fashion magazine would prefer for the cover.

b. The photographer guessed which person the chief editor of the fashion magazine would prefer for the cover.

c. The photographer guessed who the chief editor of the fashion magazine would prefer for the cover.

d. The photographer guessed that the chief editor of the fashion magazine would prefer a celebrity for the cover.

(11) a. The neighbors remember which kid the German shepherd with a red collar had bitten for no reason.

b. The neighbors remember which person the German shepherd with a red collar had bitten for no reason.

c. The neighbors remember who the German shepherd with a red collar had bitten for no reason.

d. The neighbors remembers that the German shepherd with a red collar had bitten a kid for no reason.

(12) a. Ben already forgot which client the tour guide from the travel agency had reported for his behavior.

b. Ben already forgot which person the tour guide from the travel agency had reported for his behavior.

c. Ben already forgot who the tour guide from the travel agency had reported for his behavior.

d. Ben already forgot that the tour guide from the travel agency had reported a client for his behavior.

(13) a. The press reported which visitor the African lion from the city zoo had attacked after being fed.

b. The press reported which person the African lion from the city zoo had attacked after being fed.

c. The press reported who the African lion from the city zoo had attacked after being fed.

d. The press reported that the African lion from the city zoo had attacked a visitor after being fed.

(14) a. The statement confirmed which employee the sales manager of the department store had blamed for the robbery.

b. The statement confirmed which person the sales manager of the department store had blamed for the robbery.

c. The statement confirmed who the sales manager of the department store had blamed for the robbery.

d. The statement confirmed that the sales manager of the department store had blamed an employee for the robbery.

(15) a. Emily told me which neighbor the drunk vagabond at the main entrance had scared in the afternoon.

b. Emily told me which person the drunk vagabond at the main entrance had scared in the afternoon.

c. Emily told me who the drunk vagabond at the main entrance had scared in the afternoon.

d. Emily told me that the drunk vagabond at the main entrance had scared her neighbor in the afternoon.

(16) a. The board confirmed which teacher the blonde kid with a small scar had shoved down the stairs.

b. The board confirmed which person the blonde kid with a small scar had shoved down the stairs.

c. The board confirmed who the blonde kid with a small scar had shoved down the stairs.

d. The board confirmed that the blonde kid with a small scar had shoved the teacher down the stairs.

(17) a. Tom already knew which tourist the ripped guy in the swimming pool had dated the year before.

b. Tom already knew which person the ripped guy in the swimming pool had dated the year before.

c. Tom already knew who the ripped guy in the swimming pool had dated the year before.

d. Tom already knew that the ripped guy in the swimming pool had dated the tourist the year before.

(18) a. The recording verified which driver the security guard in the parking lot had seen from his cabin.

b. The recording verified which person the security guard in the parking lot had seen from his cabin.

c. The recording verified who the security guard in the parking lot had seen from his cabin.

d. The recording verified that the security guard in the parking lot had seen the driver from his cabin.

(19) a. The designer mentioned which doctor the Russian model with an eating disorder should visit before the show.

b. The designer mentioned which person the Russian model with an eating disorder should visit before the show.

c. The designer mentioned who the Russian model with an eating disorder should visit before the show.

d. The designer mentioned that the Russian model with an eating disorder should visit a doctor before the show.

(20) a. The reports revealed which producer the British actress in the leading role had sued for sexual harassment.

b. The reports revealed which person the British actress in the leading role had sued for sexual harassment.

c. The reports revealed who the British actress in the leading role had sued for sexual harassment.

d. The reports revealed that the British actress in the leading role had sued the producer for sexual harassment.

(21) a. The assistant mentioned which nurse the blood donor with low blood pressure should call for some water.

b. The assistant mentioned which person the blood donor with low blood pressure should call for some water.

c. The assistant mentioned who the blood donor with low blood pressure should call for some water.

d. The assistant mentioned that the blood donor with low blood pressure should call a nurse for some water.

(22) a. The CEO decided which businessman the vice president of the insurance company should meet before the announcement.

b. The CEO decided which person the vice president of the insurance company should meet before the announcement.

c. The CEO decided who the vice president of the insurance company should meet before the announcement.

d. The CEO decided that the vice president of the insurance company should meet the businessman before the announcement.

(23) a. The audience decided which singer the four judges of the talent show should choose as the winner.

b. The audience decided which person the four judges of the talent show should choose as the winner.

c. The audience decided who the four judges of the talent show should choose as the winner.

d. The audience decided that the four judges of the talent show should choose the singer as the winner.

(24) a. We all read which hiker the rescue team at the mountain shelter had rescued with a helicopter.

b. We all read which person the rescue team at the mountain shelter had rescued with a helicopter.

c. We all read who the rescue team at the mountain shelter had rescued with a helicopter.

d. We all read that the rescue team at the mountain shelter had rescued a hiker with a helicopter.

(25) a. The reporter announced which player the basketball coach of the school team had selected as the captain.

b. The reporter announced which person the basketball coach of the school team had selected as the captain.

c. The reporter announced who the basketball coach of the school team had selected as the captain.

d. The reporter announced that the basketball coach of the school team had selected a player as the captain.

(26) a. The aristocrat heard which assistant the art collector in the blue blazer had hired to run errands.

b. The aristocrat heard which person the art collector in the blue blazer had hired to attend errands.

c. The aristocrat heard who the art collector in the blue blazer had hired to attend errands.

d. The aristocrat heard that the art collector in the blue blazer had hired an assistant to run errands.

(27) a. The witness saw which teenager the fat man with a big hotdog had followed down the street.

b. The witness saw which person the fat man with a big hotdog had followed down the street.

c. The witness saw who the fat man with a big hotdog had followed down the street.

d. The witness saw that the fat man with a big hotdog had followed a teenager down the street.

(28) a. The dean determined which candidate the selection committee for the open position should interview in the morning.

b. The dean determined which person the selection committee for the open position should interview in the morning.

c. The dean determined who the selection committee for the open position should interview in the morning.

d. The dean determined that the selection committee for the open position should interview the candidate in the morning.

(29) a. The journalist discovered which politician the former leader of the labor union had accused of stealing money.

b. The journalist discovered which person the former leader of the labor union had accused of stealing money.

c. The journalist discovered who the former leader of the labor union had accused of stealing money.

d. The journalist discovered that the former leader of the labor union had accused a politician of stealing money.

(30) a. The recording confirmed which journalist the Spanish player on the tennis court had insulted before the game.

b. The recording confirmed which person the Spanish player on the tennis court had insulted before the game.

c. The recording confirmed who the Spanish player on the tennis court had insulted before the game.

d. The recording confirmed that the Spanish player on the tennis court had insulted a journalist before the game.

(31) a. The magazine published which artist the rock star at the music festival had imitated on the stage.

b. The magazine published which person the rock star at the music festival had imitated on the stage.

c. The magazine published who the rock star at the music festival had imitated on the stage.

d. The magazine published that the rock star at the music festival had imitated an artist on the stage.

(32) a. The groom discovered which bridesmaid the two groomsmen wearing a red tie had texted during the ceremony.

b. The groom discovered which person the two groomsmen wearing a red tie had texted during the ceremony.

c. The groom discovered who the two groomsmen wearing a red tie had texted during the ceremony.

d. The groom discovered that the two groomsmen wearing a red tie had texted the bridesmaid during the ceremony.

(33) a. The staff witnessed which cashier the supermarket customer carrying a yellow plastic bag protected from the robber.

b. The staff witnessed which person the supermarket customer carrying a yellow plastic bag protected from the robber.

c. The staff witnessed who the supermarket customer carrying a yellow plastic bag protected from the robber.

d. The staff witnessed that the supermarket customer carrying a yellow plastic bag protected the cashier from the robber.

(34) a. The vendor saw which musician the subway passenger holding a large coffee had pushed with her elbow.

b. The vendor saw which person the subway passenger holding a large coffee had pushed with her elbow.

c. The vendor saw who the subway passenger holding a large coffee had pushed with her elbow.

d. The vendor saw that the subway passenger holding a large coffee had pushed the musician with her elbow.

(35) a. Mark finally recalled which clerk the inexperienced cleaner mopping the tiled floor had rebuked for spilling coffee.

b. Mark finally recalled which person the inexperienced cleaner mopping the tiled floor had rebuked for spilling coffee.

c. Mark finally recalled who the inexperienced cleaner mopping the tiled floor had rebuked for spilling coffee.

d. Mark finally recalled that the inexperienced cleaner mopping the tiled floor had rebuked the clerk for spilling coffee.

(36) a. The client knew which pedestrian the taxi driver listening to classical music had saved from being injured.

b. The client knew which person the taxi driver listening to classical music had saved from being injured.

c. The client knew who the taxi driver listening to classical music had saved from being injured.

d. The client knew that the taxi driver listening to classical music had saved a pedestrian from being injured.

(37) a. The publicist announced which actor the band member playing the bass guitar had kissed at the gala.

b. The publicist announced which person the band member playing the bass guitar had kissed at the gala.

c. The publicist announced who the band member playing the bass guitar had kissed at the gala.

d. The publicist announced that the band member playing the bass guitar had kissed the actor at the gala.

(38) a. The boss revealed which intern the administrative assistant writing the final report had recommended for the job.

b. The boss revealed which person the administrative assistant writing the final report had recommended for the job.

c. The boss revealed who the administrative assistant writing the final report had recommended for the job.

d. The boss revealed that the administrative assistant writing the final report had recommended an intern for the job.

(39) a. The residents determined which electrician the building janitor fixing the wall socket should call to help him.

b. The residents determined which person the building janitor fixing the wall socket should call to help him.

c. The residents determined who the building janitor fixing the wall socket should call to help him.

d. The residents determined that the building janitor fixing the wall socket should call an electrician to help him.

(40) a. Hanna quickly deduced which bartender many regular customers at the Irish pub had known for many years.

b. Hanna quickly deduced which person many regular customers at the Irish pub had known for many years.

c. Hanna quickly deduced who many regular customers at the Irish pub had known for many years.

d. Hanna quickly deduced that many regular customers at the Irish pub had known the bartender for many years.

(41) a. The officer certified which trespasser the lean man living by the park had observed climbing the fence.

b. The officer certified which person the lean man living next the park had observed climbing the fence.

c. The officer certified who the lean man living next the park had observed climbing the fence.

d. The officer certified that the lean man living next the park had observed the trespasser climbing the fence.

(42) a. The attendees explained which photographer the tall celebrity entering the movie theater had punched in the face.

b. The attendees explained which person the tall celebrity entering the movie theater had punched in the face.

c. The attendees explained who the tall celebrity entering the movie theater had punched in the face.

d. The attendees explained that the tall celebrity entering the movie theater had punched the photographer in the face.

(43) a. The book uncovered which senator the revolutionary leader fighting for social equality had inspired for his campaign.

b. The book uncovered which person the revolutionary leader fighting for social equality had inspired for his campaign.

c. The book uncovered who the revolutionary leader fighting for social equality had inspired for his campaign.

d. The book uncovered that the revolutionary leader fighting for social equality had inspired the senator for his campaign.

(44) a. The pedestrians reported which girl the skater boy lying on the ground had defended from the attackers.

b. The pedestrians reported which person the skater boy lying on the ground had defended from the attackers.

c. The pedestrians reported who the skater boy lying on the ground had defended from the attackers.

d. The pedestrians reported that the skater boy lying on the ground had defended the girl from the attackers.

(45) a. The driver guessed which cousin the sad child waving at the bus will miss during the summer.

b. The driver guessed which person the sad child waving at the bus will miss during the summer.

c. The driver guessed who the sad child waving at the bus will miss during the summer.

d. The driver guessed that the sad child waving at the bus will miss his cousin during the summer.

(46) a. Everyone already knew which classmate the tall guy sitting on the armchair had invited to the premiere.

b. Everyone already knew which person the tall guy sitting on the armchair had invited to the premiere.

c. Everyone already knew who the tall guy sitting on the armchair had invited to the premiere.

d. Everyone already knew that the tall guy sitting on the armchair had invited a classmate to the premiere.

(47) a. The girl sensed which friend the nervous teenager chatting with his parents had disappointed at the party.

b. The girl sensed which person the nervous teenager chatting with his parents had disappointed at the party.

c. The girl sensed who the nervous teenager chatting with his parents had disappointed at the party.

d. The girl sensed that the nervous teenager chatting with his parents had disappointed his friend at the party.

(48) a. The cameraman described which hunter the TV host exploring the Amazon rainforest had seen behind a tree.

b. The cameraman described which person the TV host exploring the Amazon rainforest had seen behind a tree.

c. The cameraman described who the TV host exploring the Amazon rainforest had seen behind a tree.

d. The cameraman described that the TV host exploring the Amazon rainforest had seen a hunter behind a tree.

(49) a. The cops verified which fugitive the park gardener watering the red roses had found in a bush.

b. The cops verified which person the park gardener watering the red roses had found in a bush.

c. The cops verified who the park gardener watering the red roses had found in a bush.

d. The cops verified that the park gardener watering the red roses had found the fugitive in a bush.

(50) a. Paul accidentally mentioned which boy the French babysitter looking after the kids had rewarded for tidying up.

b. Paul accidentally mentioned which person the French babysitter looking after the kids had rewarded for tidying up.

c. Paul accidentally mentioned who the French babysitter looking after the kids had rewarded for tidying up.

d. Paul accidentally mentioned that the French babysitter looking after the kids had rewarded the boy for tidying up.

(51) a. The researcher mentioned which professor the undergraduate student sleeping on the couch had met at lunch time.

b. The researcher mentioned which person the undergraduate student sleeping on the couch had met at lunch time.

c. The researcher mentioned who the undergraduate student sleeping on the couch had met at lunch time.

d. The researcher mentioned that the undergraduate student sleeping on the couch had met the professor at lunch time.

(52) a. The manager remarked which drummer the Belgian musician drinking a cold beer had recommended for the band.

b. The manager remarked which person the Belgian musician drinking a cold beer had recommended for the band.

c. The manager remarked who the Belgian musician drinking a cold beer had recommended for the band.

d. The manager remarked that the Belgian musician drinking a cold beer had recommended the drummer for the band.

(53) a. The authorities knew which visitor the zoo carer feeding the panda bears had warned for touching animals.

b. The authorities knew which person the zoo carer feeding the panda bears had warned for touching animals.

c. The authorities knew who the zoo carer feeding the panda bears had warned for touching animals.

d. The authorities knew that the zoo carer feeding the panda bears had warned a visitor for touching animals.

(54) a. The scholar recommended which academic many contemporary politicians talking about climate change should consult to check facts.

b. The scholar recommended which person many contemporary politicians talking about climate change should consult to check facts.

c. The scholar recommended who many contemporary politicians talking about climate change should consult to check facts.

d. The scholar recommended that many contemporary politicians talking about climate change should consult an academic to check facts.

(55) a. Laura could sense which undergraduate the emeritus professor giving a talk on art encouraged to be creative.

b. Laura could sense which person the emeritus professor giving a talk on art encouraged to be creative.

c. Laura could sense who the emeritus professor giving a talk on art encouraged to be creative.

d. Laura could sense that the emeritus professor giving a talk on art encouraged the undergraduate to be creative.

(56) a. The fairytale described which princess the evil queen screaming from the balcony had poisoned before her wedding.

b. The fairytale described which person the evil queen screaming from the balcony had poisoned before her wedding.

c. The fairytale described who the evil queen screaming from the balcony had poisoned before her wedding.

d. The fairytale described that the evil queen screaming from the balcony had poisoned the princess before her wedding.

(57) a. The director forgot which cook the film producer talking on the phone had employed to cater food.

b. The director forgot which person the film producer talking on the phone had employed to cater food.

c. The director forgot who the film producer talking on the phone had employed to cater food.

d. The director forgot that the film producer talking on the phone had employed a cook to cater food.

(58) a. The clown witnessed which spectator the agile acrobat hanging from the trapeze warmly greeted after the show.

b. The clown witnessed which person the agile acrobat hanging from the trapeze warmly greeted after the show.

c. The clown witnessed who the agile acrobat hanging from the trapeze warmly greeted after the show.

d. The clown witnessed that the agile acrobat hanging from the trapeze warmly greeted the spectator after the show.

(59) a. Aaron warily underlined which speaker the graduate student knitting a red scarf had distracted at the seminar.

b. Aaron warily underlined which person the graduate student knitting a red scarf had distracted at the seminar.

c. Aaron warily underlined who the graduate student knitting a red scarf had distracted at the seminar.

d. Aaron warily underlined that the graduate student knitting a red scarf had distracted the speaker at the seminar.

(60) a. The reporter corroborated which protester the conservative judge leading the riot case had imprisoned for high treason.

b. The reporter corroborated which person the conservative judge leading the riot case had imprisoned for high treason.

c. The reporter corroborated who the conservative judge leading the riot case had imprisoned for high treason.

d. The reporter corroborated that the conservative judge leading the riot case had imprisoned a protester for high treason.

(61) a. The entrepeneur recommended which concierge the stock brokers attending the business meeting should tip at the hotel.

b. The entrepeneur recommended which person the stock brokers attending the business meeting should tip at the hotel.

c. The entrepeneur recommended who the stock brokers attending the business meeting should tip at the hotel.

d. The entrepeneur recommended that the stock brokers attending the business meeting should tip the concierge at the hotel.

(62) a. The viewers noticed which quaterback the obsessed fan yelling at the referee passionately applauded for his strategy.

b. The viewers noticed which person the obsessed fan yelling at the referee passionately applauded for his strategy.

c. The viewers noticed who the obsessed fan yelling at the referee passionately applauded for his strategy.

d. The viewers noticed that the obsessed fan yelling at the referee passionately applauded the quaterback for his strategy.

(63) a. The blacksmith heard which peasant the ancient vampire craving for fresh blood had bitten in the neck.

b. The blacksmith heard which person the ancient vampire craving for fresh blood had bitten in the neck.

c. The blacksmith heard who the ancient vampire craving for fresh blood had bitten in the neck.

d. The blacksmith heard that the ancient vampire craving for fresh blood had bitten the peasant in the neck.

(64) a. Lara doesn't remember which stylist the Italian baker growing a thick beard recently visited for the opening.

b. Lara doesn't remember which person the Italian baker growing a thick beard recently visited for the opening.

c. Lara doesn't remember who the Italian baker growing a thick beard recently visited for the opening.

d. Lara doesn't remember that the Italian baker growing a thick beard recently visited a stylist for the opening.

(65) a. The participants witnessed which technician the renowned expert teaching the computer course had helped with the projector.

b. The participants witnessed which person the renowned expert teaching the computer course had helped with the projector.

c. The participants witnessed who the renowned expert teaching the computer course had helped with the projector.

d. The participants witnessed that the renowned expert teaching the computer course had helped the technician with the projector.

(66) a. The receptionist reported which reporter most American socialites staying at the hotel had identified in the lobby.

b. The receptionist reported which person most American socialites staying at the hotel had identified in the lobby.

c. The receptionist reported who most American socialites staying at the hotel had identified in the lobby.

d. The receptionist reported that most American socialites staying at the hotel had identified the reporter in the lobby.

(67) a. The auctioneer verified which buyer the antique collector selling his old comics had contacted after the auction.

b. The auctioneer verified which person the antique collector selling his old comics had contacted after the auction.

c. The auctioneer verified who the antique collector selling his old comics had contacted after the auction.

d. The auctioneer verified that the antique collector selling his old comics had contacted the buyer after the auction.

(68) a. Tyler finally confessed which painter the anonymous artist engraving on brick walls had influenced throughout her career.

b. Tyler finally confessed which person the anonymous artist engraving on brick walls had influenced throughout her career.

c. Tyler finally confessed who the anonymous artist engraving on brick walls had influenced throughout her career.

d. Tyler finally confessed that the anonymous artist engraving on brick walls had influenced the painter throughout her career.

(69) a. Bethany doesn't recall which latecomer the helpful usher holding a green flashlight had guided across the theater.

b. Bethany doesn't recall which person the helpful usher holding a green flashlight had guided across the theater.

c. Bethany doesn't recall who the helpful usher holding a green flashlight had guided across the theater.

d. Bethany doesn't recall that the helpful usher holding a green flashlight had guided the latecomer across the theater.

(70) a. The newspaper published which banker the FBI officer investigating the corruption scandal had arrested for hiding evidence.

b. The newspaper published which person the FBI officer investigating the corruption scandal had arrested for hiding evidence.

c. The newspaper published who the FBI officer investigating the corruption scandal had arrested for hiding evidence.

d. The newspaper published that the FBI officer investigating the corruption scandal had arrested the banker for hiding evidence.

(71) a. The kid explained which dog the moving shadow of the willow tree had scared during the storm.

b. The kid explained which person the moving shadow of the willow tree had scared during the storm.

c. The kid explained who the moving shadow of the willow tree had scared during the storm.

d. The kid explained that the moving shadow of the willow tree had scared the dog during the storm.

(72) a. The organizers imagine which rapper most teenage girls waiting for the concert deeply love more than anything.

b. The organizers imagine which person most teenage girls waiting for the concert deeply love more than anything.

c. The organizers imagine who most teenage girls waiting for the concert deeply love more than anything.

d. The organizers imagine that most teenage girls waiting for the concert deeply love the rapper more than anything.

(73) a. The statement confirmed which employee the executive manager smoking a Cuban cigar had dismissed after the meeting.

b. The statement confirmed which person the executive manager smoking a Cuban cigar had dismissed after the meeting.

c. The statement confirmed who the executive manager smoking a Cuban cigar had dismissed after the meeting.

d. The statement confirmed that the executive manager smoking a Cuban cigar had dismissed an employee after the meeting.

(74) a. The teacher discovered which scientist the enthusiastic student answering all the questions had stopped in the street.

b. The teacher discovered which person the enthusiastic student answering all the questions had stopped in the street.

c. The teacher discovered who the enthusiastic student answering all the questions had stopped in the street.

d. The teacher discovered that the enthusiastic student answering all the questions had stopped the scientist in the street.

(75) a. Colin convincingly demonstrated which follower the religious guru giving a long sermon had humiliated in the commune.

b. Colin convincingly demonstrated which person the religious guru giving a long sermon had humiliated in the commune.

c. Colin convincingly demonstrated who the religious guru giving a long sermon had humiliated in the commune.

d. Colin convincingly demonstrated that the religious guru giving a long sermon had humiliated the follower in the commune.

(76) a. The host knew which contestant the jury member of the cooking competition had favored in the finale.

b. The host knew which person the jury member of the cooking competition had favored in the final.

c. The host knew who the jury member of the cooking competition had favored in the final.

d. The host knew that the jury member of the cooking competition had favored a contestant in the final.

(77) a. The sheriff remembered which burglar the diligent cop nearing retiring age had previously chased for similar reasons.

b. The sheriff remembered which person the diligent cop nearing retiring age had previously chased for similar reasons.

c. The sheriff remembered who the diligent cop nearing retiring age had previously chased for similar reasons.

d. The sheriff remembered that the diligent cop nearing retiring age had previously chased a burglar for similar reasons.

(78) a. The orderly saw which patient the nursing assistant carrying a newborn baby had congratulated in the hall.

b. The orderly saw which person the nursing assistant carrying a newborn baby had congratulated in the hall.

c. The orderly saw who the nursing assistant carrying a newborn baby had congratulated in the hall.

d. The orderly saw that the nursing assistant carrying a newborn baby had congratulated a patient in the hall.

(79) a. The passengers noticed which stewardess the plane captain flying to South America greatly appreciates for her patience.

b. The passengers noticed which person the plane captain flying to South America greatly appreciates for her patience.

c. The passengers noticed who the plane captain flying to South America greatly appreciates for her patience.

d. The passengers noticed that the plane captain flying to South America greatly appreciates the stewardess for her patience.

(80) a. The committee suggested which apprentice the company manager training the new members should monitor for a month.

b. The committee suggested which person the company manager training the new members should monitor for a month.

c. The committee suggested who the company manager training the new members should monitor for a month.

d. The committee suggested that the company manager training the new members should monitor the apprentice for a month.

(81) a. The visitors noted which oprhan the charity volunteers cooking a delicious stew absolutely adore for his positivity.

b. The visitors noted which person the charity volunteers cooking a delicious stew absolutely adore for his positivity.

c. The visitors noted who the charity volunteers cooking a delicious stew absolutely adore for his positivity.

d. The visitors noted that the charity volunteers cooking a delicious stew absolutely adore the orphan for his positivity.

(82) a. The study determined which criminal many robbery victims visiting the prison regulary had forgiven for his actions.

b. The study determined which person many robbery victims visiting the prison regulary had forgiven for his actions.

c. The study determined who many robbery victims visiting the prison regulary had forgiven for his actions.

d. The study determined that many robbery victims visiting the prison regulary had forgiven the criminal for his actions.

(83) a. The historian demonstrated which general the powerful emperor expanding his vast domains had defeated in the battle.

b. The historian demonstrated which person the powerful emperor expanding his vast domains had defeated in the battle.

c. The historian demonstrated who the powerful emperor expanding his vast domains had defeated in the battle.

d. The historian demonstrated that the powerful emperor expanding his vast domains had defeated the general in the battle.

(84) a. The ambassador explained which activist the radical terrorist destroying Roman ancient monuments had kidnapped during the night.

b. The ambassador explained which person the radical terrorist destroying Roman ancient monuments had kidnapped during the night.

c. The ambassador explained who the radical terrorist destroying Roman ancient monuments had kidnapped during the night.

d. The ambassador explained that the radical terrorist destroying Roman ancient monuments had kidnapped an activist during the night.

(85) a. Mark was told which designer the mechanical engineers constructing a flying car had acknowledged for her contribution.

b. Mark was told which person the mechanical engineers designing a flying car had chidden for the delay.

c. Mark was told who the mechanical engineers designing a flying car had chidden for the delay.

d. Mark was told that the mechanical engineers designing a flying car had chidden the designer for the delay.

(86) a. The gardener witnessed which girl the small toddler flying a dragon kite had pushed at the park.

b. The gardener witnessed which person the small toddler flying a dragon kite had pushed at the park.

c. The gardener witnessed who the small toddler flying a dragon kite had pushed at the park.

d. The gardener witnessed that the small toddler flying a dragon kite had pushed the girl at the park.

(87) a. The CEO reported which associate the review board examining the difficult incident had expelled for unruly behavior.

b. The CEO reported which person the review board examining the difficult incident had expelled for unruly behavior.

c. The CEO reported who the review board examining the difficult incident had expelled for unruly behavior.

d. The CEO reported that the review board examining the difficult incident had expelled the associate for unruly behavior.

(88) a. The assistant remembered which researcher the university librarian hiding behind the bookcase had scared some time ago.

b. The assistant remembered which person the university librarian hiding behind the bookcase had scared some time ago.

c. The assistant remembered who the university librarian hiding behind the bookcase had scared some time ago.

d. The assistant remembered that the university librarian hiding behind the bookcase had scared the researcher some time ago.

(89) a. The judge announced which witness the criminal attorney working for the defendent had accused of hiding evidence.

b. The judge announced which person the criminal attorney working for the defendent had accused of hiding evidence.

c. The judge announced who the criminal attorney working for the defendent had accused of hiding evidence.

d. The judge announced that the criminal attorney working for the defendent had accused the witness of hiding evidence.

(90) a. The mayor reported which gardener the giant otter eating the park plants had injured with its claws.

b. The mayor reported which person the giant otter eating the park plants had injured with its claws.

c. The mayor reported who the giant otter eating the park plants had injured with its claws.

d. The mayor reported that the giant otter eating the park plants had injured the gardener with its claws.

(91) a. The survey revealed which comedian most middle-age citizens in favor of guns deeply hate for his jokes.

b. The survey revealed which person most middle-age citizens in favor of guns deeply hate for his jokes.

c. The survey revealed who most middle-age citizens in favor of guns deeply hate for his jokes.

d. The survey revealed that most middle-age citizens in favor of guns deeply hate the comedian for his jokes.

(92) a. Ellen subtly suggested which agent some amateur actors performing at the theater should call to manage them.

b. Ellen subtly suggested which person some amateur actors performing at the theater should call to manage them.

c. Ellen subtly suggested who some amateur actors performing at the theater should call to manage them.

d. Ellen subtly suggested that some amateur actors performing at the theater should call an agent to manage them.

(93) a. The director recognized which critic the marketing manager presenting the latest device had greeted at the event.

b. The director recognized which person the marketing manager presenting the latest device had greeted at the event.

c. The director recognized who the marketing manager presenting the latest device had greeted at the event.

d. The director recognized that the marketing manager presenting the latest device had greeted the critic at the event.

(94) a. Jeff didn't see which royal the city authorities opening the art exhibition had welcomed to the museum.

b. Jeff didn't see which person the city authorities opening the art exhibition had welcomed to the museum.

c. Jeff didn't see who the city authorities opening the art exhibition had welcomed to the museum.

d. Jeff didn't see that the city authorities opening the art exhibition had welcomed the royal to the museum.

(95) a. The interviewer highlighted which director the war veteran writing an autobiographical book had inspired to make films.

b. The interviewer highlighted which person the war veteran writing an autobiographical book had inspired to make films.

c. The interviewer highlighted who the war veteran writing an autobiographical book had inspired to make films.

d. The interviewer highlighted that the war veteran writing an autobiographical book had inspired a director to make films.

(96) a. Maggie could hear which actor the assistant director removing the stage props had thanked for his performance.

b. Maggie could hear which person the assistant director removing the stage props had thanked for his performance.

c. Maggie could hear who the assistant director removing the stage props had thanked for his performance.

d. Maggie could hear that the assistant director removing the stage props had thanked the actor for his performance.

(97) a. The traveler noticed which waitress the German tourist resting underneath a tree really likes for her smile.

b. The traveler noticed which person the German tourist resting underneath a tree really likes for her smile.

c. The traveler noticed who the German tourist resting underneath a tree really likes for her smile.

d. The traveler noticed that the German tourist resting underneath a tree really likes the waitress for her smile.

(98) a. The paparrazi uncovered which refugee the rap artist singing on national television had hosted in her mansion.

b. The paparrazi uncovered which person the rap artist singing on national television had hosted in her mansion.

c. The paparrazi uncovered who the rap artist singing on national television had hosted in her mansion.

d. The paparrazi uncovered that the rap artist singing on national television had hosted a refugee in her mansion.

(99) a. The recording revealed which hostage the bank robber speaking with an accent had shot in his leg.

b. The recording revealed which person the bank robber speaking with an accent had shot in his leg.

c. The recording revealed who the bank robber speaking with an accent had shot in his leg.

d. The recording revealed that the bank robber speaking with an accent had shot the hostage in his leg.

(100) a. The tale describes which pirate the lonely outcast living on the island kindly guided to the treasure.

b. The tale describes which person the lonely outcast living on the island kindly guided to the treasure.

c. The tale describes who the lonely outcast living on the island kindly guided to the treasure.

d. The tale describes that the lonely outcast living on the island kindly guided the pirated to the treasure.

(101) a. Tony finally confessed which architect the British developer building the new airport utterly admires for his creativity.

b. Tony finally confessed which person the British developer building the new airport utterly admires for his creativity.

c. Tony finally confessed who the British developer building the new airport utterly admires for his creativity.

d. Tony finally confessed that the British developer building the new airport utterly admires the architect for his creativity.

(102) a. Diane kindly suggested which expert the museum guide explaining American modern history should email to avoid errors.

b. Diane kindly suggested which person the museum guide explaining American modern history should email to avoid errors.

c. Diane kindly suggested who the museum guide explaining American modern history should email to avoid errors.

d. Diane kindly suggested that the museum guide explaining American modern history should email an expert to avoid errors.

(103) a. The landlord guessed which roommate the downstairs tenant complaining about the WiFi had avoided in the morning.

b. The landlord guessed which person the downstairs tenant complaining about the WiFi had avoided in the morning.

c. The landlord guessed who the downstairs tenant complaining about the WiFi had avoided in the morning.

d. The landlord guessed that the downstairs tenant complaining about the WiFi had avoided his roommate in the morning.

(104) a. Dan suddenly realized which cousin the event planner organizing the family reunion didn't invite to the dinner.

b. Dan suddenly realized which person the family member organizing the family reunion didn't invite to the dinner.

c. Dan suddenly realized who the family member organizing the family reunion didn't invite to the dinner.

d. Dan suddenly realized that the family member organizing the family reunion didn't invite our cousin to the dinner.

(105) a. The spectator guessed which runner the Jamaican athlete competing for the medal seriously intimidates for her speed.

b. The spectator guessed which person the Jamaican athlete competing for the medal seriously intimidates for her speed.

c. The spectator guessed who the Jamaican athlete competing for the medal seriously intimidates for her speed.

d. The spectator guessed that the Jamaican athlete competing for the medal seriously intimidates the runner for her speed.

(106) a. The rescuers announced which child the tired hiker looking for wild mushrooms had found in the forest.

b. The rescuers announced which person the tired hiker looking for wild mushrooms had found in the forest.

c. The rescuers announced who the tired hiker looking for wild mushrooms had found in the forest.

d. The rescuers announced that the tired hiker looking for wild mushrooms had found the child in the forest.

(107) a. Rachel silently witnessed which lawmaker the pacifist activist leading the massive protest had opposed in the debate.

b. Rachel silently witnessed which person the pacifist activist leading the massive protest had opposed in the debate.

c. Rachel silently witnessed who the pacifist activist leading the massive protest had opposed in the debate.

d. Rachel silently witnessed that the pacifist activist leading the massive protest had opposed the lawmaker in the debate.

(108) a. Monica could prove which colleague the TV host presenting the quiz show had texted before going live.

b. Monica could prove which person the TV host presenting the quiz show had texted before going live.

c. Monica could prove who the TV host presenting the quiz show had texted before going live.

d. Monica could prove that the TV host presenting the quiz show had texted a colleague before going live.

(109) a. Joey was told which bellhop the lovely couple enjoying their romantic honeymoon finally convinced to take pictures.

b. Joey was told which person the lovely couple enjoying their honey moon finally convinced to take pictures.

c. Joey was told who the lovely couple enjoying their honey moon finally convinced to take pictures.

d. Joey was told that the lovely couple enjoying their honey moon finally convinced the bellhop to take pictures.

(110) a. The doctor mentioned which specialist many hospitalized patients recovering from an accident may need to walk again.

b. The doctor mentioned which person many hospitalized patients recovering from an accident may need to walk again.

c. The doctor mentioned who many hospitalized patients recovering from an accident may need to walk again.

d. The doctor mentioned that many hospitalized patients recovering from an accident may need a specialist to walk again.

(111) a. The cook discovered which client the delivery person working the night shift had insulted for not tipping.

b. The cook discovered which person the delivery person working the night shift had insulted for not tipping.

c. The cook discovered who the delivery person working the night shift had insulted for not tipping.

d. The cook discovered that the delivery person working the night shift had insulted a client for not tipping.

(112) a. The will revealed which son the proud owner of the baseball team had disinherited for his comments.

b. The will revealed which person the proud owner of the baseball team had disinherited for his comments.

c. The will revealed who the proud owner of the baseball team had disinherited for his comments.

d. The will revealed that the proud owner of the baseball team had disinherited his son for his comments.

(113) a. Charles thoroughly explained which guard the warehouse assistant tagging the incoming products had heard open the safe.

b. Charles thoroughly explained which person the warehouse assistant tagging the incoming products had heard open the safe.

c. Charles thoroughly explained who the warehouse assistant tagging the incoming products had heard open the safe.

d. Charles thoroughly explained that the warehouse assistant tagging the incoming products had heard the guard open the safe.

(114) a. The review underlined which king the novel writer signing dozens of books had included as a character.

b. The review underlined which person the novel writer signing dozens of books had included as a character.

c. The review underlined who the novel writer signing dozens of books had included as a character.

d. The review underlined that the novel writer signing dozens of books had included the King as a character.

(115) a. The columnist remembered which congressman the talk-show guest discussing European political affairs had supported in the election.

b. The columnist remembered which person the talk-show guest discussing European political affairs had supported in the election.

c. The columnist remembered who the talk-show guest discussing European political affairs had supported in the election.

d. The columnist remembered that the talk-show guest discussing European political affairs had supported the congressman in the election.

(116) a. The children guessed which guest the gentle butler looking after the mansion had killed in the movie.

b. The children guessed which person the gentle butler looking after the mansion had killed in the movie.

c. The children guessed who the gentle butler looking after the mansion had killed in the movie.

d. The children guessed that the gentle butler looking after the mansion had killed the guest in the movie.

(117) a. The sign specifies which janitor the last person leaving the conference room must tell to avoid robberies.

b. The sign specifies which person the last person leaving the conference room must tell to avoid robberies.

c. The sign specifies who the last person leaving the conference room must tell to avoid robberies.

d. The sign specifies that the last person leaving the conference room must tell the janitor to avoid robberies.

(118) a. Luke already knew which engineer the HR director selecting the potential candidates had wanted on her team.

b. Luke already knew which person the HR director selecting the potential candidates had wanted on her team.

c. Luke already knew who the HR director selecting the potential candidates had wanted on her team.

d. Luke already knew that the HR director selecting the potential candidates had wanted an engineer on her team.

(119) a. Rob vaguely mentioned which mailman the security guard blocking the main access had warned after the gunfire.

b. Rob vaguely mentioned which person the security guard blocking the main access had warned after the gunfire.

c. Rob vaguely mentioned who the security guard blocking the main access had warned after the gunfire.

d. Rob vaguely mentioned that the security guard blocking the main access had warned the mailman after the gunfire.

(120) a. The media reported which noble the country singer hosting the philanthropic event had encouraged to donate money.

b. The media reported which person the country singer hosting the philanthropic event had encouraged to donate money.

c. The media reported who the country singer hosting the philanthropic event had encouraged to donate money.

d. The media reported that the country singer hosting the philanthropic event had encouraged the noble to donate money.

Appendix C

Supplementary materials to Chapter 3

1 Experimental materials used in Experiments 2 & 3

(1) a. A la abogada (que buscaba el secretario judicial) le han entregado la documentación (que buscaba el secretario judicial) para que la archive.

b. ¿A qué abogada (que buscaba el secretario judicial) le han entregado la documentación (que buscaba el secretario judicial) para que la archive?

(2) a. A la chica (que registró la policía secreta) le han confiscado la mochila (que registró la policía secreta) porque puede ser peligrosa.

b. ¿A qué chica (que registró la policía secreta) le han confiscado la mochila (que registró la policía secreta) porque puede ser peligrosa?

(3) a. A la residente (que supervisó el médico adjunto) le han enviado el informe (que supervisó el médico adjunto) para que lo revise.

b. ¿A qué residente (que supervisó el médico adjunto) le han enviado el informe (que supervisó el médico adjunto) para que lo revise?

(4) a. A la catedrática (que evaluó el comité científico) le han concedido el proyecto (que evaluó el comité científico) para investigar el cáncer.

b. ¿A qué catedrática (que evaluó el comité científico) le han concedido el proyecto (que evaluó el comité científico) para investigar el cáncer?

(5) a. A la investigadora (que premió la Universidad Complutense) le han publicado el estudio (que premió la Universidad Complutense) en una revista importante.

b. ¿A qué investigadora (que premió la Universidad Complutense) le han publicado el estudio (que premió la Universidad Complutense) en una revista importante?

(6) a. A la ejecutiva (que apoya la junta directiva) le han rechazado la propuesta (que apoya la junta directiva) en la última votación.

b. ¿A qué ejecutiva (que apoya la junta directiva) le han rechazado la propuesta (que apoya la junta directiva) en la última votación?

(7) a. A la cantante (que critican las asociaciones feministas) le han boicoteado el disco (que critican las asociaciones feministas) por sus letras machistas.

b. ¿A qué cantante (que critican las asociaciones feministas) le han boicoteado el disco (que critican las asociaciones feministas) por sus letras machistas?

(8) a. A la concursante (que eligieron todos los espectadores) le han hecho la pregunta (que eligieron todos los espectadores) para llevarse el bote.

b. ¿A qué concursante (que eligieron todos los espectadores) le han hecho la pregunta (que eligieron todos los espectadores) para llevarse el bote?

(9) a. A la ministra (que criticó duramente la oposición) le han publicado el libro (que criticó duramente la oposición) después de mucha polémica.

b. ¿A qué ministra (que criticó duramente la oposición) le han publicado el libro (que criticó duramente la oposición) después de mucha polémica?

(10) a. A la escritora (que elogian los críticos literarios) le han censurado la novela (que elogian los críticos literarios) porque habla de corrupción.

b. ¿A qué escritora (que elogian los críticos literarios) le han censurado la novela (que elogian los críticos literarios) porque habla de corrupción?

(11) a. A la pasajera (que retuvo el guarda jurado) le han perdido la maleta (que retuvo el guarda jurado) al llegar al aeropuerto.

b. ¿A qué pasajera (que retuvo el guarda jurado) le han perdido la maleta (que retuvo el guarda jurado) al llegar al aeropuerto?

(12) a. A la fugitiva (que buscaba la Guardia Civil) le han quitado el botín (que buscaba la Guardia Civil) cuando cruzaba la frontera.

b. ¿A qué fugitiva (que buscaba la Guardia Civil) le han quitado el botín (que buscaba la Guardia Civil) cuando cruzaba la frontera?

(13) a. A la famosa (que admiran todos los adolescentes) le han cancelado la serie (que admiran todos los adolescentes) por no tener audiencia.

b. ¿A qué famosa (que admiran todos los adolescentes) le han cancelado la serie (que admiran todos los adolescentes) por no tener audiencia?

(14) a. A la periodista (que temen todos los políticos) le han dado el programa (que temen todos los políticos) por su gran profesionalidad.

b. ¿A qué periodista (que temen todos los políticos) le han dado el programa (que temen todos los políticos) por su gran profesionalidad?

(15) a. A la señora (que visitaron los trabajadores sociales) le han embargado el piso (que visitaron los trabajadores sociales) porque no tiene recursos.

b. ¿A qué señora (que visitaron los trabajadores sociales) le han embargado el piso (que visitaron los trabajadores sociales) porque no tiene recursos?

(16) a. A la modelo (que querían muchas agencias internacionales) le han pospuesto el evento (que querían muchas agencias internacionales) por el mal tiempo.

b. ¿A qué modelo (que querían muchas agencias internacionales) le han pospuesto el evento (que querían muchas agencias internacionales) por el mal tiempo?

(17) a. A la cocinera (que elogian los críticos culinarios) le han copiado el postre (que elogian los críticos culinarios) sin que diera permiso.

b. ¿A qué cocinera (que elogian los críticos culinarios) le han copiado el postre (que elogian los críticos culinarios) sin que diera permiso?

(18) a. A la administrativa (que investiga la policía anticorrupción) le han dado el contrato (que investiga la policía anticorrupción) para que lo oculte.

b. ¿A qué administrativa (que investiga la policía anticorrupción) le han dado el contrato (que investiga la policía anticorrupción) para que lo oculte?

(19) a. A la candidata (que proponen los principales sindicatos) le han criticado el proyecto (que proponen los principales sindicatos) antes de poder presentarlo.

b. ¿A qué candidata (que proponen los principales sindicatos) le han criticado el proyecto (que proponen los principales sindicatos) antes de poder presentarlo?

(20) a. A la dependienta (que quieren todos los trabajadores) le han ofrecido el puesto (que quieren todos los trabajadores) después de tres años.

b. ¿A qué dependienta que quieren todos los trabajadores le han ofrecido el puesto después de tres años?

(21) a. A la empresaria (que boicotearon las grandes empresas) le han enseñado el plan (que boicotearon las grandes empresas) para que lo evalúe.

b. ¿A qué empresaria (que boicotearon las grandes empresas) le han enseñado el plan (que boicotearon las grandes empresas) para que lo evalúe?

(22) a. A la enfermera (que recomendó el equipo médico) le han enseñado el tratamiento (que recomendó el equipo médico) para tratar dolores musculares.

b. ¿A qué enfermera (que recomendó el equipo médico) le han enseñado el tratamiento (que recomendó el equipo médico) para tratar dolores musculares?

(23) a. A la guía (que contrataron los turistas italianos) le han cancelado el tour (que contrataron los turistas italianos) porque va a nevar.

b. ¿A qué guía (que contrataron los turistas italianos) le han cancelado el tour (que contrataron los turistas italianos) porque va a nevar?

(24) a. A la arquitecta (que contrató la promotora inmobiliaria) le han parado la obra (que contrató la promotora inmobiliaria) por no tener permiso.

b. ¿A qué arquitecta (que contrató la promotora inmobiliaria) le han parado la obra (que contrató la promotora inmobiliaria) por no tener permiso?

2 Experimental materials used in Experiment 4

(1) a. El funcionario (que buscaba el secretario judicial) ha preguntado si la abogada (que buscaba el secretario judicial) ha entregado la documentación dentro del plazo establecido.

b. El funcionario (que buscaba el secretario judicial) ha preguntado que qué abogada (que buscaba el secretario judicial) ha entregado la documentación dentro del plazo establecido.

(2) a. El pasajero (que paró la policía secreta) ha preguntado si la chica (que paró la policía secreta) ha escondido la mochila dentro de la cisterna.

b. El pasajero (que paró la policía secreta) ha preguntado que qué chica (que paró la policía secreta) ha escondido la mochila dentro de la cisterna.

(3) a. El enfermero (que supervisó el médico adjunto) ha preguntado si la residente (que supervisó el médico adjunto) ha enviado el informe para que lo revisen.

b. El enfermero (que supervisó el médico adjunto) ha preguntado que qué residente (que supervisó el médico adjunto) ha enviado el informe para que lo revisen.

(4) a. El investigador (que evaluó el comité científico) ha preguntado si la catedrática (que evaluó el comité científico) ha conseguido el proyecto para investigar el cáncer.

b. El investigador (que evaluó el comité científico) ha preguntado que qué catedrática (que evaluó el comité científico) ha conseguido el proyecto para investigar el cáncer.

(5) a. El profesor (que premió la Universidad Complutense) ha preguntado si la investigadora (que premió la Universidad Complutense) ha publicado el estudio en la revista Nature.

b. El profesor (que premió la Universidad Complutense) ha preguntado que qué investigadora (que premió la Universidad Complutense) ha publicado el estudio en la revista Nature.

(6) a. El empresario (que vieron entrando al despacho) ha preguntado si la activista (que vieron entrando al despacho) ha forzado la cerradura para acceder al edificio.

b. El empresario (que vieron entrando al despacho) ha preguntado que qué activista (que vieron entrando al despacho) ha forzado la cerradura para acceder al edificio.

(7) a. El concejal (que critican las asociaciones feministas) ha preguntado si la cantante (que critican las asociaciones feministas) ha cancelado el concierto por las quejas recibidas.

b. El concejal (que critican las asociaciones feministas) ha preguntado que qué cantante (que critican las asociaciones feministas) ha cancelado el concierto por las quejas recibidas.

(8) a. El presentador (que expulsaron la semana pasada) ha preguntado si la concursante (que expulsaron la semana pasada) ha aceptado la propuesta para volver al programa.

b. El presentador (que expulsaron la semana pasada) ha preguntado que qué concursante (que expulsaron la semana pasada) que expulsaron la semana pasada ha aceptado la propuesta para volver al programa.

(9) a. El diputado (que criticó duramente la oposición) ha preguntado si la ministra (que criticó duramente la oposición) ha escrito sus memorias aunque sigue en activo.

b. El diputado (que criticó duramente la oposición) ha preguntado que qué ministra (que criticó duramente la oposición) ha escrito sus memorias aunque sigue en activo.

(10) a. La periodista (que invitaron a la tertulia) ha preguntado si el poeta (que invitaron a la tertulia) ha enviado una carta para denunciar la corrupción.

b. La periodista (que invitaron a la tertulia) ha preguntado que qué poeta (que invitaron a la tertulia) ha enviado una carta para denunciar la corrupción.

(11) a. El taxista (que retuvo el guarda jurado) ha preguntado si el pasajero (que retuvo el guarda jurado) ha perdido la maleta al salir del aeropuerto.

b. El taxista (que retuvo el guarda jurado) ha preguntado que qué pasajero (que retuvo el guarda jurado) ha perdido la maleta al salir del aeropuerto.

(12) a. El ladrón (que buscaba la Guardia Civil) ha preguntado si el fugitivo (que buscaba la Guardia Civil) ha tirado el botín cuando cruzaba la frontera.

b. El ladrón (que buscaba la Guardia Civil) ha preguntado que qué fugitivo (que buscaba la Guardia Civil) ha tirado el botín cuando cruzaba la frontera.

(13) a. La influencer (que admiran todos los adolescentes) ha preguntado si el modelo (que admiran todos los adolescentes) ha cancelado el evento porque tiene miedo escénico.

b. La influencer (que admiran todos los adolescentes) ha preguntado que qué modelo (que admiran todos los adolescentes) ha cancelado el evento porque tiene miedo escénico.

(14) a. El colaborador (que temen todos los políticos) ha preguntado si el reportero (que temen todos los políticos) ha propuesto un programa para entrevistar a personalidades.

b. El colaborador (que temen todos los políticos) ha preguntado que qué reportero (que temen todos los políticos) ha propuesto un programa para entrevistar a personalidades.

(15) a. El señor (que visitaron los trabajadores sociales) ha preguntado si el vecino (que visitaron los trabajadores sociales) ha perdido su piso pese a las protestas.

b. El señor (que visitaron los trabajadores sociales) ha preguntado que qué vecino (que visitaron los trabajadores sociales) ha perdido su piso pese a las protestas.

(16) a. El fotógrafo (que quieren para la portada) ha preguntado si el modelo (que quieren para la portada) ha pospuesto la sesión por el mal tiempo.

b. El fotógrafo (que quieren para la portada) ha preguntado que qué modelo (que quieren para la portada) ha pospuesto la sesión por el mal tiempo.

(17) a. El chef (que elogian los críticos culinarios) ha preguntado si el cocinero (que elogian los críticos culinarios) ha copiado el postre del libro de recetas.

b. El chef (que elogian los críticos culinarios) ha preguntado que qué cocinero (que elogian los críticos culinarios) ha copiado el postre del libro de recetas.

(18) a. El contable (que investiga la policía anticorrupción) ha preguntado si el administrativo (que investiga la policía anticorrupción) ha destruido el dosier antes de la redada.

b. El contable (que investiga la policía anticorrupción) ha preguntado que qué administrativo (que investiga la policía anticorrupción) ha falsiificado el dosier antes de la redada.

(19) a. La alcaldesa (que apoyan los principales sindicatos) ha preguntado si la concejal (que apoyan los principales sindicatos) ha leído el proyecto antes de proponer cambios.

b. La alcaldesa (que apoyan los principales sindicatos) ha preguntado que qué concejal (que apoyan los principales sindicatos) ha leído el proyecto antes de proponer cambios.

(20) a. La encargada (que ascendieron hace poco tiempo) ha preguntado si la dependienta (que ascendieron hace poco tiempo) ha presentado su dimisión porque estaba muy estresada.

b. La encargada (que ascendieron hace poco tiempo) ha preguntado que qué dependienta (que ascendieron hace poco tiempo) ha presentado su dimisión porque estaba muy estresada.

(21) a. La representante (que boicotearon las grandes empresas) ha preguntado si la actriz (que boicotearon las grandes empresas) ha publicado un comunicado en sus redes sociales.

b. La representante (que boicotearon las grandes empresas) ha preguntado que qué actriz (que boicotearon las grandes empresas) ha enseñado un comunicado en sus redes sociales.

(22) a. La cirujana (que recomendó el equipo médico) ha preguntado si la enfermera (que recomendó el equipo médico) ha pedido la baja porque tiene muchos dolores.

b. La cirujana (que recomendó el equipo médico) ha preguntado que qué enfermera (que recomendó el equipo médico) ha pedido la baja porque tiene muchos dolores.

(23) a. La agencia (que contrataron los turistas italianos) ha preguntado si el guía (que contrataron los turistas italianos) ha cancelado el tour porque va a nevar.

b. La agencia (que contrataron los turistas italianos) ha preguntado que qué guía (que contrataron los turistas italianos) ha cancelado el tour porque va a nevar.

(24) a. La secretaria (que contrató la promotora inmobiliaria) ha preguntado si la arquitecta (que contrató la promotora inmobiliaria) ha solicitado una prórroga para acabar el proyecto.

b. La secretaria (que contrató la promotora inmobiliaria) ha preguntado que qué arquitecta (que contrató la promotora inmobiliaria) ha solicitado una prórroga para acabar el proyecto.

(25) a. El mendigo (que interrogaron los Mossos d'Esquadra) ha preguntado si la señora (que interrogaron los Mossos d'Esquadra) ha encontrado un cadáver dentro de la catedral.

b. El mendigo (que interrogaron los Mossos d'Esquadra) ha preguntado que qué señora (que interrogaron los Mossos d'Esquadra) ha encontrado un cadáver dentro de la catedral.

(26) a. La charcutera (que despidió el nuevo encargado) ha preguntado si la cajera (que despidió el nuevo encargado) ha puesto una reclamación por el trato recibido.

b. La charcutera (que despidió el nuevo encargado) ha preguntado que qué cajera (que despidió el nuevo encargado) ha puesto una reclamación por el trato recibido.

(27) a. La limpiadora (que vieron los nuevos inquilinos) ha preguntado si la vecina (que vieron los nuevos inquilinos) ha colgado un cartel en la entrada principal.

b. La limpiadora (que vieron los nuevos inquilinos) ha preguntado que qué vecina (que vieron los nuevos inquilinos) ha colgado un cartel en la entrada principal.

(28) a. La presentadora (que fichó la televisión pública) ha preguntado si el colaborador (que fichó la televisión pública) ha emitido un comunicado tras sus polémicas declaraciones.

b. La presentadora (que fichó la televisión pública) ha preguntado que qué colaborador (que fichó la televisión pública) ha emitido un comunidado tras sus polémicas declaraciones.

(29) a. La bedel (que expulsaron de la universidad) ha preguntado si el profesor (que expulsaron de la universidad) ha puesto una denuncia después de los altercados.

b. La bedel (que expulsaron de la universidad) ha preguntado que qué profesor (que expulsaron de la universidad) ha puesto una denuncia después de los altercados.

(30) a. La entrenadora (que fichó el Real Madrid) ha preguntado si el jugador (que fichó el Real Madrid) ha besado la portería después de marcar gol.

b. La entrenadora (que fichó el Real Madrid) ha preguntado que qué jugador (que fichó el Real Madrid) ha besado la portería después de marcar gol.

(31) a. La vendedora (que robaron en el mercadillo) ha preguntado si el cliente (que robaron en el mercadillo) ha tirado la fruta al perseguir al ladrón.

b. La vendedora (que robaron en el mercadillo) ha preguntado que qué cliente (que robaron en el mercadillo) ha tirado la fruta al perseguir al ladrón.

(32) a. La cantante (que escucha todo el mundo) ha preguntado si la canción (que escucha todo el mundo) ha batido el record de reproducciones en Spotify.

b. La cantante (que escucha todo el mundo) ha preguntado si la canción (que escucha todo el mundo) ha batido el record de reproducciones en Spotify.

(33) a. La directora (que denunciaron por acoso sexual) ha preguntado si el actor (que denunciaron por acoso sexual) ha rechazado el papel en la nueva película.

b. La directora (que denunciaron por acoso sexual) ha preguntado que qué actor (que denunciaron por acoso sexual) ha rechazado el papel en la nueva película.

(34) a. El payaso (que ficharon en el circo) ha preguntado si el trapecista (que ficharon en el circo) ha desgastado la cuerda por hacer demasiados ensayos.

b. El payaso (que ficharon en el circo) ha preguntado que qué trapecista (que ficharon en el circo) ha desgastado la cuerda por hacer demasiados ensayos.

(35) a. La jueza (que entrevistaron en el telediario) ha preguntado si el chico (que entrevistaron en el telediario) ha presenciado el atraco desde dentro del banco.

b. La jueza (que entrevistaron en el telediario) ha preguntado que qué chico (que entrevistaron en el telediario) ha presenciado el atraco desde dentro del banco.

(36) a. La camarera (que golpearon con una bandeja) ha preguntado si el cliente (que golpearon con una bandeja) ha pedido una compensación por las secuelas psicológicas.

b. La camarera (que golpearon con una bandeja) ha preguntado que qué cliente (que golpearon con una bandeja) ha pedido una compensación por las secuelas psicológicas.

3 Experimental materials used in Experiment 5

(1) a. A la abogada (que buscaba el secretario judicial) _____

b. ¿A qué abogada (que buscaba el secretario judicial) _____?

(2) a. A la chica (que registró la policía secreta) _____

b. ¿A qué chica (que registró la policía secreta) _____?

(3) a. A la residente (que supervisó el médico adjunto) _____

b. ¿A qué residente (que supervisó el médico adjunto) _____?

(4) a. A la catedrática (que evaluó el comité científico) _____

b. ¿A qué catedrática (que evaluó el comité científico) _____?

(5) a. A la investigadora (que premió la Universidad Complutense) _____

b. ¿A qué investigadora (que premió la Universidad Complutense) _____?

(6) a. A la ejecutiva (que apoya la junta directiva) _____

b. ¿A qué ejecutiva (que apoya la junta directiva) _____?

(7) a. A la cantante (que critican las asociaciones feministas) _____

b. ¿A qué cantante (que critican las asociaciones feministas) _____?

(8) a. A la concursante (que eligieron todos los espectadores) _____

b. ¿A qué concursante (que eligieron todos los espectadores) _____?

(9) a. A la ministra (que criticó duramente la oposición) _____

b. ¿A qué ministra (que criticó duramente la oposición) _____?

(10) a. A la escritora (que elogian los críticos literarios) _____

b. ¿A qué escritora (que elogian los críticos literarios) _____?

(11) a. A la pasajera (que retuvo el guarda jurado) _____

b. ¿A qué pasajera (que retuvo el guarda jurado) _____?

(12) a. A la fugitiva (que buscaba la Guardia Civil) _____

b. ¿A qué fugitiva (que buscaba la Guardia Civil) _____?

(13) a. A la famosa (que admiran todos los adolescentes) _____

b. ¿A qué famosa (que admiran todos los adolescentes) _____?

(14) a. A la periodista (que temen todos los políticos) _____

b. ¿A qué periodista (que temen todos los políticos) _____?

(15) a. A la señora (que visitaron los trabajadores sociales) _____

b. ¿A qué señora (que visitaron los trabajadores sociales) _____?

(16) a. A la modelo (que querían muchas agencias internacionales) _____

b. ¿A qué modelo (que querían muchas agencias internacionales) _____?

(17) a. A la cocinera (que elogian los críticos culinarios) _____

b. ¿A qué cocinera (que elogian los críticos culinarios) _____?

(18) a. A la administrativa (que investiga la policía anticorrupción) _____

b. ¿A qué administrativa (que investiga la policía anticorrupción) _____?

(19) a. A la candidata (que proponen los principales sindicatos) _____

b. ¿A qué candidata (que proponen los principales sindicatos) _____?

(20) a. A la dependienta (que quieren todos los trabajadores) _____

b. ¿A qué dependienta (que quieren todos los trabajadores) _____?

(21) a. A la empresaria (que boicotearon las grandes empresas) _____

b. ¿A qué empresaria (que boicotearon las grandes empresas) _____?

(22) a. A la enfermera (que recomendó el equipo médico) _____

b. ¿A qué enfermera (que recomendó el equipo médico) _____?

(23) a. A la guía (que contrataron los turistas italianos) _____

b. ¿A qué guía (que contrataron los turistas italianos) _____?

(24) a. A la arquitecta (que contrató la promotora inmobiliaria) _____

b. ¿A qué arquitecta (que contrató la promotora inmobiliaria) _____?

Appendix D

Supplementary materials to Chapter 4

1 Experimental materials used in Experiments 6 & 7

(1) Edurnek (herriko lehiaketan sari guztiak jaso dituen) aizkolariari (herriko lehiaketan sari guztiak jaso dituen) pastela ekarri dio eta ziztu bizian jan du.

(2) Itziarrek (diru laguntza gehiegi jaso omen dituen) pilotariari (diru laguntza gehiegi jaso omen dituen) polikiroldegia erakutsi dio bertan frontoirik ez badago ere.

(3) Amaiak (goiz osoan gure bulegoan egon den) etorkinari (goiz osoan gure bulegoan egon den) dokumentua eman dio behingoz eskaera ofiziala egin dezan.

(4) Kepak (azken boladan asko aipatu izan den) marrazkilariari (azken boladan asko aipatu izan den) komikia bidali dio lehenbailehen bigarren alea prestatzen hasteko.

(5) Kirmenek (Eibarren eta inguruko herrietan ezaguna den) gizonari (Eibarren eta inguruko herrietan ezaguna den) garagardoa atera dio eta lagunekin bazkaltzera gonbidatu du.

(6) Aitorrek (nazioartean ospe handia izan omen duen) pailazoari (nazioartean ospe handia izan omen duen) jolasa saldu dio eta dirutza handia irabazi du.

(7) Asierrek (aurreko egunean Andoniren dendan ikusi genuen) neskari (aurreko egunean Andoniren dendan ikusi genuen) diska erosi dio eta Wallapopen saldu du segituan.

(8) Iratik (jende askoren artean ezaguna omen den) taldeari (jende askoren artean ezaguna omen den) kanta eskatu dio baina azkenean ez dute jo.

(9) Zuriñek (gure auzoko emakume guztien gustukoa den) gizonari (gure auzoko emakume guztien gustukoa den) alkandora urratu dio baina beste bat erosiko dio.

(10) Jonek (entzule guztien artean sumina sortu duen) epaileari (entzule guztien artean sumina sortu duen) iruzkina zuzendu dio eta epailea ez da lotsatu.

(11) Amak (arratsalde osoan gela hartan egon den) irakasleari (arratsalde osoan gela hartan egon den) tarta ekarri dio eta poz handiz hartu du.

(12) Lehendakariak (Zabalgana eta Salburuako auzokideen babesa daukan) zinegotziari (Zabalgana eta Salburuako auzokideen babesa daukan) plana erakutsi dio aldaketa gehiagorik gabe onar dezan.

(13) Argitaratzaileak (duela bi urte arrakasta izan zuen) itzultzaileari (duela bi urte arrakasta izan zuen) liburua eskaini dio baina ez du interesik erakutsi.

(14) Alkateak (hiriko kaleetan barrena galduta ibili den) haurrari (hiriko kaleetan barrena galduta ibili den) txakurra oparitu dio udaleko txakurtegira eraman ez dezaten.

(15) Zientzialariak (irrati eta telebista-kate guztietan agertu den) aurkezleari (irrati eta telebista-kate guztietan agertu den) aurkikuntza erakutsi dio eta hitzik gabe utzi du.

(16) Boluntarioak (nazioarteko hedabideetan oso kritikatua izan den) politikariari (nazioarteko hedabideetan oso kritikatua izan den) erabakia azaldu dio eta arazo sozialak azpimarratu ditu.

(17) Musikariak (nire lagunaren iritziz izugarri polita den) abeslariari (nire lagunaren iritziz izugarri polita den) balada idatzi dio hurrengo estudioko albumean sar dezan.

(18) Zuzendariak (gaurko administrazio batzordean goretsia izan den) ordezkariari (gaurko administrazio batzordean goretsia izan den) proposamena itzuli dio eta langileei bidaltzeko eskatu du.

(19) Tenoreak (opera munduan lagun asko omen dituen) sopranoari (opera munduan lagun asko omen dituen) musikaria aurkeztu dio eta bakarrik utzi ditu biak.

(20) Entrenatzaileak (Athleticeko zale guztiei asko gustatzen zaien) futbolariari (Athleticeko zale guztiei asko gustatzen zaien) kamiseta hartu dio eta honek zaletuari oparitu dio.

(21) Ertzainak (ondoko etxebizitzan odolez zikinduta aurkitu duen) lapurrari (ondoko etxebizitzan odolez zikinduta aurkitu duen) pistola kendu dio eta komisaldegi zentralera eraman du.

(22) Editoreak (aurtengo Euskadi Saria irabazteko aukera duen) kazetariari (aurtengo Euskadi Saria irabazteko aukera duen) artikulua kritikatu dio batere originaltasunik ez duela esanez.

(23) Begiraleak (familian ia denok aspalditik ezagutzen dugun) mutilari (familian ia denok aspalditik ezagutzen dugun) sekretua aitortu dio presaka etxera alde egin aurretik.

(24) Erizainak (medikuen artean kezka handia piztu duen) gaixoari (medikuen artean kezka handia piztu duen) botika agindu dio laster onera egingo duelako esperantzarekin.

2 Experimental materials used in Experiment 8

(1) Edurnek (herriko lehiaketan sari guztiak jaso dituen) aizkolariari _____

(2) Itziarrek (diru laguntza gehiegi jaso omen dituen) pilotariari _____

(3) Amaiak (goiz osoan gure bulegoan egon den) etorkinari _____

(4) Kepak (azken boladan asko aipatu izan den) marrazkilariari _____

(5) Kirmenek (Eibarren eta inguruko herrietan ezaguna den) gizonari _____

(6) Aitorrek (nazioartean ospe handia izan omen duen) pailazoari _____

(7) Asierrek (aurreko egunean Andoniren dendan ikusi genuen) neskari _____

(8) Iratik (jende askoren artean ezaguna omen den) taldeari _____

(9) Zuriñek (gure auzoko emakume guztien gustukoa den) gizonari _____

(10) Jonek (entzule guztien artean sumina sortu duen) epaileari _____

(11) Amak (arratsalde osoan gela hartan egon den) irakasleari _____

(12) Lehendakariak (Zabalgana eta Salburuako auzokideen babesa daukan) zinegotziari _____

(13) Argitaratzaileak (duela bi urte arrakasta izan zuen) itzultzaileari _____

(14) Alkateak (hiriko kaleetan barrena galduta ibili den) haurrari _____

(15) Zientzialariak (irrati eta telebista-kate guztietan agertu den) aurkezleari _____

(16) Boluntarioak (nazioarteko hedabideetan oso kritikatua izan den) politikariari _____

(17) Musikariak (nire lagunaren iritziz izugarri polita den) abeslariari _____

(18) Zuzendariak (gaurko administrazio batzordean goretsia izan den) ordezkariari _____

(19) Tenoreak (opera munduan lagun asko omen dituen) sopranoari _____

(20) Entrenatzaileak (Athleticeko zale guztiei asko gustatzen zaien) futbolariari _____

(21) Ertzainak (ondoko etxebizitzan odolez zikinduta aurkitu duen) lapurrari _____

(22) Editoreak (aurtengo Euskadi Saria irabazteko aukera duen) kazetariari _____

(23) Begiraleak (familian ia denok aspalditik ezagutzen dugun) mutilari _____

(24) Erizainak (medikuen artean kezka handia piztu duen) gaixoari _____

References

Aldezabal, I., Aranzabe, M., Atutxa, A., Gojenola, K., Sarasola, K., & Zabala, I. (2003). *Hitz-hurrenkeren azterketa masiboa corpusean*. UPV/EHU/LSI/TR 2-2003. Department of Language and Informatic Systems.

Alexopoulou, T. & Keller, F. (2013). What vs. who and which: Kind-denoting fillers and the complexity of whether-islands. In J. Sprouse & N. Hornstein (Eds.), *Experimental Syntax and Island Effects* (pp. 310–340). Cambridge, UK: Cambridge University Press.

Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An erp study. *Language and Cognitive Processes*, *18*(4), 405–430.

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, *106*(4), 748.

Almor, A. (2004). A computational investigation of reference in production and comprehension. In J. Trueswell & M. Tanenhaus (Eds.), *Approaches to Studying World-Situated Language Use: Bridging the Language-As-Product and Language-As-Action Traditions* (pp. 285–301). Cambridge, MA: MIT Press.

Altmann, G. T. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science*, *29*(3), 313–341.

Anderson, J. R. & Lebiere, C. J. (1998). *The Atomic Components of Thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Anderson, J. R., Qin, Y., Stenger, V. A., & Carter, C. S. (2004). The relationship of three cortical regions to an information-processing model. *Journal of Cognitive Neuroscience*, *16*(4), 637–653.

Anderson, J. R. & Reder, L. M. (1979). An elaborative processing explanation of depth of processing. In L. S. Cermak & F. I. M. Craik (Eds.), (pp. 385–404). Hillsdale, NJ: Lawrence Erlbaum.

Anderson, M. C. & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of Perception and Cognition* (pp. 237–313). San Diego, CA: Academic Press.

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.

Baddeley, A. D. (1986). *Working Memory*. Oxford, UK: Oxford University Press.

Baddeley, A. D. & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1178–1198.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bever, T. G. (1970). The cognitive basis for linguistic structures. *Cognition and the Development of Language*, *279*(362), 1–61.

Bhatt, R., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D. M., & Xia, F. (2009). A multi-representational and multi-layered treebank for Hindi/Urdu. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)* (pp. 186–189).

Bloomfield, L. (1933). *Language*. New York, NY: Henry Holt.

Bock, J. K. & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*(1), 45–93.

Bradshaw, G. L. & Anderson, J. R. (1982). Elaborative encoding as an explanation of levels of processing. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 165–174.

Brainard, D. H. & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, *52*(1), 144–169.

Choi, H.-W. (2007). Length and order: A corpus study of Korean dative-accusative construction. *Discourse and Cognition*, *14*(3), 207–227.

Chomsky, N. (1973). Conditions on transformations. In S. R. Anderson & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 232–286). New York, NY: Holt, Rinehart and Winston.

Chomsky, N. (1977). On wh-movement. In P. W. Culicover, T. Wasow, & A. Akmajian (Eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 71–132). New York, NY: Academic Press.

Chomsky, N. (1983). *Lectures on Government and Binding*. Berlin, Germany: Mouton de Gruyter.

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, J. Uriagereka, & S. J. Keyser (Eds.), *Step by Step. Essays on Minimalist Syntax in Honor of Howard Lasnik* (pp. 89–155). Cambridge, MA: The MIT Press.

Chomsky, N. & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology, Vol. 2* (pp. 269–321). Amsterdam: Wiley.

Christiansen, M. H. & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning*, *59*, 126–161.

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

Clifton, C. & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In *Linguistic Structure in Language Processing* (pp. 273–317). Dordrecht, NL: Springer.

Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. Oxford, UK: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57.

Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Memory and Language*, *19*(4), 450.

De Rijk, R. P. (1969). Is Basque an SOV language? *Fontes Linguae Vasconum: Studia et Documenta*, *1*(3), 319–352.

Delorme, A. & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Eusko Jaurlaritza. (2016). *VI Encuesta Sociolingüística*. Retrieved from https://www.euskara.euskadi.eus/contenidos/informacion/argitalpenak/es_6092/adjuntos/2016%20VI%20INK%20SOZLG%20-%20Euskal%20Herria%20gaz.pdf

Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. *Brain and Language*, *87*(3), 345–354.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*(2), 210–233.

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.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10).

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9*(3), 427–473.

Foraker, S. & McElree, B. (2011). Comprehension of linguistic dependencies: Speed-accuracy tradeoff evidence for direct-access retrieval from memory. *Language and Linguistics Compass*, *5*(11), 764–783.

Frank, S. L., Trompenaars, T., & Vasishth, S. (2016). Cross-linguistic differences in processing double-embedded relative clauses: Working-memory constraints or language statistics? *Cognitive Science*, *40*(3), 554–578.

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, *5*(4), 519–559.

Frazier, L. & Clifton Jr, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, *4*(2), 93–126.

Frazier, L. & Flores d'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*(3), 331–344.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183–192.

Futrell, R., Mahowald, K., & Gibson, E. (2015a). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341.

Futrell, R., Mahowald, K., & Gibson, E. (2015b). Quantifying word order freedom in dependency corpora. In *Proceedings of the Third International Conference on Dependency Linguistics (DepLing 2015)* (pp. 91–100).

Gallego, Á. (2010). *Phase Theory*. Amsterdam, NL: John Benjamins Publishing.

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language*, *58*(4), 1095–1111.

Gernsbacher, M. A. (1990). *Language Comprehension as Structure Building*. Hillsdale, NJ: Lawrence Erlbaum.

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. In A. P. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain* (pp. 95–126). Cambridge, MA: The MIT Press.

Gibson, E. & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, *14*(3), 225–248.

Gibson, E. & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, *7*(1), 55–78.

Gildea, D. & Jaeger, T. F. (2015). Human languages order information efficiently. *arXiv preprint arXiv:1510.02823*.

Gildea, D. & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science*, *34*(2), 286–310.

Goodall, G. (2015). The D-linking effect on extraction from islands and non-islands. *Frontiers in Psychology*, *5*(1493), 1–11.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the p600. *Language and Cognitive Processes*, *25*(2), 149–188.

Grodner, D. & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, *29*(2), 261–290.

Grodner, D., Gibson, E., & Tunstall, S. (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language*, *46*(2), 267–295.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439–483.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies* (pp. 1–8). Association for Computational Linguistics.

Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. New York, NY: Cambridge University Press.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. New York, NY: Oxford University Press.

Hofmeister, P. (2007). Retrievability and gradience in filler-gap dependencies. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* (Vol. 43, *1*, pp. 109–123). Chicago Linguistic Society.

Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, *26*(3), 376–405.

Hofmeister, P., Jaeger, T. F., Sag, I. A., Arnon, I., & Snider, N. (2007). Locality and accessibility in *wh*-questions. In S. Featherston & W. Sternefeld (Eds.), *Roots: Linguistics in Search of Its Evidential Base* (pp. 185–206). Berlin: Mouton de Gruyter.

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

Hofmeister, P. & Vasishth, S. (2014). Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology*, *5*, 1237.

Hsiao, F. & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, *90*(1), 3–27.

Hualde, J. I. & Ortiz de Urbina, J. (2003). *A Grammar of Basque*. Berlin, Germany: Mouton de Gruyter.

Hunt, R. R. (1995). The subtlety of distinctiveness: What von restorff really did. *Psychonomic Bulletin & Review*, *2*(1), 105–112.

Husain, S., Bhatt, R., & Vasishth, S. (2013). Towards a psycholinguistically motivated dependency grammar for Hindi. In *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)* (pp. 108–117).

Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong expectations cancel locality effects: Evidence from Hindi. *PLoS ONE*, *9*(7).

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Jaeger, T. F., Fedorenko, E., Hofmeister, P., & Gibson, E. (2008). Expectation-based syntactic processing: Anti-locality effects outside of head-final languages. In *21st Annual CUNY Sentence Processing Conference*. University of North Carolina at Chapel Hill, NC.

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*(10), 1745–1758.

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

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201.

Kaan, E. & Swaab, T. Y. (2003). Electrophysiological evidence for serial sentence processing: A comparison between non-preferred and ungrammatical continuations. *Cognitive Brain Research*, *17*(3), 621–635.

Kachru, Y. (2006). *Hindi*. Amsterdam, NL: John Benjamins Publishing.

Kaiser, E. & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147.

Karimi, H. & Ferreira, F. (2016). Informativity renders a referent more accessible: Evidence from eyetracking. *Psychonomic Bulletin & Review*, *23*(2), 507–525.

Karimi, H., Fukumura, K., Ferreira, F., & Pickering, M. J. (2014). The effect of noun phrase length on the form of referring expressions. *Memory & Cognition*, *42*(6), 993–1009.

Karimi, H., Swaab, T. Y., & Ferreira, F. (2018). Electrophysiological evidence for an independent effect of memory retrieval on referential processing. *Journal of Memory and Language*, *102*, 68–82.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, *1*(1), 3–44.

Kim, A. & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225.

King, J. & Kutas, M. (1995). Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*(3), 376–395.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1.

Kolers, P. A. (1973). Remembering operations. *Memory & Cognition*, *1*(3), 347–355.

Kolk, H. H., Chwilla, D. J., Van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*(1), 1–36.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627–645.

Konieczny, L. & Döring, P. (2003). Anticipation of clause-final heads: Evidence from eye-tracking and SRNs. In *Proceedings of ICCS/ASCS* (pp. 13–17). Sydney, NSW.

Kuno, S. & Robinson, J. J. (1972). Multiple wh questions. *Linguistic Inquiry*, *3*(4), 463–487.

Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kutas, M. & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Laka, I. (1996). *A Brief Grammar of Euskara, the Basque Language*. Retrieved from https://www.ehu.eus/documents/2430735/0/A-brief-grammar-of-euskara.pdf

Lau, E., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502.

Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Levy, R., Fedorenko, E., & Gibson, E. (2013). The syntactic complexity of Russian relative clauses. *Journal of Memory and Language*, *69*(4), 461–495.

Levy, R. & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, *68*(2), 199–222.

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.

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159–191.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. L. (1976). An elaborative processing explanation of depth of processing. In J. Brown (Ed.). London, UK: Wiley.

Lopez-Calderon, J. & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.

MATLAB. (2017). *version 9.2.0 (R2017a)*. Natick, MA: The MathWorks Inc.

McDaniel, M. A. (1981). Syntactic complexity and elaborative processing. *Memory & Cognition*, *9*(5), 487–495.

McDaniel, M. A., Dunay, P. K., Lyman, B. J., & Kerwin, M. L. E. (1988). Effects of elaboration and relational distinctiveness on sentence memory. *The American Journal of Psychology*, 357–369.

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. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation: Advances In Research and Theory* (pp. 155–200). Oxford, UK: Elsevier Academic Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Nakatani, K. & Gibson, E. (2008). Distinguishing theories of syntactic expectation cost in sentence comprehension: Evidence from Japanese. *Linguistics*, *46*(1), 63–87.

Nakatani, K. & Gibson, E. (2010). An on-line study of Japanese nesting complexity. *Cognitive Science*, *34*(1), 94–112.

Ness, T. & Meltzer-Asscher, A. (2017). Working memory in the processing of long-distance dependencies: Interference and filler maintenance. *Journal of psycholinguistic research*, *46*(6), 1353–1365.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of cognitive Neuroscience*, *3*(2), 151–165.

Nicol, J. & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 5–19.

Nivre, J., De Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., … Silveira, N., et al. (2016). Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 1659–1666).

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Osterhout, L. & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806.

Osterhout, L. & Swinney, D. A. (1993). On the temporal course of gap-filling during comprehension of verbal passives. *Journal of Psycholinguistic Research*, *22*(2), 273–286.

Pastor, L. (2019). Sobre el orden básico de palabras en euskera: Un nuevo estudio de corpus. In I. Laka (Ed.), *Pello Salaburuari esker onez* (pp. 159–180). Bilbao, Spain: UPV/EHU Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Pesetsky, D. (1987). Wh-in-situ: Movement and unselective binding. In E. J. Reuland & A. G. ter Meulen (Eds.), *The Representation of (In) definiteness* (pp. 98–129). Cambridge, MA: The MIT Press.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407–428.

Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, *86*(4), 777.

R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Ros, I. (2018). *Minimizing Dependencies Across Languages and Speakers: Evidence from Basque, Polish and Spanish and Native and Non-Native Bilinguals* (Doctoral dissertation, University of the Basque Country UPV/EHU).

Ros, I., Santesteban, M., Fukumura, K., & Laka, I. (2015). Aiming at shorter dependencies: The role of agreement morphology. *Language, Cognition and Neuroscience*, *30*(9), 1156–1174.

Ross, J. (1967). *Constraints on Variables in Syntax* (Doctoral dissertation, Massachusetts Institute of Technology).

Ruchkin, D. S., Johnson Jr, R., Canoune, H., & Ritter, W. (1990). Short-term memory storage and retention: An event-related brain potential study. *Electroencephalography and Clinical Neurophysiology*, *76*(5), 419–439.

Safavi, M. S., Husain, S., & Vasishth, S. (2016). Dependency resolution difficulty increases with distance in Persian separable complex predicates: Evidence for expectation and memory-based accounts. *Frontiers in Psychology*, *7*, 403.

Sasano, R. & Okumura, M. (2016). A corpus-based analysis of canonical word order of Japanese double object constructions. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 2236–2244).

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038.

Sprouse, J. (2007). *A Program for Experimental Syntax: Finding the Relationship between Acceptability and Grammatical Knowledge* (Doctoral dissertation, University of Maryland).

Stepanov, A. & Stateva, P. (2015). Cross-linguistic evidence for memory storage costs in filler-gap dependencies with wh-adjuncts. *Frontiers in Psychology*, *6*, 1301.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*(3736), 652–654.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245.

Szmrecányi, B. (2004). On operationalizing syntactic complexity. In *Le poids des mots. 7th international conference on textual data statistical analysis* (Vol. 2, pp. 1032–1039). Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Bulletin*, *30*(4), 415–433.

Tesnière, L. (1959). *Eléments de syntaxe structurale*. Paris: Librairie C. Klincksieck.

Tily, H. J. (2010). *The role of processing complexity in word order variation and change* (Doctoral dissertation, Stanford University).

Ueno, M. & Garnsey, S. M. (2008). An ERP study of the processing of subject and object relative clauses in Japanese. *Language and Cognitive Processes*, *23*(5), 646–688.

Van Dyke, J. A. & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A retrieval interference theory of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(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.

Van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*(2), 241–255.

Vasishth, S. (2003). *Working Memory in Sentence Comprehension: Processing Hindi Center Embeddings*. New York, NY: Routledge.

Vasishth, S. & Drenhaus, H. (2011). Locality in German. *Dialogue & Discourse*, *2*(1), 59–82.

Vasishth, S. & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, 767–794.

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, *103*, 151–175.

Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, *25*(4), 533–567.

Villata, S., Rizzi, L., & Franck, J. (2016). Intervention effects and Relativized Minimality: New experimental evidence from graded judgments. *Lingua*, *5*(1493), 1–11.

Von Restorff, H. (1933). Ueber die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung*, *18*(1), 299–342.

Wagers, M., Lau, E., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237.

Wagers, M. & Phillips, C. (2013). Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology*, *67*(7), 1274–1304.

Wanner, E. & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic Theory and Psychological Reality* (pp. 119–161). Cambridge, MA: The MIT Press.

Warren, T. & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79–112.

Wasow, T. (2002). *Postverbal Behavior*. Stanford, CA: CSLI Publications.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org