How long can you hold the filler

How long can you hold the filler: Maintenance and Retrieval

Nayoun Kim[1*], Laurel Brehm[2], Patrick Sturt[3], Masaya Yoshida[1]

[1] *Department of Linguistics, Northwestern University, Evanston, IL, USA*

[2] *Max Planck Institute for Psycholinguistics, The Netherlands*

[3] *Department of Psychology, University of Edinburgh*

*Corresponding Author
Department of Linguistics, Northwestern University
2016 Sheridan Rd.
Evanston, IL 60208 USA
+1 847 868 5236
NayounKim2018@u.northwestern.edu

How long can you hold the filler: Maintenance and Retrieval

This study attempts to reveal the mechanisms behind the online formation of Wh-Filler-Gap Dependencies (WhFGD). Specifically, we aim to uncover how maintenance and retrieval work in WhFGD processing by paying special attention to the information associated with the wh-filler that is retrieved when the gap is recognized. We use the agreement attraction phenomenon (Pearlmutter, Garnsey & Bock, 1999; Wagers et al., 2009) as a probe. The first and second experiments examined the type of information that is maintained and how maintenance is motivated, investigating the retrieved information at the gap for reactivated fillers and definite NPs. The third experiment examined the role of the retrieval associated with reactivated and active fillers. We contend that the information being accessed reflects the extent to which the filler is maintained, where the reader is able to access fine-grained information including category information as well as a representation of both the head and the modifier at the verb. This suggests that both retrieval and maintenance components play a role in resolving online whFGD (Fiebach et al., 2002; Wagers & Phillips, 2014) and for parsing at large.

Key words: Filler Gap Dependency, Syntax, Maintenance, Retrieval, Agreement Attraction.

1. Introduction

Resolving Wh-Filler-Gap Dependencies (WhFGD) involves linking a wh-phrase to a verb, preposition, or gap. An example of aWhFGD construction is (1).

(1) Which mistake in the program will be disastrous for the company?

In (1), neither the interpretation nor the grammatical status of the wh-phrase *which mistake in the program* is determined solely by the wh-phrase itself. The wh-phrase, which is the subject of the *disastrous*, is interpreted as the theme argument of the predicate *disastrous*, and bears Nominative Case, assigned from *disastrous*. In general, the grammatical status and the interpretation of a wh-phrase are determined in relation to other elements, such as the verb or preposition, or the gap, a controlling element. The dependent element is often referred to as a filler (e.g., the wh-phrase in *which mistake in the program*), and the controlling element which hosts the grammatically mandatory yet hidden argument is referred to as a gap.[1]

One of the important properties of long-distance dependencies is that they can span across a large number of words or clauses. In online WhFGD resolution, the parser needs to link the wh-filler to the gap in order to achieve the interpretation of the WhFGD sentence; for a wh-phrase to be interpreted, the wh-phrase needs to be linked to the gap. In other words, to resolve WhFGD, the parser needs to 'recover' the information of the filler after encountering the gap, in order to achieve the right interpretation of the sentence (Bever & McElree, 1988; Crain & Fodor, 1985; Fodor, 1978; McElree & Bever, 1989; Nicol, Fodor. & Swinney, 1989). This implies that in order to resolve a WhFGD online, the parser needs to perform two processes. One is the storage or maintenance of a wh-filler (Gazdar & Mellish, 1989; Gibson, 1998; Gibson & Warren, 2004; Wagers & Phillips, 2014; Wanner & Maratsos, 1978; Warren & Gibson, 2002), and the

[1] Note that we do not commit to a specific analysis of WhFGD constructions. Specifically, we are agnostic about whether it involves a phonetically empty gap or not. We customarily call the controlling element as *gap*, but our conclusions do not necessarily require a gap-based analysis.

other is the retrieval or reactivation of the wh-filler (Lewis & Vasishth, 2005; McElree, 2001; McElree & Dosher, 1989; McElree, 2006; Nicol et al., 1989; Nicol, Fodor, & Swinney, 1994; Oberauer & Kliegl, 2006; Van Dyke, 2007; Van Dyke & McElree, 2006).

This study attempts to reveal the mechanisms behind online WhFGD formation. Specifically, we aim to uncover how maintenance and retrieval operate in WhFGD processing, by paying special attention to what information is retrieved from the wh-filler when the gap is recognized. We contend that if the wh-filler is released from maintenance and retrieved at a later point, its activation in memory will be lower, and its retrieval will be less successful, relative to a situation where it is maintained.

2. Processing Wh-Filler-Gap Dependencies

2.1. Maintenance and Retrieval

Let us look at the maintenance and retrieval components in more detail. First, it is possible that the wh-filler is maintained in memory until it is assigned a thematic role from the verb (Gibson, 1998; Gibson & Warren, 2004; Wagers & Phillips, 2014). Due to its morphological properties (i.e., *wh*-morphology, e.g., *which*) the parser can immediately recognize that the wh-filler is an element that will be linked to the gap somewhere downstream, or otherwise it cannot be interpreted. Note this is not true for other non-wh-NPs, like *the program*: a definite determiner *the* does not signal movement. The gap is not guaranteed to be adjacent to the wh-phrase, as it can appear in the subject position, the direct object position, the indirect object position, or the object position of a preposition. As such, the wh-phrase itself does not signal where the gap should be located. Thus, the parser needs to maintain the wh-filler in

memory until the gap is identified and the wh-filler is successfully linked to the gap.

Numerous studies have shown that upon encountering the filler, the parser actively posits a gap in advance of confirming evidence. This is known as active dependency formation (Aoshima, Phillips, & Weinberg, 2004; Crain & Fodor, 1985; Frazier & Flores D'Arcais, 1989; Lee, 2004; Omaki et al., 2015; Phillips, 2006; Pickering & Barry, 1991; Stowe, 1986; Traxler & Pickering, 1996). Active dependency formation can be understood as a consequence of the parser's maintenance of a wh-filler in memory. That is, while a wh-filler must be linked to a gap, the distance between the filler and the gap is potentially long. This means that the parser needs to maintain the wh-filler in memory for a potentially long distance until the gap is encountered and the wh-filler can be linked to the gap. If the wh-filler is maintained in memory, it would be costly for long dependencies, which in turn would lead the parser to resolve the dependency as quickly as possible (Gibson, 1998; Gibson & Warren, 2002).

Evidence for maintenance comes from studies showing larger processing costs when the head of the dependency is not resolved immediately due to many intervening words. For example, Chen, Gibson, & Wolf (2005) showed that the readers have difficulty in maintaining multiple unresolved dependencies when the right hand element of the dependency has not been encountered yet, such that the reading times at the most deeply embedded NP () position were faster for the relative clause constructions than the sentential complement constructions.

(2) a. SC structure: The announcement that the baker from a small bakery in New York City received the award helped the business of the owner.

b. RC structure: The announcement which the baker from a small bakery in New York City received helped the business of the owner.

This is because in sentence complement constructions, readers need to store the wh-element in memory until the dependency is resolved, whereas the wh-element in the relative clause does not need to be stored. In WhFGD processing, the reading time of the verb (that hosts the gap) is faster when the parser can form short filler-gap dependencies successively, versus when the parser needs to hold the filler for a longer time (see Keine, 2015 for related observations). This claim is bolstered by the findings of Gibson and Warren (2004), who observed that the reading times were slower when the number of words intervening between the wh-filler and the gap increased. When sentences involving wh-extraction are compared to those that do not involve wh-extraction, reading times of the words between the filler and gap increase for longer dependencies (see also Chen et al., 2005 and Stepanov & Stateva, 2015).

Once the gap is recognized, information associated with the wh-filler needs to be recovered or retrieved (McElree, 2006; McElree, Foraker, & Dyer, 2003; McElree & Dosher, 1989). This is necessary for the parser to check the case, thematic role and other morphological features of the filler and to achieve its proper interpretation (Bever & McElree, 1988; Crain & Fodor, 1985; Fodor, 1978; McElree & Bever, 1989; Nicol & Swinney, 1989; Nicol, Fodor, & Swinney, 1994). Fillers may contain different kinds of information, including morphological features, syntactic category, and lexical-semantic content. Some of these properties may be subject to memory decay (King & Just, 1981; Wagers & Phillips, 2014; Wanner & Maratsos, 1978). Different information could be maintained during the resolution of the dependency, or could decay and then be retrieved when the gap is recognized.[2]

[2] As an anonymous reviewer points out, some previous studies have suggested that decay is no

Wagers and Phillips (2014) investigated which aspects of the filler are maintained, and which are susceptible to decay. They observed a filled-gap effect (e.g. Boland, Tanenhaus, Garnsey, & Carlson, 1995; Crain & Fodor 1985; Frazier & Clifton 1989; Stowe, 1986; Tanenhaus, Boland, Garnsey & Carlson, 1989; Wagers & Phillips, 2014) when the wh-filler is an NP and the potential gap site is also an NP, but not when the wh-filler is a PP and the potential gap site is an NP, regardless of whether the WhFGD spans a short or long distance, as illustrated in (3) (Wagers & Phillips 2014; 1282); n.b. "FGE" stands for "filled gap effect".

(3)

					Plausibility Mismatch.		*FGE*
					↓		↓
a.	Wh-NP	...	SHORT DISTANCE	...	**V**	...	**NP**
b.	Wh-PP	...	SHORT DISTANCE	...	V	...	NP
							FGE

longer a useful explanatory concept in the retrieval literature (see Nairne 2002; McElree 2006, Berman et al., 2009; Lewandowsky, 2009). Wagers and Phillips (2009) pointed out that not all the features of the elements that are retrieved at the head of the dependency or are fully reactivated at the verb position (e.g., semantic features of the wh-filler in Wagers and Phillips's (2009) study). Such findings can be accounted for by memory decay. Thus, for present purposes, we hypothesize that some of the information associated with the filler is subject to decay or interference. We assume that the success of retrieval is related to the amount of material intervening between the filler and the gap.

								↓
c.	Wh-NP	...	LONG DISTANCE	...	V	...		**NP**
d.	Wh-PP	...	LONG DISTANCE	...	V	...		NP

This suggests that category information of the wh-filler is maintained throughout the dependency formation process. However, the semantic incongruity between the wh-filler and the verb (e.g. Boland et al., 1995; Traxler & Pickering, 1996) was not recognized. That is, the readers cannot detect the semantic incompatibility of the filler and the verb when the dependency spans a long distance, nor is the mismatch between the preposition attached to the wh-phrase and the verb recognized in long distance dependencies. This suggests that syntactic category information of the fillers is actively maintained during the online WhFGD formation process, but semantic content and lexical information are released from maintenance. In sum, the implication is that resolving filler gap dependencies involves both maintenance and retrieval, and the information that is retrieved at the verb position reflects what information of the filler is maintained and what information of the filler is released from maintenance.

Note that Chow & Zhou (2018) recently suggested that the lack of a plausibility effect is not because the content of the wh-filler is released from maintenance, but because of the lack of statistical power in earlier studies. They conducted an eye-tracking experiment with high statistical power. Like Wagers and Phillips (2014), they found that readers actively insert a gap regardless of dependency length whenever one is grammatically possible, suggesting an active gap filling effect. In addition, they found a plausibility effect in regression path duration as well as total reading times for all dependency lengths. Their findings therefore provide evidence for

the maintenance of semantic features. Furthermore, they found a weaker plausibility mismatch effect after the critical region for long dependencies, relative to short dependencies. Therefore, it is possible that, contrary to Wagers and Phillips (2014), thematic information can be maintained in memory. However, distance still impacts the retrieval of thematic information as the observed plausibility effects show, which suggests that memory decay may be in effect.

In the current studies, we investigate this claim, asking what sort of information from the wh-filler can be maintained: just category information, or something more detailed? Through a series of studies on online WhFGD formation, we show that, like in Chow & Zhou (2018), the wh-filler needs to be maintained in memory throughout the processing of WhFGD sentences, but if the wh-filler is released from maintenance and retrieved later, the relative strength of the filler, and thus the degree of success of its retrieval, is reduced.

We also posit an additional question: what motivates the maintenance of an element? In the case of WhFGD, wh-fillers involve distinctive wh-morphology. In a language like English, which is a wh-movement language, a phrase bearing wh-morphology provides strong evidence for the presence of WhFGD, i.e., if there is a wh-phrase, there must be a gap somewhere in the sentence (Chomsky, 1977). Therefore, it is plausible that wh-morphology signals the presence of a filler-gap dependency and thus leads the parser to maintain the wh-phrase. On the other hand, when a phrase does not bear wh-morphology, it is unclear whether the phrase is part of a filler-gap dependency. Assuming that maintaining an element in memory is costly (Gibson, 1998; Wanner & Maratsos, 1978), it is plausible that the parser does not maintain non-wh-phrases in memory in the same way as wh-phrases. We investigate these points by examining in detail the processing of coordinated structures involving WhFGDs.

2.2. Active and Reactivated Fillers

Many of the previous studies of wh-dependency processing have either adopted the maintenance view (Fiebach, Schlesewsky, & Friederici, 2002; Gibson, 1998; Gibson & Warren, 2004; Grodner & Gibson, 2005; Wagers & Phillips, 2014; Wanner & Maratsos, 1978; Warren & Gibson, 2002) or the retrieval view (Lewis & Vasishth, 2005; McElree, 2001; McElree & Dosher, 1989; McElree, 2006; Nicol, Fodor, & Swinney, 1994; Nicol & Swinney, 1989; Van Dyke, 2007; Van Dyke & McElree, 2006). The maintenance and retrieval views are mostly motivated on empirical grounds. Storage cost effects (Chen et al., 2005; Gibson & Warren, 2004; Nakatani & Gibson, 2008) and active dependency formation (Stowe, 1986; Traxler and Pickering, 1996; Phillips, 2006), as reviewed earlier, provide motivation for the maintenance view. On the other hand, it has been shown that many effects attributed to storage cost can instead be understood as retrieval effects (Nicenboim et al., 2016). There are also some findings that are not compatible with the expectation (and storage) theories, such as their difficulty in predicting particular distance effects where facilitation is stronger for modifiers of the head of the dependency (Nicenboim, et al., 2016; Vasishth & Lewis, 2006). For example, (4b) should lead to the facilitation in reading times as there are more materials associated with VP (Vasishth &Lewis, 2006).

(4) a. that paper which that boy-ERG saw very old was ‘That paper which that boys saw was very old.’

b. that paper which that boy-ERG table-GEN behind fallen saw very old was. ‘That paper which that boy saw fallen behind a/the table was very old.’ (translation of German to English)

These observations have motivated the retrieval view. We argue that there is a possible mechanism that incorporates both retrieval and maintenance components; this has not been extensively investigated (Fiebach et al., 2002; Wagers & Phillips, 2014). In such a mechanism, maintained information is easier to access and unmaintained information is less accessible for retrieval when a gap is recognized. If some information associated with the filler is maintained and is less susceptible to decay, we expect it to be retrieved easily (Wagers & Phillips, 2014).[3] On the other hand, if some information is susceptible to decay, we expect its retrieval to be more difficult. Another goal of the present study is to uncover the mechanism working behind both the maintenance and retrieval components by testing what aspects of a filler are retrieved in different WhFGD constructions: we refer to these as the "reactivated" WhFGD in (5a) and the "active" WhFGD in (5b).[4]

(5) a. *Reactivated Wh-Filler-Gap Dependency*

[3] As an anonymous reviewer pointed out, Lewis & Vasishth (2005) suggests that retrieval could occur in such a way that the parser can re-instate information into comprehenders' focus of attention, in order to process that information. In this sense, if information were already in comprehenders' focus of attention due to maintenance, there is no need for it to be retrieved. However, following Wagers & Phillips (2014), we argue that comprehenders discharge some components associated with the features from focal attention and this information must be retrieved at when the verb is processed.

[4] Gaps are indicated by an underscore '__' in a sentence.

Which mistake in the **program/programs** __ will be disastrous for the company and certainly __ is/are harmful for everyone involved?

b. *Active Wh-Filler-Gap Dependency*

Which mistake in the **program/programs** [$_{RC}$ that will be disastrous for the company] certainly __ is/are harmful for everyone involved?

In (5a), the wh-filler must be linked to two gaps in the coordinate structure. When a sentence like (5a) is processed, the wh-filler is first linked to the gap in the first conjunct. Before the coordination connective *and* is encountered, the first conjunct can be understood as an independent sentence (*Which mistake in the programs will be disastrous for the company?*), thus the WhFGD can be resolved and interpreted at the point of the first gap. However, when the connective *and* is recognized, the wh-filler needs to be reactivated so that another WhFGD can be formed. This is so because the WhFGDs in the coordination construction obey grammatical constraints known as the Coordinate Structure Constraint (CSC) and the Across-the-Board (ATB) movement restriction (Ross, 1967). Specifically, wh-phrases cannot be extracted from only one conjunct in a coordinate structure, as a single conjunct in the coordinate structure is an island for wh-extraction (Ross, 1967). However, Ross (1967) has shown that wh-extraction from a conjunct is possible when the wh-phrase is extracted from all conjuncts. Thus, as shown in an example (6a), if any conjunct in a coordinate structure contains a gap, then all conjuncts must contain a gap, i.e., the wh-phrase needs to be extracted in an across-the-board (ATB) fashion (Gazdar, Klein, Pullum & Sag, 1985; Ross, 1967; Williams, 1978). If not, the example is unacceptable, as (6b) shows.

(6) a. Which mistake __will be disastrous for the company and certainly __ is harmful for everyone involved?

b. *Which mistake __ will be disastrous for the company and certainly this mistake is harmful for everyone involved?

This suggests that in order to construct a grammatical WhFGD in a coordinated structure, the parser needs to posit the gap in the second conjunct subsequently to the first conjunct, and link the wh-phrase to the gap again in the second conjunct (see Wagers & Phillips, 2009 and Wagers & Phillips, 2014 for related experimental investigations). Thus, it should be the case that when the parser encounters the coordinating connective *and* the wh-phrase must be "reactivated" (*Reactivated Filler*).

On the other hand, (5b) involves a simple WhFGD construction. Although the wh-phrase is modified by a relative clause, the wh-verb dependency is established only at the main verb (the second verb *is/are*).[5] In (5b), the NP, which is the head of the relative clause, is linked to the gap within the relative clause. Thus, a filler-gap dependency is formed. However, unlike in (3a), the first half of the sentence (the wh-NP and the relative clause: *[$_{NP}$ Which mistake in the **program** [$_{RC}$ that will be disastrous for the company]]*) cannot be understood as an independent sentence. Furthermore, even though the head of the relative clause is linked to the gap within the relative

[5] Note, the relative head needs to be linked to the embedded verb, but this is not relevant to the wh-gap dependency formation in terms of wh-question formation.

clause, no WhFGD has been established at the point of the first gap position: the Wh-filler needs to be linked to the gap in the matrix clause for proper interpretation. Assuming that the parser engages in active dependency formation in a case like (5b), we call the wh-filler in (5b) the *Active Filler*.

If, as we have discussed earlier, active fillers are maintained in memory, then it means that they are immediately accessible to the parser to use in online structure building. This means that an active filler should be easier to access, compared to a reactivated filler, at the point of processing the verb. This is because reactivated fillers are released from memory and need to be retrieved when the gap or the verb is recognized. Thus, detailed information from reactivated fillers should be harder to access at the point of processing the verb and completing the whFGD. Consider the difference between (5a) and (5b) from the perspective of online sentence processing. In (5a), the wh-phrase is linked to the gap in the first conjunct, meaning that the wh-gap dependency has been formed and the wh-filler no longer needs to be maintained. This may mean that the wh-filler can be released from memory and no longer impacts memory resources. Subsequently, when the coordinating connective is encountered, the wh-phrase would need to be reactivated. On the other hand, in (5b), the wh-phrase must be linked to the gap in the matrix clause directly. Therefore, the wh-phrase must be maintained until the gap is encountered. If the element that is maintained is retrieved more easily, then we expect that the information associated with wh-filler in (5b) will be retrieved more easily than in (5a).

2.3. How Do We Approach Maintenance and Retrieval?

How can one examine maintenance and retrieval differences between active and reactivated fillers? The current work appeals to the agreement attraction effect, where the local

noun (e.g., a noun other than the head) erroneously licenses agreement (Pearlmutter, Garnsey & Bock, 1999; Wagers et al., 2009, among many others). We use this as a probe to examine what aspects of the filler are retrieved.

One of the important features of agreement attraction is that it is sensitive to grammatical properties of the subject NP that triggers the erroneous agreement relation (Lago, Shalom, Sigman, Lau, & Phillips, 2015, Parker & Phillips, 2017; Tanner, Nicol, & Brehm, 2014; Wagers, Lau, & Phillips, 2009). When the number on the head noun and the verb mismatch, i.e., when grammatical agreement is not established (e.g., *the mistake in the programs *are*), then a clear interference effect from the local noun (*programs*) is typically present. This facilitation in ungrammatical conditions is often called an *Illusion of Grammaticality* (Dillon, Mishler, Sloggett, & Phillips, 2013; Lago et al., 2015; Nicol, Forster, & Veres, 1997; Parker & Phillips, 2017; Pearlmutter et al., 1999; Tanner, Grey, & van Hell, 2017; Tanner et al., 2014; Thornton & MacDonald, 2003; Wagers et al., 2009). When the number of the head noun matches the number of the verb, i.e., when number agreement is grammatical (e.g., *the mistake in the programs is*), there is typically no interference observed from the local noun within the modifier (*programs*), though inhibitory effects are observed in some studies (Acuña-Fariña, Meseguer, & Carreiras, 2014; Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008; Jäger, Engelmann & Vasishth, 2017; Nicenboim, Vasishth, Engelmann and Suckow, 2018; Pearlmutter et al., 1999).

These data suggest that when subject-verb agreement is computed, the parser first computes the agreement relation between the head noun and the verb, and only when this fails, the local noun embedded within the modifier phrase is retrieved. In other words, the initial stage of subject-verb agreement processing is guided by the grammatical structure of the subject NP,

i.e., the parser identifies the head noun and specifically refers to its number information, not the number from other nouns embedded within the subject NP (Phillips, Lau, & Wagers, 2010; Kim, Brehm, & Yoshida 2019). We use this aspect of agreement processing to investigate the extent to which the information on the NP is accessed. If only the category information is maintained and the details about the content of NP are released from the maintenance, then we expect no illusion of grammaticality. On the other hand, if detailed information about the NP (such as information about the head and the modifier) is maintained, then we expect an illusion of grammaticality to be present. With this "selective fallibility" aspect of parsing (Phillips, Wagers, & Lau, 2011) in mind, let us consider the processing of active and reactivated fillers.

If the active-filler is less susceptible to memory decay, and full details about the wh-filler are maintained, we expect parser to be able to access detailed information about the filler when the verb is processed. For example, in (5b) the wh-phrase contains category information (NP), and the representation of the noun head (*mistake*) and the modifier phrases ([$_{PP}$ *in the programs*]). If maintenance of a wh-phrase leads to easier retrieval, all of these pieces of information may be retrieved. If this is the case, then an illusion of grammaticality effect should appear in active filler constructions.

The reactivated filler in (5a), on the other hand, is linked to the gap in the first conjunct, forming a dependency, meaning that the parser no longer needs to maintain the wh-filler. Thus, the wh-filler could be released from maintenance. Given that already-processed elements are susceptible to memory decay (Lewis & Vasishth, 2005; McElree et al., 2003), it is plausible that less detailed information about the filler will be retrieved at the second gap position in the second conjunct. Wagers & Phillips (2014) argued that lexical/semantic information is lost at a long

distance. We could ask what other information is lost, and specifically whether the filler's internal structure remains at a long distance. If the filler is maintained, then the internal structure will be more available for the parser and can lead to an illusion of grammaticality effect. If not, only the category information will be available. If only the category of the filler is retrieved, this would lead simply to a grammatical mismatch effect without the illusion of grammaticality, and interference from the local noun regardless of whether the grammatical subject-verb agreement is established.

Specifically, differences in retrieval and maintenance indexed by the illusion of grammaticality effect are predicted for items involving *Reactivated WhFGD formation* (the filler is linked to the verb once and the wh-filler is reactivated later) and *Active WhFGD formation.*

(7) a. Which mistake in the **program/programs** will be disastrous for the company and certainly is/are harmful for everyone involved? (=5a)

b. Which mistake in the **program/programs** that will be disastrous for the company certainly is/are harmful for everyone involved? (=5b)

Both involve a complex wh-NP, composed of a head noun modified by a prepositional phrase (PP) containing another noun. In both, the wh-phrase serves as the subject of the first and second clause. For the subject-verb agreement dependency to be resolved, the number feature of the verb (i.e., *is*/*are*) in the second clause and the silent gap should agree; differences in processing at the verb in the second clause inform what is maintained versus needs reactivation.

If the parser needs to reactivate the wh-filler again in the second clause, we do not expect

detailed information of the wh-NP to be accessible (this includes the internal structure, including category information and a representation of both the head and the modifier). Thus, when encountering a matrix verb that mismatches the number feature of the head noun, we expect a cost in the ungrammatical conditions, without any agreement attraction.

On the other hand, if information associated with the filler is maintained and thus not susceptible to decay, we expect information about the internal structure to be accessed more easily. The parser may maintain sufficiently-detailed information associated with the filler, including the representation of both the head and the modifier, until the wh-dependency is completed. When the parser encounters a matrix verb (e.g., *are*) that does not license the number feature of the head noun phrase (e.g., *program*), the parser could activate another noun that would fix the number mismatch. Thus, the ungrammatical matrix verb could be erroneously licensed by the local noun *programs,* consistent with the typical agreement attraction effect (Wagers et al., 2009) observed with overt subject noun phrases. If the wh-NP is sufficiently detailed to enable readers to make use of the head vs. non-head information, an agreement attraction effect is expected, and is predicted to be selective to ungrammatical conditions. As such, assuming that stronger maintenance leads to easier retrieval, we expect more detailed information about the filler to be retrieved in (7b) compared to (7a), leading to more agreement attraction for (7b) than for (7a).

2.4. This Study

To address the question of what content is maintained and accessed at the gap, we directly compare the differences in agreement attraction between constructions that involve a relative clause (active filler) and active dependencies based on reactivation (reactivated filler). We conducted three acceptability rating experiments accompanied by three self-paced reading

experiments.

The first two sets of experiments serve the purpose of understanding the processing of the WhFGD within coordinated structures, in order to approach the question of what is maintained and what motivates the maintenance. The purpose of the first experiment is to examine what information is retrieved at the gap in the coordinated structure, testing the hypothesis that Wagers & Phillips (2014) held: in the reactivated filler constructions (i.e., the WhFGD in a coordinated structure), only coarse-grained information of the filler is retrieved (e.g., category information). Agreement attraction serves as a diagnostic for to what extent details about the wh-filler are accessible: If only coarse-grained information such as category is accessible, we expect no agreement attraction. On the other hand, if detailed information of the wh-filler, including the filler's internal structure, is accessible, we expect an illusion of grammaticality.

The second experiment examines what motivates the maintenance of a filler. We compared coordinated structures that involve a wh-filler with ones that do not involve a wh-filler. In coordinated structures involving a wh-filler, like (7a), the reader can recognize that the gap should be inserted in the second conjunct upon encountering the coordinating connective. However, when no wh-element is included and when the subject of the sentence is a simple definite NP (e.g., *The mistake in the program/programs __ will be disastrous for the company and certainly __* is *harmful for everyone involved*), the presence of the filler-gap dependency is not signaled. Thus, the reader can recognize the movement structure only when the gap in the subject position of the second conjunct is recognized. The second experiment shows that there is indeed such a difference between a wh-phrase and a definite NP. This suggests that in wh constructions, the wh-filler is reactivated and made more accessible for the parser at the point that the verb is processed. In other

words, the wh-filler in the coordinated construction is initially released, but is subsequently reactivated and maintained again in memory. In the definite NP construction on the other hand, detailed information about the filler is not maintained, and thus needs to be retrieved at the verb, making it harder to access and leading to less agreement attraction and no illusion of grammaticality. We argue that, if both the wh-phrase and the definite NP were retrieved at the second verb position in the same way, then no such difference should be observed for the illusion of grammaticality effect.

The aim of the third experiment is to examine the role of the maintenance associated with wh-fillers. In a reactivated filler, the wh-gap dependency is completed in the first conjunct, thus, the wh-filler is released from maintenance in the first conjunct. The recognition of the gap in the second clause triggers the retrieval of the wh-element. Assuming that the element released from the maintenance is subject to decay, the reactivated filler is not immediately accessible for the parser when the second verb is processed. On the other hand, in the active filler construction, the wh-filler is maintained in memory and thus it is immediately accessible for the parser when the second verb is processed. As a result, the prediction is for stronger agreement attraction for the active filler than the reactivated filler, as the active filler is better maintained and likely to result in easier accessibility of more information about the internal structure, carried over a long distance.

3. Experiment 1a/1b: Wh-Filler-Gap Dependency in Coordinated Structures

This study investigates what information associated with the wh-filler is retrieved in resolving WhFGD. When the verb is recognized, information associated with the filler should be retrieved from memory. We use the presence of agreement to disclose what information from the filler is retrieved at the verb position. If information regarding the representation of both the head and the

modifier is retrieved, then a verb might agree with the local noun that has been accessed at the verb position. Thus ungrammatical plural verbs after plural local nouns would result in high acceptability ratings and decreased processing difficulty in comparison to ungrammatical plural verbs with singular local nouns. On the other hand, if no sufficiently detailed information of the filler is retrieved, the parser should have difficulty retrieving both the head and the modifier. Thus, ungrammatical verbs would be considered as ungrammatical regardless of the local noun, resulting in low acceptability ratings and increased processing difficulty.

3.1. Experiment 1a: Acceptability Rating

3.1.1. Participants, Materials and Design

Participants were 39 native speakers of English from Northwestern University with no history of reading disorders. All participants provided informed consent and received credit (1 credit/45 minutes) in an introductory Linguistics class.

Critical items consisted of 32 sentence sets arranged in a 2x2 within-subjects factorial design, in which *Local noun* (singular vs. plural) and *Grammaticality* (grammatical vs. ungrammatical) were manipulated as independent factors. A sample set of stimuli is summarized in Table 1. Items were distributed in a pseudo-randomized manner to ensure that participants did not receive the same type of experimental items sequentially. Experimental items were combined with 82 filler sentences with manipulations irrelevant to the experimental items. The experiment took around 30 minutes to complete.

[Table 1 near here]

3.1.2. ***Procedure***

Stimuli were displayed on a desktop PC using the Linger software package (Rohde, 2003). For

each stimulus, participants observed a single sentence on the screen until they pressed the button to move on. After each sentence, they selected a numbered button from 1 to 7, where 1 indicated totally unacceptable and 7 totally acceptable. Ten practice items were presented before the actual experimental items. Participants were instructed that there were no right or wrong answers.

*3.1.3. **Analysis***

Data were analyzed using an ordinal mixed-effects model performed with the *ordinal* package in R version 3.2.3 (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Bates, Maechler, Bolker, & Walker, 2014; Jaeger, 2008). A cumulative logit model was used instead of the linear model as the linear model assumes a continuous and unbounded dependent variable. Each model included simple difference sum-coded fixed effects of *Local noun* (singular vs. plural; contrasts -0.5 and 0.5) and *Grammaticality* (grammatical vs. ungrammatical; contrasts -0.5 and 0.5) and their interactions. All models contained the maximal random effects structure justified by the data (Barr, Levy, Scheepers, & Tily, 2013), including random intercepts for participants and items and random slopes for fixed effects where they converged; the random effects that accounted for the least variance were removed in the case of non-convergence. See model tables for random effect structures.

*3.1.4. **Results & Discussion***

The quantiles of residuals were relatively small and symmetrical about zero (Min: -3.05, Median: 0.05, Max=3.24). Mean acceptability scores are shown in Table 2 and in Figure 1, and the ordinal mixed effect model outputs are shown in Table 3.

[Table 2 near here]

[Figure 1 near here]

[Table 3 near here]

A significant main effect of *Grammaticality* was observed such that ungrammatical items were rated significantly less acceptable than grammatical ones. This was qualified by an interaction between *Local noun* and *Grammaticality* such that the difference between plural local nouns and singular local nouns was larger in ungrammatical conditions. This was further confirmed by a subset analysis, where the main effect of *Local noun* was larger in ungrammatical ($\beta = 0.51$, SE= 0.14, t= 3.75, $p < 0.001$) than in grammatical conditions ($\beta = -0.30$, SE= 0.14, t= -2.17, $p < 0.05$) and grammatical sentences were read significantly slower than ungrammatical sentences. This observed illusion of grammaticality provides evidence for the retrieval of grammatical properties, such as information about the internal structure such as the head and the modifier, in reactivated WhFGD constructions[6].

3.2. Experiment 1b: A Self-paced reading Experiment

3.2.1. ***Participants, Materials and Design***

Participants were 58 native speakers of English from Northwestern University with no history of reading disorders. All participants provided informed consent and received credit (1 credit/ 45 minutes) in an introductory Linguistics class. Nine participants were excluded due to very low accuracy (<65%) in answering questions after each stimulus.

---

[6] As an anonymous reviewer pointed out, decay should have less impact on the offline experiments as readers can look back at the left context anytime, to remember the content of the antecedent. Our purpose of the offline experiments was to understand how the availability of the contexts can influence the retrieval of different kinds of information.

The same critical items were used as in Experiment 1a. Items were distributed in a pseudo-randomized manner to make sure that participants did not receive experimental items from the same experiment more than once in a row. Two experimental items were excluded from the analysis due to typographical errors. The experimental items were combined with 74 filler sentences of similar complexity. Fillers included items related to ambiguity resolution, passive sentences and locative constructions, all of which are irrelevant to processing either agreement attraction or coordinate structures.

3.2.2. Procedure

Stimuli were displayed on a desktop PC using the Linger software package (Rohde, 2003). A self-paced word-by-word moving window paradigm (Just, Carpenter, & Woolley, 1982) was employed. Participants saw a row of dashes, masking the words in the sentence. Participants pressed the space bar to proceed to the next sentence. After reading each sentence, they were asked to answer comprehension questions. To answer comprehension questions, participants were asked to press F (yes) or J (no) keys. An example comprehension question is "Was the word drawer mentioned in the story?". They were provided with immediate feedback in terms of their accuracy. Six practice items were given to participants at the beginning of the experiment. The experiment took each participant about 30-45 minutes to complete.

*3.2.3. **Analysis***

Data were analyzed using linear mixed effect regression, performed with the *lme4* package in R version 3.2.3 (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Bates, Maechler, Bolker, & Walker, 2014; Jaeger, 2008). Reading times were log-transformed to minimize non-normality (Box & Cox, 1965; Vasishth & Chen, Li & Guo, 2013) and data that fell outside 2.5 standard

deviations from the overall mean for the each region was excluded from the analysis. The critical regions were the verb, the following word (spillover region 1) and one word following the spill over region 1 (spillover region 2). The by-region exclusion percentages due to outlier removal were 1.73 % (verb region), 2.59 % (spillover region 1), and 1.5% (spillover region 2).

3.2.4. Results & Discussion

Figure 2 shows region-by-region reading times, Figure 3 shows the interaction plot at the critical region (spillover region 1), and Table 4 shows the mixed effect model. Mean accuracy for critical trial comprehension questions was 72.0%.

[Figure 2 near here]
[Figure 3 near here]
[Table 4 near here]

At the spillover region 1,[7] a marginal main effect of *Local noun* was observed, such that items paired with singular local nouns were read significantly slower than their plural counterparts. This was qualified by an interaction between *Local noun* and *Grammaticality,* such that constructions with singular local nouns were read slower than those with plural local nouns in ungrammatical

[7] As an anonymous reviewer has pointed out, it is possible that the absence of an agreement attraction at the verb might be due to the nature of the self-paced reading experiment. It has been well known that in self-paced reading experiments, the expected effect can be observed in one or two regions after the critical region (the spill-over effect; Vasishth, 2006). Therefore, it is possible that, even if the agreement attraction effect is caused at the verb region, it would not be observed right on the verb region but in spill-over regions.

conditions but no differences were detected in grammatical conditions. Subset analyses confirmed a main effect of *Local noun* (β = -0.04, SE= 0.02, t=-2.77, p <0.01) in ungrammatical conditions, which was absent in grammatical conditions (β = 0.02, SE= 0.02, t= 0.36).[8] This again shows an illusion of grammaticality effect that provides evidence for the retrieval of grammatical properties in processing reactivated WhFGD constructions.

3.2.4. Discussion

We investigated what information associated with the filler is retrieved from memory in resolving reactivated WhFG dependencies. Ungrammatical sentences that included plural verbs resulted in high acceptability ratings as well as in decreased reading time, in comparison to ungrammatical singular verbs, eliciting an illusion of grammaticality similar to that seen in overt sentences (Lago et al., 2015; Parker et al., 2016; Tanner et al., 2014). This suggests that grammatical information of the wh-filler is retrieved, including the representation regarding the head and the modifier (*which mistake in the programs*), allowing the verb to erroneously agree with the local noun as a last resort. In contrast, if detailed information associated with the antecedent had not been recovered, all ungrammatical verbs would have been processed similarly, with no amelioration and

[8] Following an anonymous reviewer's suggestion, we also examined the region immediately preceding the verb (i.e., the pre-critical region). The results showed a main effect of *Grammaticality* (β = -0.03, SE= 0.01, t=-2.49, p <0.05) but no main effect of *Local noun* (β = 0.00, SE= 0.01, t=0.07, p >0.05) as well as no interaction between *Local noun* and *Grammaticality* (β = 0.04, SE= 0.02, t=1.49, p >0.05). This further suggests that the effects we observe are not due to spillover effects from the prior regions.

reading time facilitation by a local plural noun.

It is possible that rather than a pure maintenance view, it is the presence of the coordinating connective *and* that triggers the reactivation and maintenance of the wh-filler and the active dependency formation. In other words, while retrieval happens at the gap, how much information is retrieved depends on how accessible the information is. The agreement attraction at spillover region 1 indicates that the grammatical and lexical content of the wh-NP are readily reactivated once the verb is processed. However, the lack of attraction at the verb region suggests that differences between conditions appear after processing the verb, and after processing the gap.

Our results are less compatible with the view that only the category information of the filler is accessible at the verb position. If only the category information were accessible, we would not expect agreement attraction to be present. The results are compatible with the view that the whole NP including category information (e.g., NP) and grammatical information (information about the internal structure; the representation regarding the head and the modifier) are retrieved, leading to an agreement attraction effect at the verb region.

4. Experiment 2a/2b: Wh-filler vs. Definite NP

Experiment 1 showed that readers retrieve detailed category and grammatical information, including the internal structure of the noun head and its modifier phrase. This led to an illusion of grammaticality effect. In the current experiment, we compare coordinated structures that involve a wh-filler, (8a) with those that do not involve a wh-filler, (8b).

(8) a. Which mistake in the program/programs __ will be disastrous for the company and certainly__ is harmful for everyone involved?

b. The mistake in the program/programs __ will be disastrous for the company and

certainly__ is harmful for everyone involved.

One major difference between the two types of coordinated construction is that the former involves a wh-element that can signal the presence of the filler-gap dependency prior to encountering the gap.[9] Therefore, in the wh construction, the presence of the filler-gap dependency is recognized immediately upon encountering the wh-phrase and thus the parser can compute any grammatical constraints that apply to the WhFGDs such as CSC and the ATB restriction. If Wagers and Phillips (2009, 2014) are correct, then this means that the wh-filler can be reactivated upon encountering the coordinating connective *and*. On the other hand, the definite NP subject (e.g., *the mistake in the program/s*) does not signal the presence of a filler-gap dependency, and thus the coordinating connective should not reactivate the definite NP subject. As the presence of the filler-gap dependency is recognized when the gap in the second conjunct is recognized, the recognition of

[9] An anonymous reviewer suggested that CSC/ATB constraints are not the only relevant assumptions. It could be the case that the readers recognize the presence of the gap due to the combination of the coordinating connective, *and*, and an adverb. If the combination of the coordinating connective and an adverb (... *and certainly* ...) helps reactivate the filler, then our assumption must be weakened, i.e., the reactivation of the filler is not due to the grammatical constraints. However, as Wagers & Phillips (2009) showed, the gap in the coordinated structure and parasitic gap within an adjunct clause, which is optional, show different reactivation profiles. Therefore, it is still plausible to that ATB/CSC plays a role in the reactivation of the wh-filler. As we do not have any evidence to distinguish the two hypotheses, we would like to leave this point open at this point.

the gap and the retrieval of the subject NP in the first conjunct may occur at the same time. The prediction is that the definite NP subject, should not be reactivated by the coordinating connective. Thus, retrieving a definite NP subject at the gap position in the second conjunct could be more difficult than retrieving the wh-filler, leading to a reduced agreement attraction effect.

4.1. Experiment 2a: Acceptability Rating

4.1.1. Participants, Materials and Design

Participants were 39 native speakers of English from Northwestern University with no history of language disorders. All participants provided informed consent and received credit (1 credit/ 45 minutes) in an introductory Linguistics class.

Critical items consisted of 32 sentence sets arranged in a 2x2x2 within-subjects factorial design, in which *Local noun* (singular vs. plural), *Grammaticality* (grammatical vs. ungrammatical) and *Filler type* (the definite NP vs. wh-filler) were manipulated as independent factors. A sample set of stimuli is summarized in Table 5. Items were distributed in a pseudo-randomized manner to ensure that participants did not receive the same type of experimental conditions sequentially. The experimental items were combined with 48 filler sentences, with manipulations irrelevant to the current experiment. The experiment took around 30 minutes to complete.

[Table 5 near here]

4.1.2. Procedure

The similar procedure was employed as with Experiment 1a.

4.1.3. ***Analysis***

A similar analysis was employed as in Experiment 1a. Each model included simple difference sum-coded fixed effects of *Local noun* (singular vs. plural; contrasts -0.5 and 0.5), *Grammaticality* (grammatical vs. ungrammatical; contrasts -0.5 and 0.5), *Filler type* (the definite NP vs. wh-filler; contrasts 0.5 and -0.5) and their interactions. All models contained the maximal random effects structure justified by the data (Barr, Levy, Scheepers, & Tily, 2013), including random intercepts for participants and items and random slopes for fixed effects where they converged; see model tables for random effect structures.

4.1.4. ***Results & Discussion***.

The quantiles of residuals were relatively small and symmetrical about zero (Min: -3.40, Median: -0.04, Max=3.68). Mean acceptability scores are shown in Figure 4 and Table 6, and ordinal mixed effect model outputs are shown in Table 7.

[Figure 4 near here]

[Table 6 near here]

[Table 7 near here]

Local noun, *Grammaticality*, and *Filler type* were all significant as main effects. A main effect of *Local noun* was observed, such that items with singular local nouns were rated lower than those with plural local nouns. A main effect of *Grammaticality* was observed such that ungrammatical items were rated as significantly less acceptable than grammatical ones. Finally, a main effect of *Filler type* was observed, such that items with the wh-filler were rated as significantly less acceptable than those containing the definite NP.

These main effects were qualified by an interaction between *Local noun* and

Grammaticality such that constructions with singular local nouns were rated less acceptable than those with plural local nouns, in the ungrammatical conditions only. This was further supported by subset analyses which confirmed a main effect of *Local noun* (β = 0.71, SE= 0.18, t= 3.87, p <0.001) in ungrammatical conditions but not in grammatical conditions (β = -0.19, SE= 0.15, t= -1.29, p >0.05). An interaction between *Filler type* and *Grammaticality* was also observed such that Definite NP *Filler type*s were judged to be significantly less acceptable than Wh-*Filler type*s in grammatical sentences only. This was confirmed with a subset analysis that revealed a main effect of *Filler type* (*β = 0.70, SE= 0.18, t= 3.87,* p<0.001) in grammatical conditions only. There were no interactions observed between *Local noun* and *Filler type*, or between *Local noun*, *Filler type*, and *Grammaticality*.

The pattern of increased acceptability for ungrammatical verbs following local plural nouns regardless of filler type indicates an illusion of grammaticality: ungrammatical definite NPs and Wh-Fillers are considered equally acceptable in offline ratings, despite the increase in acceptability for grammatical definite NPs over Wh-Fillers.

4.2. Experiment 2b: Self-paced reading Experiment

4.2.1. Participants, Materials and Design

Participants were 81 native speakers of English from Northwestern University with no history of language disorders. All participants provided informed consent and received credit (1 credit/ 45 minutes) in an introductory Linguistics class. Six subjects were excluded due to their very low accuracy in answering comprehension questions about the sentences (<60%).

Critical items consisted of 32 sentence sets arranged in a 2x2x2 within-subjects factorial design, in which *Local noun* (singular vs. plural; contrasts 0.5 and -0.5), *Grammaticality* (grammatical vs. ungrammatical; contrasts -0.5 and 0.5) and *Filler type* (wh-filler vs. definite NP; contrasts 0.5 and -0.5) were manipulated as independent factors. Items were distributed in a pseudo-randomized manner to ensure that participants did not get two experimental items of the same type in a row. The experimental items were combined with 48 filler sentences irrelevant to the current experiment.

4.2.2. Procedure

A similar procedure was employed as with Experiment 1b.

4.2.3. Analysis

The same factors and contrasts were used as in Experiment 2a. The rest of the analysis mirrored Experiment 1b, with the critical regions of the verb, the following word (spillover region 1) and one word after the spill over region 1 (spillover region 2). The by-region exclusion percentages due to outlier removal was 1.43 % (verb region), 1.89 % (spillover region 1), and 1.74% (spillover region 2).

4.2.4. Results & Discussion

Region-by-region reading times for Wh-Filler conditions are presented in Figure 5, the Definite NP conditions are presented in Figure 6, the interaction plot at the critical region in Figure 7, and mixed effect model outputs are presented in Table 8. Mean accuracy for critical trial comprehension questions was 81.0%.

[Figure 5 near here]

[Figure 6 near here]

[Figure 7 near here]

[Table 8 near here]

At the verb region, a main effect of *Grammaticality* was observed, such that ungrammatical constructions were read slower than the grammatical constructions. This was qualified by the critical interaction between *Local noun* and *Grammaticality,* such that the difference between plural nouns and singular local nouns was larger in ungrammatical constructions. A planned subset analysis showed that this interaction between *Local noun* and *Grammaticality* was significant only in the wh-filler NP (β = *-0.06, SE= 0.02, t= 2.53,* $p<0.05$) but not in the definite NP (β = *-0.03, SE= 0.02, t= -1.15,* $p>0.05$), indicating that the illusion of grammaticality was at least numerically driven by the reactivated wh-filler conditions, although the three-way interaction failed to reach significance.

At the spillover region 1, an interaction between the *Grammaticality* and the *Filler type* was observed such that the differences between the definite NP and the wh-filler were larger in grammatical conditions (β = *0.03, SE= 0.01, t= 2.46,* $p<0.05$), indicating that the definite NP was read significantly slower than the wh-filler in grammatical conditions.

At the spillover region 2, the critical interaction between *Local noun* and *Grammaticality* was again observed, such that the differences between plural local nouns and singular local nouns were larger in the ungrammatical conditions. A subset analysis confirmed that this was carried by a marginal main effect of *Local noun* in ungrammatical conditions (β = *-0.05, SE=*

0.02, t= -1.93). An interaction between *Local noun* and *Filler type* was also observed. Further subset analysis revealed no main effect of *Local noun* in the wh-filler (*β = -0.15, SE= 0.02, t= -0.52*) but a marginal main effect of Local noun in the definite NP (*β = 0.02, SE= 0.01, t= 1.63*). This indicates that the singular local noun was read faster than the plural local noun in the definite NP.

4.2.5. Discussion

In this experiment, we tested whether coordination leads wh-NPs and definite NPs to be reactivated similarly at the gap (the verb) in the second conjunct. Although the three-way interaction did not reach significance, the results of the planned subset analysis are compatible with the idea that attraction was reduced for definite conditions relative to wh-fillers, suggesting that details about the grammatical information of the definite NP might not be retrieved at the verb. Assuming that this is correct, we argue that these differences in attraction are due to differences in how these two kinds of fillers are processed. While the wh-filler should be reactivated at the coordinating connective and put into maintenance again, this should not occur for the definite NP. This can be understood by considering the time-course of processing the verb in the second conjunct.

In the definite NP condition, when the reader encounters the coordinating connective *and*, the parser may expect a clausal conjunct which involves an overt subject and a verb, or another NP, due to the local attachment bias (Frazier & Clifton 2005; Staub & Clifton 2006). If so, when the parser encounters the verb, the parser needs to abandon this expected structure and build a structure with a subject gap. Because the gap is not expected upon encountering the coordinate structure, the parser could posit a gap only after the bottom-up evidence (the verb) is encountered leading to a reanalysis. In other words, in definite NP sentences, the definite NP itself does not

signal the presence of a filler-gap dependency and the coordinating *and* does not provide a cue to actively complete the dependency: the parser does not maintain the definite NP subject. The lack of a significant illusion of grammaticality in the definite NP conditions is plausibly due to the fact that the information associated with the definite NP was not maintained and thus is subject to memory decay. At the same time, this may be due to that the reanalysis difficulty that we have mentioned above. In other words, the reanalysis processes and the reactivation might happen at the same point (at the verb), and thus we may not be able to observe the effect of reactivation or the reanalysis effect could hide the reactivation effect.

In contrast, the presence of a significant interaction indicating the illusion of grammaticality in the wh-filler conditions suggests that the detailed information from the wh-filler was readily accessible at the second verb position. This observation leads to the following conclusions. First, it is possibly the case that grammatical constraints such as ATB movement restriction and CSC in the coordinate structures could lead the parser to the formation of the wh-dependencies in the second conjunct. If the parser is sensitive to the ATB extraction from coordinate phrases, upon encountering the coordinating connective *and*, the parser would be sensitive to the constraints on WhFGD formation in the context of coordinate structures, such as CSC and the. ATB restriction. These constraints lead to actively searching for the gap in the second conjunct (Wagers & Phillips, 2009, 2014), which could lead to the more robust illusion of grammaticality effect in the wh-filler condition.

Assuming that there is a genuine processing difference between the wh-construction and the definite NP construction, then this suggests that the combination of the wh-filler and the coordinate structure is crucial. This means either that the wh-filler should be affected by the

presence of the coordinating connective or that the processing of the wh-construction fundamentally does not involve the reanalysis process that would mask the illusion of grammaticality effect. If the lack of the illusion of grammaticality effect in the definite NP constructions is due to a lack of reactivation of the definite NP, then the presence of the illusion effect in the wh-filler construction should be due to the reactivation of the wh-filler by the coordinating connective. On the other hand, if the lack of the illusion effect in the definite NP construction is due to reanalysis (the parser initially expected an NP-conjunct after the coordinated connective and had to change the structure to the clausal conjunct with a gap), then, in the wh-filler construction, such reanalysis process should not have taken place. We contend that the reanalysis hypothesis predicts that the adverb or the verb in the second conjunct should be read slower in the definite NP conditions than in the Wh-filler conditions because the adverb or the verb disambiguate the structure and therefore trigger reanalysis. As has been long known, reanalysis incurs a processing cost (Frazier 1987; Schneider & Phillips, 2001; Sturt et al., 2001). Therefore, if reanalysis takes place in the definite NP conditions, masking the agreement attraction effect, then we expect slower reading of the verb and/or the adverb in the second conjunct in the definite NP conditions than in the Wh-filler conditions. In our data, this effect was not borne out, with no main effect of filler type in either region (Adverb: β = *-0.00, SE= 0.01, t= -0.36,* p>0.05; Verb: β = *0.01, SE= 0.01, t= 1.16,* p>0.05). This suggests against the reanalysis hypothesis. Therefore, we conclude that it is more likely that the wh-phrase is reactivated at the connective position and put into maintenance again.

In contrast to the pattern observed in the online data, note that in the offline rating exepriment (Experiment 2a), we observed clear evidence for agreement attraction in the definite

NP as well as the Wh-filler conditions. We argue that this discrepancy may arise from the availability of the contexts for the offline rating experiment: readers had more time to go back and read the first conjunct in the rating experiment, leading to an agreement attraction effects. For the online experiment (Experiment 2b), we argue that the parser recognizes these grammatical constraints in real-time, leading to an expectation of the upcoming gap position upon encountering the coordinating connective, and actively linking the wh-filler and the subsequent gap site. Thus, when the reader encounters the connective and the verb sequence, the parser could readily reactivate the wh-filler and the wh-filler is maintained. If some information about the filler is more accessible and less susceptible to decay, we expect information to be retrieved easily (Wagers & Phillips, 2014). Thus, reactivation at the coordinating *and* could suggest that the parser retrieves detailed information at the verb. This could lead to retrieval of fine-grained information at the gap, such that the plural local noun is read faster than the singular local noun in ungrammatical conditions.

Another possibility for the differences between the retrieval of the definite NP and the wh-filler is that they could behave differently in terms of encoding. Wh-words could be intrinsically more prominent than the definite NP because they have special morphology, function and semantics (Jäger et al., 2017). Although this is indeed a possibility, we have to note that it is difficult to distinguish the effects of prominence from maintenance.

5. Experiment 3a/3b: Active Filler vs. Reactivated Filler

The results of the previous experiments showed that reactivation of fillers could not be the sole cause of agreement attraction. In this experiment, we ask how active versus reactivated wh-fillers may differ in processing. We compare how the information retrieved at the matrix verb (*is/are*)

could differ by changing the dependency configuration as in (9).[10]

(9) a. Which mistake in the program/programs will be disastrous for the company and certainly is harmful for everyone involved? (=5a)

b. Which mistake in the program/programs that will be disastrous for the company certainly is harmful for everyone involved? (=5b)

As we noted earlier, (9a) involves the parser posting a gap in the second conjunct subsequent to the first conjunct, and linking the wh-phrase to the gap in the second conjunct. This indicates that when the parser encounters the coordinating connective *and*, the wh-phrase must be

[10] As an anonymous reviewer pointed out, there is an alternative explanation for Experiment 3 that would not rely on the reactivated vs. active distinction, but rather, on differences in cue-based retrieval. In (3b), the attachment site of the RC is actually ambiguous, such that "that will be disastrous" could modify either "mistake" or "program(s)". If readers prefer to attach the RC low, to "program(s)", then according to cue-based retrieval this noun phrase will be reactivated, rendering it more active in memory. This would yield stronger attraction rates at the main verb ("is/are"), since the local noun will have higher activation (and thus interfere more) in (3b) than in (3a). However, if the attachment of RC modulates the accessibility of the lower noun, we also predict a similarity-based interference effect. In other words, the local noun should be more accessible across-the-board and thus should give rise to an interference effect whether the agreement is grammatical or ungrammatical. This should not predict the illusion of grammaticality we observed, but rather an agreement attraction effect in both grammatical and ungrammatical conditions.

reactivated. If the release from maintenance is subsequently followed by retrieval of decayed information, then we expect that the wh-filler will not be immediately accessible for the parser. This would suggest weaker agreement attraction. Conversely, (9b) involves an active filler where the wh-filler needs to be maintained until the matrix verb in order to resolve the dependency. If the parser could avoid the release from maintenance, we expect that detailed information associated with the wh-filler will be accessible for the parser, leading to stronger agreement attraction in (9b).

5.1. *Experiment 3a: Acceptability Rating Task*

5.1.1. Participants, Materials and Design

Participants were 43 native speakers of English from Northwestern University with no history of language disorders. All participants provided informed consent and received credit (1 credit/ 45 minutes) in an introductory Linguistics class.

Critical items consisted of 32 sentence sets arranged in a 2x2x2 within-subjects factorial design, in which *Local noun* (singular vs. plural) and *Grammaticality* (grammatical vs. ungrammatical) and *Dependency type* (Active Filler vs. Reactivated Filler) were manipulated as independent factors. A sample set of stimuli is summarized in Table 9. Items were distributed in a pseudo-randomized manner to ensure that participants did not receive same type of experimental conditions sequentially. The experimental items were combined with 88 filler sentences, irrelevant to the current experiment. The experiment took around 30 minutes to complete.

[Table 9 near here]

5.1.2. Procedure

A similar procedure was employed as in Experiment 1a.

5.1.3. Analysis

The same analysis was employed as Experiment 1a. Each model included simple difference sum-coded fixed effects of *Local noun* (singular vs. plural; contrasts: -0.5 and 0.5), *Grammaticality* (grammatical vs. ungrammatical; contrasts: -0.5 and 0.5), *Dependency type* (active filler vs. reactivated filler; contrasts: -0.5 and 0.5) and their interactions. All models contained the maximal random effects structure justified by the data (Barr, Levy, Scheepers, & Tily, 2013), including random intercepts for participants and items and random slopes for fixed effects where they converged; see model tables for random effect structures.

5.1.4. Results & Discussion

The quantiles of residuals were relatively small and symmetrical about zero (Min: 3.48, Medi an: 0.01, Max=4.00). Mean acceptability scores are shown in Table 10 and Figure 8, and an ordinal mixed effect model outputs are shown in Table 11.

[Table 10 near here]

[Figure 8 near here]

[Table 11 near here]

All three factors disclosed main effects. A main effect of *Local noun* was observed such that items with singular local nouns were rated lower than their plural counterparts. A main effect of *Grammaticality* was observed such that ungrammatical items were rated significantly less acceptable than grammatical ones. Finally, a main effect of *Dependency type* was observed such that items with active Fillers were rated significantly less acceptable than those containing the reactivated Fillers.

These main effects were qualified by an interaction between *Local noun* and

Grammaticality such that items containing singular local nouns were rated less acceptable than those with plural local nouns in ungrammatical conditions, but the same in grammatical conditions. This was further supported by a main effect of *Local noun* ($\beta = 0.67$, SE= 0.10, z= 6.50, $p < 0.001$) and a main effect of *Dependency type* ($\beta = -0.63$, SE= 0.10, z=-6.10, $p < 0.001$) in ungrammatical but not grammatical conditions. This indicates an illusion of grammaticality effect consistent with agreement attraction.

However, an interaction between *Dependency type* and *Grammaticality* was also observed such that the differences between the active filler and reactivated filler were larger in grammatical sentences ($\beta = -1.32$, SE= 0.31, z=-3.98, $p < 0.001$) than in ungrammatical sentences ($\beta = -0.54$, SE= 0.25, z=-2.14, $p < 0.05$), suggesting that when considered in light of the grammatical sentence baseline, reactivated filler sentences elicit relatively more agreement attraction, with a reduced difference between the grammatical and ungrammatical plural conditions in the active filler (M = 0.05) than the reactivated filler conditions (M= 0.54). An interaction between *Local noun* and *Dependency type* was also observed such that items containing singular local nouns were judged less acceptable than those containing plural local nouns in the Reactivated Filler condition ($\beta = 0.36$, SE= 0.16, z= 2.18, $p<0.05$), as well as, marginally, in the Active Filler condition ($\beta = 0.34$, SE= 0.19, z= 1.74, p=0.08). Finally, the three-way interaction between *Local noun*, *Filler type*, and *Grammaticality* did not reach significance.

In combination, these results show evidence for attraction in an offline measure for both active and reactivated wh-fillers. The results are consistent with the idea that the difference between the two types of filler was stronger in grammatical than ungrammatical conditions.

5.2. Experiment 3b: A Self-paced reading Experiment

5.2.1. Participants, Materials and Design

Participants were 76 native speakers of English from Northwestern University with no history of language disorders. All participants provided informed consent and received credit (1 credit/ 45 minutes) in an introductory Linguistics class.

Critical items were similar to Experiment 3a. Items were distributed in a pseudo-randomized manner to ensure that participants did not receive same type of experimental conditions sequentially. The experimental items were combined with 64 filler sentences of similar length. The experiment took around 30 minutes to complete.

5.2.2. Procedure

As in Experiment 1b.

5.2.3. Analysis

Factors are as described in Experiment 3a. The analysis was conducted as described in Experiment 1a. The by-region exclusion percentages due to outlier removal was 2.3 % (verb region), 3.16 % (spillover region 1), and 2.76 % (spillover region 2).

5.2.4. Results & Discussion

Region-by-region reading times for Active Filler conditions are presented in Figure 9, the Reactivated Filler conditions are presented in Figure 10, and interaction plots for spillover regions 1 and 2 are presented in Figure 11 and 12 respectively. Mixed effect model outputs are presented in Table 12. Mean accuracy for critical trial comprehension questions was 84.0%.

[Figure 9 near here]

[Figure 10 near here]

[Figure 11 near here]

[Figure 12 near here]

[Table 12 near here]

At the verb region, a main effect of *Dependency type* was observed such that such that items with the active filler were read significantly slower than those containing the reactivated filler.

At the spillover region 1, a main effect of *Grammaticality* was observed such that ungrammatical sentences were read significantly slower than grammatical sentences. This was qualified by an interaction between *Grammaticality* and *Local noun*, and an interaction between *Grammaticality, Local noun*, and *Dependency type.* Further subset analysis suggest that these differences were driven by the active filler dependency condition, which showed a significant main effect of *Grammaticality* ($\beta = 0.04$, $SE = 0.01$, $t = 2.70$, $p<0.05$) and an interaction between *Local noun* and *Grammaticality* ($\beta = -0.08$, $SE = 0.02$, $t = -3.18$, $p<0.001$). In contrast, for the reactivated filler, there was a marginal main effect of Grammaticality ($\beta = 0.02$, $SE = 0.01$, $t = 1.84$, $p<0.05$) but no significant interaction between *Local noun* and *Grammaticality* ($\beta = -0.01$, $SE = 0.02$, $t = -0.41$, $p>0.05$). This indicates more agreement attraction for active versus reactivated wh-fillers at the spillover region 1. Importantly, we found an interaction between *Local noun* and *Dependency type* in the ungrammatical conditions ($\beta = -0.05$, $SE = 0.02$, $t = -2.69$, $p<0.05$) but not in grammatical conditions ($\beta = 0.01$, $SE = 0.02$, $t = 0.39$, $p<0.05$).

At the spillover region 2, the critical interaction between *Local noun* and *Grammaticality* was observed, such that the differences between plural and singular local nouns were larger in the ungrammatical conditions than in the grammatical conditions. We also report the reading times at the adverb and at the verb. At the adverb and the verb, there was a main effect of *Dependency type*, such that active fillers were read significantly slower than the reactivated fillers (Adverb: $\beta = 0.14$,

SE= 0.02, t= 8.61, p<0.05; Verb: $\beta = 0.03$, *SE= 0.01, t*= 2.38).

5.2.5. Discussion

The current experiment addresses the question of differences between wh-fillers that are linked to the gap in the matrix clause verb directly (active filler) versus wh-fillers linked to the gap in the first conjunct and subsequently reactivated in the coordinate structure. Offline acceptability results show that the interaction between *Local noun* and *Grammaticality* was numerically larger in the active filler conditions, relative to the reactivated filler conditions. We also observed that in reading time measures, agreement attraction was significantly larger for the active filler than the reactivated filler in spillover region 1, as indexed by the three-way interaction in this region; however, both filler types led to attraction in the following region (spillover region 2), with a two-way interaction between *Local noun* and *Grammaticality*. This suggests, although both the reactivated filler and active filler may lead to an agreement attraction effect, the effect was stronger for active fillers and manifested at an earlier stage, than it did for the reactivated fillers.

We have further observed that the second verb and the adverb preceding the second verb were read significantly more slowly in the active filler conditions than in the reactivated filler conditions. We contend that this means that the active Filler was maintained in memory. As we have discussed earlier, one of the motivations for the maintenance of the filler is the integration cost effect (Gibson 1998; Grodner & Gibson, 2005; Warren & Gibson, 2002). The observation that the adverb and the verb are read significantly slower in the active filler conditions than the reactivated filler conditions is the following. The active filler caused a larger integration cost because it was maintained in memory for a long distance and it has been observed that the longer dependency gives rise to the more processing cost at the end of the dependency due to the

integration cost (Gibson 1998; Grodner & Gibson, 2005; Warren & Gibson, 2002). The verb region is where the whFGD is completed. Furthermore, the adverb can clearly signal the presence of the verb and thus the parser can expect that the verb which can terminate whFGD is upcoming. As a result, as early as the adverb position, the parser can recognize that the whFGD is being completed, leading to an integration cost at this point. The reactivated filler, on the other hand, was released from memory, and then reactivated and put into maintenance again at the coordinating connective. The distance between the point where the wh-filler was reactivated (coordinating connective, *and*) and the point where the whFGD is completed (i.e., the second verb position) was short. Therefore, the integration cost should be smaller accordingly.

If we only assume that retrieval plays a role, we would not predict such difference at the second verb position, as both in the active filler and the reactivated filler conditions, the wh-filler should be retrieved at the second verb position, and the distance between the point where the wh-filler is recognized and the second verb where the wh-filler is to be reactivated are basically the same.

As suggested earlier, the differences in the strength and the timing of agreement attraction could be due to whether or not the parser has previously released the wh-filler from maintenance. For the active filler, information associated with the wh-filler is well preserved because the filler has not been released from maintenance and subsequently reactivated. The maintenance of the wh-filler could make available the detailed information of the wh-filler where the parser could access both the head and the modifier, leading to stronger agreement attraction when there is a mismatch between the verb and the head noun but a match between the head and the local noun. On the other hand, for reactivated fillers, the parser releases the wh-filler from memory and subsequently

reactivates the wh-filler by means of the coordinating connective or the recognition of the gap. Therefore, given that the wh-filler is released from maintenance at an early point in the sentence, the released wh-filler is subject to memory decay. We then expect that the wh-filler is less accessible compared to the active filler and thus the information associated with the filler is not accessible for the parser when the second verb is processed. As a result, the structure of the wh-filler, including information about the head noun and the modifier is less accessible, leading to a lower degree of the agreement attraction in the ungrammatical constructions, and a delay in the timing of attraction as the filler is reactivated in processing.

6. General Discussion

In three experiments, we examined how the wh-filler is maintained and accessed in two WhFGD configurations. These studies argue for a processing architecture that incorporates both maintenance and retrieval components. Our assumption is that if information about the filler is maintained, and less susceptible to decay, it will be accessed easily when the verb is processed (Wagers & Phillips, 2014). On the other hand, if some information from the filler is susceptible to decay because it is released from active maintenance, we expect it to be less accessible for the parser. Differences in what is accessible at the verb lead to differences in agreement attraction for different types of wh-fillers.

The first experiment tested WhFGD within coordinated structures, in order to examine what information about the wh-filler is accessed at the verb region. According to Wagers & Phillips (2014), information about the category of the wh-filler is maintained throughout the dependency formation process, but thematic and semantic information is not. We investigated whether only category information is maintained, or if details about the content of NP are released from

maintenance

Within the coordinated structure, the wh-filler can be linked to the gap in the first conjunct, and can thus be released from memory. However, the wh-filler should be reactivated when the coordinating connective *and* is processed, due to the CSC and ATB restriction. The results showed that the verb was read faster in the ungrammatical plural local noun conditions than the ungrammatical singular local noun conditions, i.e., we observed an illusion of grammaticality effect. Thus, detailed information associated with the filler (i.e. grammatical information) is readily accessed at the verb, for reactivated wh-fillers.

In the second experiment, we compared definite subject NPs with reactivated wh-fillers, in order to understand what motivates the maintenance of an element. In a coordinated structure involving a definite NP in the subject position, the presence of a filler-gap dependency is not signaled, and thus the coordinating connective does not initiate the parser to form a filler-gap dependency in the second conjunct. Thus, until the gap in the second conjunct is encountered, the parser should not construct the structure that involves the filler-gap dependency. Only by recognizing the gap in the second conjunct does the parser register that the definite NP is part of a filler-gap dependency. Thus in this configuration, the parser should not initially register that a filler-gap dependency is involved, and therefore the parser needs to reanalyze the structure as such. As the filler-gap dependency is not constructed initially, the definite NP should not be maintained in memory and, therefore, should be subject to memory decay. The results of Experiment 2 are compatible with the view that a wh-element is different from a definite NP with respect to retrieval; attraction effects are obtained at the main verb, supporting the idea that memory encoding of the filler includes richer information than just its category information (c.f., Wagers & Phillips, but

see also Chow et al., 2018).[11] A numerically weaker illusion of grammaticality effect was observed in the definite NP conditions than in the wh-filler conditions, which follows from the premise that the definite NP is not maintained in memory and needs to be reactivated upon encountering the coordinative connective *and.* However, we did not find a three-way interaction between *Local noun*, *Grammaticality*, and *Filler Type*, which is expected if the accessibility to the wh-filler is different. This means that the magnitude of the attraction effect did not significantly differ depending on the type of the dependency, it only differed numerically. Also, note that there were differences between the online and offline experiments in that the results of the acceptability rating experiment revealed an agreement attraction both in wh-filler and the definite NP. This could have been due to the possibility for readers to look back to the prior context in the acceptability rating experiment, and may also have been due to the lack of power/more noise in the reading experiment.

The third experiment examined how active fillers and reactivated fillers differed in terms of their maintenance, comparing the accessibility of the wh-filler in these two constructions, when the verb is processed. Both the active filler and the reactivated filler showed an illusion of grammaticality effect, with the reactivated filler eliciting attraction later in processing, as revealed by the three-way interaction of *Local noun*, *Grammaticality*, and *Dependency type* at the spillover region 1 and the interaction between *Local noun* and *Grammaticality* at the spillover region 2. Importantly, at the spillover region 1, we found an interaction between *Local noun* and *Dependency Type* in ungrammatical sentences. We contend that this is because in the reactivated filler construction with the coordination structure, the wh-filler is released from memory and the

[11] We thank an anonymous reviewer for pointing this out.

parser reactivates the wh-filler at the point of the coordinating connective, or at the second verb position. The unmaintained information is thus subject to memory decay. In contrast, information associated with the active filler is likely to be maintained, because there is no gap which can complete the WhFGD prior to the gap in the matrix clause, i.e., the active filler is not released from memory and, thus, not subject to memory decay.

. We argue that both maintenance and retrieval play crucial roles in the resolution of whFGD, and adopting either the retrieval or the maintenance view cannot account for the data (see Wagers and Phillips 2014 for related discussion). To understand this point, let us first assume that only retrieval plays a role in dependency resolution (Nicenboim, Vasishth, Gattei, Sigman, & Kliegl, 2015). In this case, for both the active and reactivated filler, the wh-filler is expected to be reactivated at the same point in the sentence. In the case of the active filler (*Which mistake in the program/s that will be disastrous for the company certainly is/are harmful for everyone involved*), the recognition at the matrix verb and the recognition of the gap in the second conjunct at the same time triggers retrieval of the wh-filler. In the case of the reactivated filler (*Which mistake in the program/s will be disastrous for the company and certainly is/are harmful for everyone involved?*), the coordinating connective *and* triggers the reactivation of the wh-filler due to CSC and ATB restrictions or the second verb triggers the reactivation of the wh-filler. Thus, in terms of retrieval, we do not expect any difference between the active filler and the reactivated filler. In the cue-based retrieval model, this reactivation prior to the gap should increase activation specifically for the head of the wh-phrase, not the modifier. This would mean that less attraction should be predicted

for the reactivated filler, relative to active filler.[12] However, the results show that the active filler reveals agreement attraction at an earlier stage. Thus, the differences between the active and the reactivated filler suggest a role for maintenance in parsing, as information about the active filler should be maintained relatively well whereas that of the reactivated filler should not. The earlier agreement attraction for the active filler suggests that details about the content of the NP is not released from the maintenance.

These results suggest that both the maintenance and retrieval are involved in the online whFGD formation. We showed that category information and the internal structure associated with the filler are accessed at the verb position. However, differences in the accessibility of information with respect to different types of fillers and dependency types cannot be explained if we only posit that information is retrieved from the content-addressable memory store based on the cue-based retrieval mechanism. Our results support that some information associated with the filler is easily accessed at the verb, whereas some information associated with the filler is hard to access. Namely, the differences in the retrieval information between different fillers (wh-filler vs. the definite NP) and dependency types (reactivated filler vs. active filler) could be attributed to differences in maintenance. If we assume that maintained information leads to greater accessibility of

[12] Note that, under the cue-based retrieval model, if the gap increases activation specifically for the head of the wh-phrase but not the local noun (Nicenboim et al., 2015), then the weaker agreement attraction is predicted for the reactivated filler relative to the active filler. We would like to note this as a possible alternative hypothesis.

information to the parser, we could account for the differences in the retrieved information between different filler types and dependency types. We showed that both of these two components are used for online WhFGD formation process, where detailed information associated with the filler can be maintained in memory, making it less susceptible to memory decay. On the other hand, if the filler is not maintained in memory, detailed information can still be retrieved at a later stage, though it is subject to decay.

The crucial differences between these two dependency types are that for an active filler, the wh-phrase is linked to the gap in the first conjunct and thus the wh-dependency is completed once the parser reaches the first conjunct. This means that the wh-filler can be released from memory when the wh-gap dependency is formed in the first conjunct. When the coordinating connective *and* is encountered, the wh-phrase needs to be reactivated, but the wh-filler was released from the maintenance before *and* is processed. Unmaintained information is subject to memory decay, thus before *and* is processed, the reactivated filler has been subject to decay. *And* indeed reactivates the filler, but some information associated with the filler could have decayed due to the release from the maintenance. Therefore, detailed information associated with the wh-filler could not be accessed at the verb position. Conversely, the active filler must be linked to the gap in the matrix clause directly in the second conjunct. This may allow for stronger maintenance for detailed information associated with the wh-phrase such as category information and the internal structure of the wh-phrase. Thus, to account for the difference between active and reactivated fillers, we need to consider both maintenance and retrieval mechanisms.

7. Conclusion

This study has attempted to reveal the mechanisms behind online WhFGD formation. Specifically, we aimed to uncover the way in which maintenance and retrieval operate in WhFGD processing, paying special attention to the information that is retrieved when the gap is recognized. The first and second experiments examined the kind of information that is maintained, and examined how maintenance is motivated, by investigating the retrieved information at the gap for reactivated fillers and the definite NPs. The third experiment closely examined the role of retrieval, comparing reactivated and active fillers. We contend that the information that is maintained is reflected in the extent to which information associated with the filler is retrieved at the verb position. We showed that the reader is able to access fine-grained information, including the category information as well as the representation of both the head and the modifier. This suggests that both retrieval and maintenance components play a role in resolving online whFGD (Fiebach et al., 2002; Wagers & Phillips, 2014) and for parsing at large.

References

Acuña-Fariña, J. C., Meseguer, E., & Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua*, 143, 108-128.

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.

Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items. *Journal of Memory and Language, 59*(4), 390-412. doi:10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal. *Journal of Memory and Language, 68*(3), 255-278. doi:10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4, https://cran.r-project.org/web/packages/lme4/index.html. *R package version.*

Bever, T. G., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry*, *19*, 35–43.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in

parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, *34*(6), 774-806. Chen, Gibson, & Wolf, 2005
Box, G. E., & Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological), 211–252.
Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language, 52*(1), 144-169.
Chomsky, N. (1977). On Wh-Movement. In P. W. Culicover, T. Wasow, & A. Akmajian (Eds.), Formal syntax (pp. 71–132). New York: Academic Press.
Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.), Natural language parsing (pp. 94–128). Cambridge: Cambridge University Press.
Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language, 69*(2), 85-103.
Drummond, A. (2011). Ibex Farm, spellout.net/ibexfarm.
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.
Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9*, 427–473.
Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, *4*(2), 93-126.
Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes, 23*(3), 329-374.
Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-Based Interference in Sentence Comprehension: Literature Review and Bayesian Meta-Analysis. *Journal of Memory and Language, 94*(C), 316-339.
Gazdar, G., Klein, E., Pullum, G., & Sag, I. (1985). Generalized phrase structure grammar. Cambridge, MA: Harvard University Press.
Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.
Gibson, E., & Warren, T. (2004). Reading-Time Evidence for Intermediate Linguistic Structure in Long-Distance Dependencies. *Syntax, 7*(1), 55-78.
Grodner, D., & Gibson, E. (2005). Consequences of the Serial Nature of Linguistic Input for Sentenial [Sentential] Complexity. *Cognitive Science: A Multidisciplinary Journal of Artificial Intelligence, Linguistics, Neuroscience, Philosophy, Psychology, 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.
Keine, Stefan (2015): Locality domains in syntax: Evidence from sentence processing. Ms., University of Massachusetts Amherst.
Kim, N., Brehm, L., & Yoshida, M. (In press). The online processing of noun phrase ellipsis and mechanisms of antecedent retrieval. Language, Cognition and Neuroscience. https://doi.org/10.1080/23273798.2018.1513542
King, J. W., & 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.
Kush, D., Lidz, J., & Phillips, C. (2015). Relation-sensitive retrieval: Evidence from bound variable pronouns. *Journal of Memory and Language, 82*, 18-40.

Lee, M.W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, *33*(1), 51–73.
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.
McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*(3), 817–835.
McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), The psychology of learning and motivation (Vol. 46, pp. 155–200). San Diego: Academic Press.
McElree, B., & Bever, T. G. (1989). The psychological reality of linguistically defined gaps. Journal of Psycholinguistic Research, *18*(1), 21-35.
McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Psychology*: General, 118, 346–373.
McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*(1), 67-91.
Nicenboim, B., Vasishth, S., Gattei, C., Sigman, M., & Kliegl, R. (2015). Working memory differences in long-distance dependency resolution. Frontiers in Psychology, *6*(312).
Nicenboim, Bruno, Vasishth, Shravan, Engelmann, Felix, and Suckow, Katja. 2018. Exploratory and Confirmatory Analyses in Sentence Processing: A Case Study of Number Interference in German.
Cognitive Science 42:1075-1100.
Nicol, J. L., Fodor, J. D., & Swinney, D. (1994). Using cross–modal lexical decision tasks to investigate sentence processing. Journal of Experimental Psychology: Learning, Memory & Cognition, 20(5), 1229–1238.
Nicol, J., Forster, K. I., & Veres, C. (1997). Subject-verb agreement processes in comprehension. Journal of Memory and Language, 36(4), 569-587. doi:DOI 10.1006/jmla.1996.2497
Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601-626.
Omaki, A., Lau, E. F., Davidson White, I., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology, 6*:384.
Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language,* 94, 272-290.
Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41*(3), 427-456.
Phillips, C. (2006). The real-time status of island phenomena. *Language*, *82*(4): 795-823.
Phillips, C., Wagers, M. W., & Lau, E. F. (2011). 5: Grammatical Illusions and Selective Fallibility in Real-Time Language Comprehension, *Experiments at the Interfaces* (pp. 147-180): Brill.
Pickering, M., & Barry, G. (1991). Sentence Processing without Empty Categories. *Language and Cognitive Processes, 6*(3), 229-259.
Pickering, M. J., & Traxler, M. J. (2001). Strategies for processing unbounded dependencies: lexical information and verb argument assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1401–1410.
Rohde, D. (2003). Linger: a flexible platform for language processing experiments, version 2.94, Online: http://tedlab.mit.edu/~dr/Linger.
Ross, J. R. (1967). Constraints on variables in syntax. (Doctoral dissertation). Massachusetts Institute of Technology.
Staub, A., & Clifton Jr, C. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2),

425-436.
Stepanov, A., & Stateva, P. (2015). Cross-linguistic evidence for memory storage costs in filler-gap dependencies with wh-adjuncts. *Frontiers in Psychology, 6*, 1301.
Stowe, L. (1986). Parsing wh–constructions: Evidence for on–line gap location. *Language and Cognitive Processes*, *1*(3), 227–246.
Tanenhaus, M. K., Boland, J., Garnsey, S. M., & Carlson, G. N. (1989). Lexical structure in parsing long-distance dependencies. Journal of Psycholinguistic Research, *18*(1), 37-50.
Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language,* 76, 195-215.
Tanner, D., Grey, S., & van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology, 54*(2), 248-259. doi:10.1111/psyp.12788
Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the Processing of Unbounded Dependencies: An Eye-Tracking Study. *Journal of Memory and Language*, *35*(3), 454-475.
Van Dyke, J.A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 407-430.
Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language, 49*(3), 285-316.
Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*, 157–166.
Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2013, 10). Processing Chinese relative clauses: Evidence for the Subject relative advantage. PLoS ONE, *8*(10), 1–14.
Vasishth, S., and Nicenboim, B. (2016) Statistical Methods for Linguistic Research: Foundational Ideas –Part I. *Language and Linguistics Compass*, *10*: 349age andoi: 10.1111/lnc3.12201.
Wagers, M. W., & Phillips, C. (2014). Going the Distance: Memory and Control Processes in Active Dependency Construction. *Quarterly Journal of Experimental Psychology, 67*(7), 1274-1304.
Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*(2), 206-237.
Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), Linguistic theory and psychological reality (pp. 119–161). Cambridge, MA: MIT Press.
Warren, T., & Gibson, E. (2002). The Influence of Referential Processing on Sentence Complexity. *Cognition: International Journal of Cognitive Science, 85*(1), 79-112.
Williams, E. (1978). Across-the-Board Rule Application. *Linguistic Inquiry*, 9, 31-42.