Behavioral and Neural Correlates of Deep and Surface Anaphora

by

Rebecca R. Woodbury

S.B., Engineering Science
Smith College, 2006

Submitted to the Harvard-MIT Division of Health Sciences and Technology,
Speech and Hearing Bioscience and Technology Program,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy of Health Sciences and Technology

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

Signature of Author: _____

Harvard-MIT Division of Health Sciences and Technology
Speech & Hearing Bioscience and Technology Program
December 16, 2010

Certified by: _____

David Caplan, M.D., Ph.D.
Professor of Neurology
Harvard Medical School
Thesis Supervisor

Acccepted by: _____

Ram Sasisekharan, Ph.D.
Director, Harvard-MIT Division of Health Sciences and Technology
Edward Hood Taplin Professor of Health Sciences & Technology and Biological Engineering

Behavioral and Neural Correlates of "Deep" and "Surface" Anaphora

by

Rebecca Woodbury

Submitted to the Harvard-MIT Division of Health Sciences and Technology,
Speech and Hearing Bioscience and Technology Program,
on December 16, 2010,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Health Sciences and Technology

Abstract

Anaphora in language is defined as an expression that refers to another expression. Hankamer & Sag 1976 and Sag & Hankamer 1984 proposed that anaphors can be divided into deep anaphors, which are resolved using a non-linguistic discourse-level interpretation of the antecedent, and surface anaphors, which are resolved by accessing their antecedents at a linguistic level which is highly determined by surface syntactic structure. Previous behavioral studies of the differences between deep and surface anaphors have conflicting and inconsistent results. Additionally, no neuroimaging studies have previously been conducted on deep and surface anaphors or on verb-phrase anaphora in general.

Using two sets of materials which differed in whether they used a surface or a deep anaphor, the behavioral and neural responses as a function of anaphor type were determined. One set of materials was used to examine the effect of placing an intervening sentence between the antecedent and anaphor (distance materials), and one set was used to examine the effect of shifted word order (particle shift materials), both of which were expected to affect surface anaphors more than deep anaphors. Behavioral responses were measured using naturalness ratings and self-paced reading times, and neural responses were measured using blood-oxygen-level dependent (BOLD) signal differences obtained with functional magnetic resonance imaging (fMRI).

Increasing the distance between the antecedent and anaphor affected surface anaphors more than deep anaphors in naturalness ratings, question response times, and BOLD signal, while altering word order had similar or insignificant effects on surface and deep anaphors. These results are consistent with Sag and Hankamer's idea that surface and deep anaphora are distinct categories that are processed differently, but are not consistent with the exact level of access of surface anaphora proposed by Sag and Hankamer. Instead, the results suggest that surface anaphors are more dependent on syntactic information that decays over distance than deep anaphors, but do not differ from deep anaphors in terms of accessing exact surface word order information.

Thesis Supervisor: David Caplan
Title: Professor of Neurology, Harvard Medical School

Acknowledgements

First, I would like to thank my advisor, David Caplan, for his guidance, support, and vast knowledge. I would also like to thank my thesis committee chair, Ted Gibson, for his help in setting up the initial experiments and his support along the way, and my third committee member, Frank Guenther, for his assistance and advice.

I would also like to thank Satrajit Ghosh and Tom Zeffiro for their help with fMRI analysis techniques, Larry White for his assistance with setting up and running fMRI protocols, Rebecca Hufford for teaching me how to run MRI experiments, Jenna Calvino for assisting me with all the fMRI data collection, and Will Evans for helping with behavioral studies at BU.

I am also grateful to my HST and SHBT professors for sharing their incredible wealth of information and inspiration. I would also like to thank Susan Voss, my undergraduate advisor, for inspiring me to pursue scientific research and for introducing me to the SHBT program.

Special thanks goes to my SHBT classmates - Miriam Makhlouf, Adam Furman, and Michael Slama - who were always able to provide support, commiseration, and an excuse to go to Ma Soba for lunch. I would also like to thank my labmates - Rebecca Hufford, Ricky Sachdeva, Joshua Levy, Jason MacMore, and Laura Horton - for providing a supportive, intellectually exciting, and fun lab environment to work in over the past few years.

Finally I would like to thank my family - my mom, Sara, my dad, Sid, and my sister, Maya, who always believed in me, and my fiancé, Jeff, for his encouragement, support, and superhuman patience.

Table of Contents

1. Introduction

1.1. Anaphora in language

Anaphora in language is defined as an expression that refers to another expression. One common example of anaphora is the use of pronouns to refer to an antecedent, such as:

1. Rachel$_i$ takes the bus to school. She$_i$ is in third grade.

where "Rachel" is the antecedent and "she" is the anaphoric pronoun that refers to the antecedent. Anaphora can also involve verb phrases, such as in verb phrase ellipsis:

2. Ben [drinks coffee]$_i$, and Laura does []$_i$ too.

where it is understood that the meaning of the second clause is that Laura does [drink coffee] too.

The processes by which anaphors identify and access their antecedents and the form of the antecedent's information that is accessed by anaphors have been active areas of study since the 1960s. This thesis focuses on two supposedly different types of anaphora, "deep" and "surface" anaphora, and on the form of the antecedent that is accessed by the anaphor in these two cases.

Early theories of anaphoric processing lumped all anaphoric processes together in discussing how anaphors are resolved. One major group of early theories, referred to as the transformational or classical position, posited that anaphors are transformationally derived

through deletion of a syntactically complete noun phrase under identity with the antecedent noun phrase (e.g. Ross 1967, Ross 1969, Postal 1970, Postal 1972). A second major group of theories, called the interpretive position, instead claimed that all anaphoric relationships are created interpretively, with no syntactic deletion occurring and with no underlying syntactic form present (e.g. Jackendoff 1972, Wasow 1972, Shopen 1972, Fiengo 1974). Importantly, neither group of theories allowed for any potential difference in processing between subcategories of anaphora.

1.2. Sag and Hankamer's theory of deep and surface anaphora

Hankamer and Sag (1976) and Sag and Hankamer (1984) proposed that anaphors could be divided into two groups, "deep" and "surface", according to whether their antecedents had to be represented linguistically, and also how they were processed. They claimed that certain anaphors required a linguistic antecedent to be acceptable, whereas other anaphors did not. For example:

3. a. *Scenario: a man sees a woman about to jump off a bridge.*

Man: "Don't do it!" ("do it" - deep anaphor)

*Man: "Don't do so!" ("do so" - surface anaphor)

b. Man: "A woman was about to jump off a bridge...

...and I told her not to do it."

...and I told her not to do so."

In this example, they argued that "do so" is acceptable only after the antecedent has been stated linguistically as in 3b, whereas "do it" is acceptable with both a contextual (3a) and a linguistic (3b) antecedent. Hankamer and Sag classified anaphors that require linguistic antecedents (such as "do so") as surface anaphors, and anaphors than do not require linguistic antecedents (such as "do it") as deep anaphors.

In addition, H&S/S&H claimed that surface anaphors are sensitive to syntactic parallelism between the antecedent and the anaphor, whereas deep anaphors do not require this parallelism, as in the example here from S&H 1984:

4. The children asked to be squirted with the hose, so

   a. they were []. (VPE, surface)

   b. *we did []. (VPE, surface)

   c. we did it. [sentential it, deep].

In this example, the antecedent phrase is passive, and Sag and Hankamer claimed that the surface anaphor is acceptable only when it is also passive, as in 4a, and not when it is active, as in 4b. They claimed that the deep anaphor is acceptable when it is active even when the antecedent is passive.

Additionally, S&H 1984 claimed that surface anaphors were affected by intervening discourse between the antecedent and the anaphor, whereas deep anaphors were not. An example of this is as follows[1]:

[1] This particular example was given by Belanger 2004.

5. a. John raked the leaves in the back yard.

i. Later, Bill did too. (surface)

ii. Later, Bill did it too. (deep)

b. John raked the leaves in the back yard.

This was much more fun than studying for exams.

i. ?Later, Bill did too. (surface)

ii. Later, Bill did it too. (deep)

In 5a, they suggested that anaphors were acceptable directly after the antecedent sentence, but introducing additional discourse information between the antecedent and the anaphor as in 5b rendered the surface anaphor less acceptable.

Sag and Hankamer (1984) proposed that these three pieces of evidence (which were all subjective claims unsupported by quantitative data) suggested that the deep and surface anaphora are processed differently. They claimed specifically that deep anaphors access their antecedents using a non-linguistic discourse-level interpretation of the antecedent, while surface anaphors access their antecedents at a linguistic level which is highly determined by surface syntactic structure. Specifically, they claimed that surface anaphors are processed by assigning an interpretation of the antecedent's full VP structure to the site of the null anaphoric VP, where the structure that is assigned corresponds to surface syntactic units.

They noted that the differences in acceptability between anaphor types following nonparallel antecedents, situation-related antecedents, and additional distance between the

anaphor and the antecedent can all be explained by this processing difference. In the case of a situation-based antecedent, they argued that a surface anaphor is unacceptable because it accesses the antecedent at a linguistic level, which is unavailable in the case of a situation-based antecedent.

Similarly, they claimed that syntactic parallelism affects surface anaphors more than deep anaphors because surface anaphors are processed using a mechanism that copies the surface structure of the antecedent at the anaphoric site, and copying a non-parallel antecedent into the anaphoric site would result in a grammatically unacceptable anaphoric sentence. For example, in the sentence "She told me the oats had to be taken down to the bin, so I did", Sag and Hankamer predict that the anaphor will be unacceptable because copying the surface structure of the antecedent VP at the anaphoric site results in an interpretation along the lines of "so I had to be taken down to the bin" or "so I taken the oats down to the bin", neither of which are acceptable interpretations of this sentence. Sag and Hankamer claim that deep anaphora is processed using references to elements of the discourse model (where the discourse model includes situational information) that do not include surface form, even when the antecedent is presented linguistically, and that as long as the action of taking the oats down to the bin in the previous example is present in the discourse model, it will be available as an acceptable antecedent for a deep anaphor. Sag and Hankamer do not explain exactly how they believe elements in the discourse model are accessed or selected during the resolution of the deep anaphor.

Finally, the supposed phenomenon by which surface anaphors are less acceptable than deep anaphors when intervening material is placed between the anaphor and the antecedent, as in Example 5, can be explained using the results of previous studies showing that surface syntactic

structure and exact word information decays rapidly while discourse-level information remains more salient over time (e.g. Sachs 1967, Sachs 1974). Sag and Hankamer present this argument anecdotally without citing these relevant studies (some of which had not been published at the time of their work), by simply stating that "propositional representations... ....lead a temporary life in a limited short-term memory register" while discourse models remain salient.

Several studies have provided evidence that the representation of surface syntax decays over intervening material while semantic or discourse information does not decay. Sachs 1974 presented initial sentences to subjects followed by a sentence of variable length, after which a target sentence was presented that was either identical to the first, changed syntactically or lexically, or changed semantically. Subjects were asked to judge whether the target sentences were the same as the initial sentences. Syntactic and lexical changes were detected significantly less often than semantic changes after even the shortest intervening sentence, indicating that syntactic and lexical information decays more quickly than semantic information with the introduction of additional linguistic material. This result was also obtained by Poppenk et al 2008, who conducted a similar experiment changing either surface syntactic structure, semantic relationships, or introducing entirely novel episodes, and found that, over time, changes in surface sentence structure were detected significantly less often than changes in semantic relationships or entirely new episodes, indicating that surface syntactic information is less salient over time than semantic information. Sag and Hankamer claim that surface anaphors access surface structure, which is thought to decay with intervening material, and deep anaphors access discourse information, which is thought to remain salient with intervening material, and thus they explain their intuitional judgment that anaphors that access distant surface information are

less acceptable than anaphors that access closer surface information or distant or close discourse-level information.

Sag and Hankamer's theory that surface and deep anaphors are processed differently, with surface anaphors depending on surface syntactic representations and deep anaphors depending on discourse information, was based on the subjective acceptability judgments outlined above and was not supported by quantitative data. Subsequently, several research groups have attempted to gather quantitative data on the acceptability claims that Sag and Hankamer made. These experiments are summarized later in the Previous Distance Experiments and Previous Parallelism Experiments sections.

1.3. Verb-Phrase Ellipsis (VPE) Studies

Subsequent work on theoretical differences between surface and deep anaphors focuses on quantitatively testing the claims of Sag and Hankamer, which is discussed in further sections. Other research has focused on anaphoric processes without distinguishing between surface and deep anaphora. This literature includes work on noun phrase anaphora (e.g. Almor 1999, Ariel 1990, Dell et al 1983, Garnham et al 1997, Garnham 2001, Garrod and Sanford 1994, Gernsbacher 1989, Gordon et al 1993, Gordon and Chan 1995, Gordon and Scearce 1995, Grosz et al 1983, Hammer et al 2008, Kaiser et al 2009, McKoon and Ratcliff 1980, Myers and O'Brien 1998, Nieuwland and Van Berkum 2008, O'Brien et al 1997, Sanford and Garrod 1989) and verb phrase anaphora, the type of anaphora studied in this thesis. Almost all work on verb phrase anaphora has specifically examined the processing of verb-phrase ellipsis (VPE - the type of surface anaphor used in this thesis). These studies will be addressed briefly here, as this thesis

focuses on identifying differences between behavioral and neural measures of surface and deep anaphora as proposed by Sag and Hankamer, rather than isolating processing mechanisms associated with one or the other. Deep anaphora in general as well as specific examples of deep anaphors have received almost no further attention in terms of their processing.

Several pieces of evidence exist indicating that at least some representation of syntactic structure exists in VPE. Clifton and Frazier (2010) present three examples from previous literature that succinctly argue for the existence of syntactic structure at the ellipsis site. The first example they cite is that Fiengo and May (1994) argue that for the sentence

6. I know which book Max read, and which Oscar didn't.

to be processed correctly, the ellipsis must contain a trace. The second example cited by Clifton and Frazier is that Grinder and Postal (1971) argue that in the example

7. My aunt doesn't have a spouse but your aunt does and he is lying on the floor.

the ellipsis site must contain a representation of the antecedent "have a spouse" for the pronoun "he" to be processed correctly. Finally, Lasnik (2001) indicates that the ellipsis site must include some information about scope, as in the example

8. Each of the linguists criticized some of the other linguists but I'm not sure how many of the other linguists [each of them criticized].

where the scope of "many" is dependent on syntactic structure existing at ellipsis.

Frazier and Clifton (2005) presented empirical data suggesting that syntactic structure is present at the ellipsis site. In Experiment 2 of their paper, they showed that ellipses with two-clause relative clauses, in which the antecedent syntactic structure is inadequate unless the determiner phrase (DP) is raised using quantifier raising (QR), such as

9. Dulles investigated some man Bill said you did.

were rated as being less natural than ellipses with one-clause relative clauses, such as

10. Dulles investigated some man you did, Bill said.

Importantly, similar sentences without ellipses with one- or two-clause relative clauses, such as in the following example,

11. One-clause RC: Dulles investigated some man you investigated, Bill said.

Two-clause RC: Dulles investigated some man Bill said you investigated.

were rated as being equally natural. Frazier and Clifton claim that these results indicate that syntactic structure must be available at the ellipsis site, since the antecedent is inadequate unless QR is performed on the DP (a syntactic operation) to create an appropriate antecedent. In

Experiment 3 of the same study, Frazier and Clifton showed in an acceptability judgment task that sentences with an elided interrogative clause with island violations such as

12. Sally was impressed after some lecture, but I don't know what.

were significantly less acceptable than sentences with elided interrogative clauses without island violations, such as

13. Sally was impressed with some lecture, but I don't know what.

and argued that the penalty for ellipses with island violations indicated the presence of syntactic structure at the ellipsis.

Despite the evidence that syntactic structure exists in VPE, several groups have pointed out examples of acceptable VPE in which the antecedent is not syntactically parallel, such as in the example

14. This information could have been released by Gorbachov, but he chose not to (Daniel Shorr, NPR, 10/17/92, cited in Hardt 1993).

Arregui et al 2006 proposed a syntactic account of VPE resolution which aimed to explain cases of acceptable mismatching antecedents syntactically. They used acceptability judgment paradigms with antecedents in which the verb phrase was variably available (ranging from

completely available to completely unavailable). The acceptability of these materials varied with the availability of the verb phrase, such that materials with completely available verb phrases were judged to be the most acceptable, materials in which the verb phrase was available but not located in the verb phrase position were judged to be slightly less acceptable, materials in which the verb phrase was available but required revision were judged to be even less acceptable, and materials in which the verb phrase was unavailable were judged as largely unacceptable. From this data, they proposed a "VP recycling hypothesis", in which a syntactically matching antecedent is used to process VPE when available, but when a syntactically matching antecedent is unavailable, the listener/reader repairs the antecedent at LF to create an acceptable antecedent. If the repair of the flawed antecedent can be done easily using available grammatical operations, the ellipsis will be relatively acceptable. The acceptability of the antecedent is determined by the number of operations needed to repair the antecedent, such that antecedents that require fewer operations for repair will be more acceptable than antecedents that require more operations for repair.

Other groups claim that the cases of acceptable syntactically mismatching antecedents indicate that discourse factors are at least moderately important in VPE. Several groups argue that VPE relies only on a semantically salient antecedent in order to be grammatical (e.g. Dalrymple et al 1991, Hardt 1992, Hardt 1993, Kehler 1993, Hardt and Romero 2004). However, these accounts overgenerate in terms of acceptability, in that they suggest that examples such as "The pig was caught and John did" should be perfectly acceptable.

Kehler (2000) proposed a different model which took into account both syntactic and discourse factors. He proposed that discourse coherence relationships between the antecedent

and the anaphor affect whether syntactically matching antecedents are necessary. Specifically, he proposed that only resemblance relations, such as

15. #This problem was looked into by John, and Bob did too.

require syntactically matching antecedents, while cause-effect relations, such as

16. The problem was to have been looked into, but obviously nobody did.

allow syntactically unmatching antecedents and account for cases of acceptable syntactic mismatches.

However, Kehler's argument was based on a small number of examples that were subjectively determined to be acceptable or unacceptable. Frazier and Clifton (2006) tested these claims by obtaining naturalness rating judgments with syntactically matching and mismatching antecedents using both resemblance and cause-effect relations. They found that syntactically matching antecedents were equally preferred for both discourse coherence relations, which directly counters Kehler's theory that syntactically mismatching antecedents are acceptable for cause-effect discourse coherence relations but not for resemblance relations. They suggested that Kehler's use of "too" in his resemblance relation conditions may have been responsible for forcing a greater parallelism requirement in the resemblance relation conditions, rather than there being any difference in requirements between the actual coherence relations.

Kertz (2008) presented a different argument for the varying acceptability of VPE with mismatching antecedents, claiming that the difference between acceptable and unacceptable mismatching ellipses arises at least partially from the information structure of the ellipses. Specifically, she claimed that focusing the subject argument of the antecedent target clause results in an unacceptable mismatch, such as in example (14) or in the example

17. The problem was looked into by the committee, just like the chair did,

while focusing the auxiliary verb of the antecedent target clause results in an acceptable mismatch, such as in the example

18. The problem was to have been looked into, but obviously nobody did.

Using acceptability judgments, she showed that although mismatching antecedents were less acceptable than matching antecedents in both focused auxiliary and focused argument cases, the difference in acceptability between match and mismatch was greater in the argument-focused condition, as predicted.

Clifton and Frazier (2010) presented another argument for discourse structure having the ability to influence syntactic constraints on ellipsis. They argued that mismatching antecedents are more acceptable when the antecedent doesn't imply that something actually occurred, which they referred to as a "nonactuality implicature". Using an acceptability judgment paradigm, they found that materials with nonactuality implicatures such as

19. If John had gone to the store he would have bought Twinkies. George would have too.

were more acceptable than materials that implied that an action had occurred, such as

20. If John went to the store he bought Twinkies. George did too.

and

21. If John went to the store he bought Twinkies. George would have too.

They use this result to argue that discourse constraints can override or influence syntactic constraints, so that syntactic constraints can be reduced under certain discourse structures.

There is therefore evidence arguing for the existence and importance of syntactic structure at the point of ellipsis in VPE (consistent with Sag and Hankamer's original theory), as well as compelling results showing that discourse information can influence syntactic constraints.

1.4. Overview and significance of this dissertation

Anaphora is frequently used in language, and its accurate processing is crucial for understanding discourse. The mechanisms by which antecedents are accessed and the types of representations that are accessed by anaphors are not well understood. Understanding how

anaphoric relationships are resolved can lead to a greater understanding of which aspects of language are maintained and re-accessed in discourse.

Sag and Hankamer developed an elegant theory of anaphoric processing, dividing anaphora into "deep" and "surface" anaphora. Behavioral studies attempting to objectively study Sag and Hankamer's claims regarding deep and surface anaphora (to be reviewed in Chapters 2 and 3) have yielded conflicting and inconsistent results, with potential confounds in the materials and with important statistical comparisons frequently omitted.

The aim of this dissertation is to quantitatively test Sag and Hankamer's claims that distance affects surface anaphora but not deep anaphora (distance experiments), and that syntactic parallelism (specifically surface word order) is necessary for surface anaphora but not for deep anaphora (particle shift experiments). In testing these claims, we hope to provide data relevant to the question of whether deep and surface anaphora access different representations of their antecedents, as claimed by Sag and Hankamer. These claims are tested using naturalness ratings, reading times, and fMRI experiments.

2. The Effect of Distance on Deep and Surface Anaphora

2.1. Previous studies: distance

Several research groups have attempted to quantitatively test Sag and Hankamer's claim that surface anaphors are affected more than deep anaphors by distance between the anaphor and the antecedent, with varying and inconclusive results. The first two experiments in Murphy (1985) varied the length of the antecedent and the distance between the antecedent and the anaphor. The length of the antecedent was varied by including or excluding the bracketed segments:

22. Jimmy swept the [tile] floor [behind the chairs free of hair and cigarettes.]

    a. Later his uncle did too. (surface)
    b. Later his uncle did it too. (deep)

The reading times for both surface and deep anaphors increased in the lengthened antecedent condition, but no significant interaction was found between anaphor type and antecedent length. Murphy concluded that both surface and deep anaphors access linguistic levels and are therefore affected equally by the increased length. However, it is also possible that surface and deep anaphors access different levels of representation in which antecedent length affects processing in different ways, resulting in slowed reading times for both anaphor types. For example, the increased length of the antecedent results in a more complex linguistic representation of the antecedent, but it could also result in a more complex discourse model for the antecedent. This could increase processing times for surface anaphors due to the more complex linguistic nature

of the antecedent while also increasing processing times for deep anaphors due to the more complex discourse representation of the antecedent.

In a second experiment, Murphy (1985) added an intervening sentence between the antecedent and anaphor sentences, while varying the length of the antecedent as in the first experiment. The reading times for the anaphors were compared to the reading times from the first experiment to determine how the addition of an intervening sentence affected surface and deep anaphors with both short and long antecedents. With the addition of the intervening sentence, as compared to the first experiment, reading times increased for surface and deep anaphors by 874 and 483 ms respectively for short antecedents and by 520 and 399 ms respectively for long antecedents, indicating that the addition of an intervening sentence appears to affect surface anaphors more than deep anaphors, but no statistics on the interaction between anaphor type and distance were provided. Additionally, no examples or materials were provided for this experiment, making it impossible to know whether these results were affected by the type of intervening sentence which may have introduced confounding factors such as topic shifts.

Tanenhaus et al (1985) also studied the effect of distance on surface and deep anaphors by placing an intervening sentence between the antecedent and the anaphor, similar to Murphy (1985) Experiment 2. The two experiments differed in that Murphy (1985) used verb-phrase ellipsis as a surface anaphor, while Tanenhaus et al (1985) used sluicing. An example of Tanenhaus et al (1985)'s materials is shown in 23, where the sentence in brackets was included to increase the distance between the anaphor and antecedent:

23. Somebody has to paint the garage.

[The paint is peeling and the wood is beginning to rot.]

a. Let's take a vote and see who has to do it. (deep)

b. Let's take a vote and see who. (surface)

Tanenhaus et al determined that in the increased distance condition, the reading times increased by 207 ms for the surface anaphor and decreased by 84 ms for the deep anaphor. However, no statistics, including the interaction between anaphor type and distance condition, were included. Additionally, as pointed out by Murphy (1990), these results were potentially confounded by the fact that the intervening sentence is a syntactically possible antecedent for the surface anaphor, but not the deep anaphor, making the surface anaphor potentially interpretable as "Let's take a vote and see who [is peeling and beginning to rot]". While this is semantically improbable, the syntactic possibility of this alternate reference may account for some or all of the increased reading time for the surface anaphor in the distant condition.

Murphy (1990) also conducted two experiments that investigated the effect of distance on anaphor type while also varying the ambiguity of the intervening sentences and the type of task used. He conducted a comprehension task experiment and a plausibility judgment experiment using the same materials. In these materials, an antecedent was presented first. In the "close" condition, an anaphoric sentence containing either a VPE (surface) or a "do it" (deep) anaphor was presented directly after the antecedent sentence. In the "distant" condition, an intervening sentence was presented between the antecedent and the anaphor. This sentence could either be

ambiguous in that it served as a syntactically (although semantically implausible) possible alternative antecedent for the VPE anaphor (but not the deep anaphor), or unambiguous in that it could not be a potential alternative antecedent for either anaphor. An example of one complete item is shown below:

24. Ellen Marcovitz was flying from Seattle to Washington, D.C.

She was a sales representative for Acme Aviation, Inc.,

and she was trying to get a big government contract for her company.

Because she was nervous, she started a conversation with the man next to her.

The man asked her a question about Acme Aviation. [antecedent]

-This was more relaxing than her previous worrying. [unambiguous filler]

-She felt more relaxed almost immediately. [ambiguous filler]

Later, Ellen wondered why he did/did it. [anaphor]

She worried that she might have told something important to this stranger.

In the comprehension task experiment, the surface and deep anaphors were affected similarly by the presence of an intervening sentence for both the unambiguous and the ambiguous distant conditions (for the unambiguous distance condition, reading time increased by 307 msec for surface anaphors and by 292 msec for deep anaphors; for the ambiguous distance condition, reading time increased by 209 msec for surface anaphors and by 232 msec for deep anaphors; the interaction between anaphor type and distance was insignificant, with Fs <1.0).

Murphy then conducted a plausibility judgment experiment using the same materials. Both deep and surface anaphors were equally affected by distance in terms of judgment accuracy. However, the reaction time data did show differences between surface and deep anaphors, such that the response time for the surface anaphors increased over conditions so that close < distant-unambiguous < distant-ambiguous, while the response time for deep anaphors decreased over conditions such that close > distant-unambiguous > distant-ambiguous. While the increase in reading time for the surface anaphor conditions was consistent with the processing model expectations, the decreasing response time for deep anaphors was not predicted and was not explained by Murphy 1990. The overall interaction of antecedent condition and anaphor type reached statistical significance in the reaction time data ($F(2,58)=3.18$, $p<0.05$, $F(2,40)=3.19$, $p<0.052$), but no analysis was performed to determine whether distance/anaphor type interactions within either unambiguous or ambiguous conditions were individually significant. The presence and direction of the distance/anaphor type interaction was therefore partially consistent with the processing model for the plausibility judgment task, but not for the comprehension task. Murphy suggested that the difference across tasks occurs because subjects are consciously processing the formal differences between the deep and surface anaphors in the plausibility judgment task, whereas in the comprehension task they are only relating each sentence to the previous sentence without focusing consciously on each sentence.

Overall, the effects of increased distance between the antecedent and the anaphor on processing of deep and surface anaphora are unclear. Most studies that have investigated these differences have either found no difference between the two anaphora types (Murphy 1985 Experiment 1, Murphy 1990 Experiment 1) or have not reported important statistics necessary

for drawing conclusions (Murphy 1985 Experiment 2, Tanenhaus et al 1985). Additionally, the materials used in several of these studies contained potential confounds (such as intervening sentences that might induce topic shifts or are semantically/syntactically possible antecedents for the anaphor) or were not provided and thus have unknown potential confounds.

2.2. Distance materials

The distance materials used in the rating study, reading time study, and fMRI study varied anaphor type (surface or deep) and presence/absence of intervening sentence (short or long). A full list of the distance target items can be found in Appendix A. Each condition consisted of an introductory antecedent sentence, which described a subject performing an action. In the long conditions, an intervening sentence was then presented which was related to the introductory sentence. Then an anaphoric sentence was presented, where the surface anaphor was in the form of verb-phrase ellipsis (VPE), and the deep anaphor was in the form of a "do it" anaphor. The anaphoric sentence was always in the form of "Then/Later/The same day/etc, [second subject] did/did it too." The items were constructed unambiguously, in that the anaphor always referred to the first sentence, and the intervening sentence could not serve semantically as a possible antecedent for the anaphor. The stimulus items were sometimes followed by comprehension questions, where comprehension questions were balanced as to which part of the item they referred to. The frequency of comprehension questions depended on the experiment (see individual experimental methods for details). Half of the comprehension questions were true, and half were false. An example of one item is shown below in Table 1.

Anaphor type	Intervening sentence	Antecedent sentence	Intervening sentence	Anaphoric sentence	Comprehension question
surface	no	John raked the leaves in the back yard.		Later, Bill did too.	Did John rake the leaves?
deep	no	John raked the leaves in the back yard.		Later, Bill did it too.	Did John rake the leaves?
surface	yes	John raked the leaves in the back yard.	The oak and maple leaves covered the ground.	Later, Bill did too.	Did John rake the leaves?
deep	yes	John raked the leaves in the back yard.	The oak and maple leaves covered the ground.	Later, Bill did it too.	Did John rake the leaves?

Table 1: Example of one item for the distance experiment.

The items also varied the discourse relation between the first sentence and the intervening sentence, as discourse structure has been thought to be important in anaphoric interpretation (e.g. Kehler 2000). The discourse relationships are divided between elaborative (e.g. "John raked the leaves in the back yard./ The oak and maple leaves covered the ground."), cause-effect/effect-cause ("Lauren got a flu shot last Monday./ She didn't want to get sick and miss school."), and temporal ("Rebecca bought a new bike./ A week later, it was stolen."). These three discourse relations were chosen out of many discourse relations that exist, as they resulted in natural-sounding sets of sentences and were compatible with a third anaphoric sentence.

The subjects in these sentences were always proper names, and the items varied as to whether the subjects were male and female, female and male, or both male or both female.

The target items were interspersed with 120 filler items, designed to reduce strategies and to make the items less predictable in format. A full list of the distance fillers can be found in Appendix B. The fillers followed the format of the items, in that they consisted of a similar antecedent sentence and an anaphoric sentence, with another sentence in between in half of the fillers. The fillers differed from the target items with respect to the anaphoric sentence. The

filler anaphors contained various anaphoric references to the antecedent sentence that did not use VPE or "do it" anaphora. Several examples are as follows:

25. Samantha took the dog for a walk./ Then Bob took it for a walk.
26. Ruby bought a new house last year./ This year, Eleanor bought one too.
27. Ellen threw out her old shoes./ They were smelly and falling apart./ Emma threw hers out too.
28. Flora downloaded a new software program./ It was very helpful for work./ Today, Gerald did the same thing.

The filler sentences also included similar comprehension questions and were balanced for gender of the subjects in the sentences.

2.3. Experiment 1a: Distance rating study

2.3.1. Introduction

Although several groups have attempted to test Sag and Hankamer's predictions that only surface anaphors are affected by distance between the antecedent and the anaphor, none of these studies have looked for quantitative differences in the acceptability of surface and deep anaphors with increased distance. Sag and Hankamer claim that surface and deep anaphors with no intervening sentence between the antecedent and anaphor and deep anaphors with intervening sentences are both perfectly acceptable, whereas surface anaphors with intervening sentences are not acceptable. This statement is based purely on an intuitive judgment made by Sag and Hankamer, not on any quantitative evidence that this is the case. Gibson and Federenko (2010)

point out that intuitive judgments such as this are subject to cognitive biases by the researcher or may result from lexical properties rather than syntactic properties, and it is therefore important to obtain quantitative data on acceptability judgments using multiple items and multiple naive subjects.

The goal of the distance naturalness rating study was therefore to obtain objective, quantitative data on the differences in acceptability of surface and deep anaphors with and without intervening sentences between the anaphor and the antecedent. Sag and Hankamer claim that all conditions are acceptable except the long surface anaphor condition, because the surface anaphor relies on surface syntactic information that decays over the intervening sentence, and that deep anaphors rely on discourse information that does not decay over the intervening sentence, so deep anaphors should be equally acceptable in long and short conditions. If this is the case and no other factors are involved, we would expect surface short, deep short, and deep long conditions to have equal naturalness ratings, while surface long conditions would have lower naturalness ratings than the other three conditions.

However, Sag and Hankamer do not consider any other possible contributing factors such as baseline differences between the acceptability of surface and deep anaphors. For example, it may be possible that surface anaphors are used much more frequently in language than deep anaphors and are therefore more familiar, easier to process, and more natural-sounding overall than deep anaphors. If this was the case, we would expect to see a main effect of naturalness rating, such that surface anaphors were rated more natural overall than deep anaphors. If Sag and Hankamer's processing model was still confirmed in this case, we would expect deep short and deep long conditions to be equally natural, with surface short conditions being more natural

than deep short conditions. Surface long conditions would be less acceptable than surface short conditions, and depending on the baseline difference in acceptability between surface and deep anaphors and the size of the decrease between surface short and surface long conditions, would either be more natural, equally natural, or less natural than deep long conditions.

It is also possible that Sag and Hankamer's theory is unfounded and the quantitative rating data could not support their original claims. In this case, we could expect several patterns of results. One such pattern would be if there was no significant difference in naturalness between surface short and surface long conditions, or if surface long conditions were judged to be more natural than surface short conditions. These patterns could indicate that surface anaphors do not depend on linguistic information that decays over intervening material. Another pattern that would show Sag and Hankamer's model to be incorrect is if there was a significant difference between deep short and deep long conditions. This could indicate that deep anaphors also depend on linguistic representations that decay over intervening material, implying that Sag and Hankamer's model would be incorrect because deep anaphors do not refer to discourse representations. However, this pattern could also arise if deep anaphors depend on discourse information as proposed by Sag and Hankamer, but introducing an intervening sentence makes the discourse information more complex and therefore equally difficult to access. This would indicate that Sag and Hankamer's model is not incorrect, but that they did not consider the effect of an intervening sentence on the complexity of discourse information.

2.3.2. Methods

Participants 320 volunteers were recruited from Amazon.com's Mechanical Turk (e.g. Gibson and Fedorenko in press; Munro et al 2010) and paid for their involvement in the study. These volunteers were native speakers of English from the USA who were naive to the purposes of the study. Informed consent was obtained prior to participation.

Materials 120 items as described in the Distance Materials section were used in the naturalness judgment experiment. The 120 target items were divided into four different groups to make the task shorter. Each group of 30 items were further split into four lists including only one condition from each item, and two versions of this list were created, resulting in a total of eight lists for each group and 32 lists total. Each list was rated by 10 subjects, so that each distinct set of sentences was seen by 20 subjects, with a total of 320 subjects for the entire experiment. 120 fillers of the format described in the Distance Materials section were interspersed among the items. Every target item and filler was followed by a comprehension question. A list of the items and fillers used in the naturalness judgment study can be found in Appendices A and B, respectively.

Procedures In the naturalness judgment study, subjects read the distance materials and rated the anaphoric sentence as to its naturalness in the context of the preceding sentences. This experiment appeared as a survey on Mechanical Turk, a service operated by Amazon.com. An example of one item as seen by the subjects is shown below in Figure 1.

Context sentence(s): Frank went ice-skating on the pond yesterday. It was a beautiful winter day.

Target sentence: Later Sarah did too.

Rating of target sentence in this context:

○ Extremely unnatural ○ Somewhat unnatural ○ Possible ○ Somewhat natural ○ Extremely natural

Question: Did Frank ice-skate on the pond yesterday?

○ Yes ○ No

Figure 1: Example of one item as seen by subjects on Mechanical Turk in the distance naturalness judgment study.

Subjects were instructed to read the "context" for each set of sentences, which consisted of the anaphoric sentence and the intervening sentence in the "long" conditions and only the anaphoric sentence in the "short" conditions. When two sentences were presented as a context, they were both shown on the same line. The subject was then asked to read the "target sentence", which is the anaphoric sentence. After reading the target sentence, they were asked to respond to how natural the target sentence sounds given the context, with possible responses of "extremely unnatural", "somewhat unnatural", "possible", "somewhat natural", and "extremely natural". These ratings were assigned a numerical value from 1 (extremely unnatural) to 5 (extremely natural) in the data analysis phase. After rating the naturalness of the target sentence, the subject then answered a comprehension question. Subjects were eliminated from the analysis if their accuracy on the comprehension questions was less than 75%, as this indicated that they were not performing the tasks as requested.

2.3.3. Results

239 subjects were retained in the naturalness judgment study. 81 subjects were removed for having comprehension question accuracy of less than 75%, attempting to complete more than one list, or not being native English speakers from the USA.

On average, 96% of the comprehension questions for the target items were answered correctly. Comprehension question accuracy collapsed over coherence condition is shown below in Figure 2.

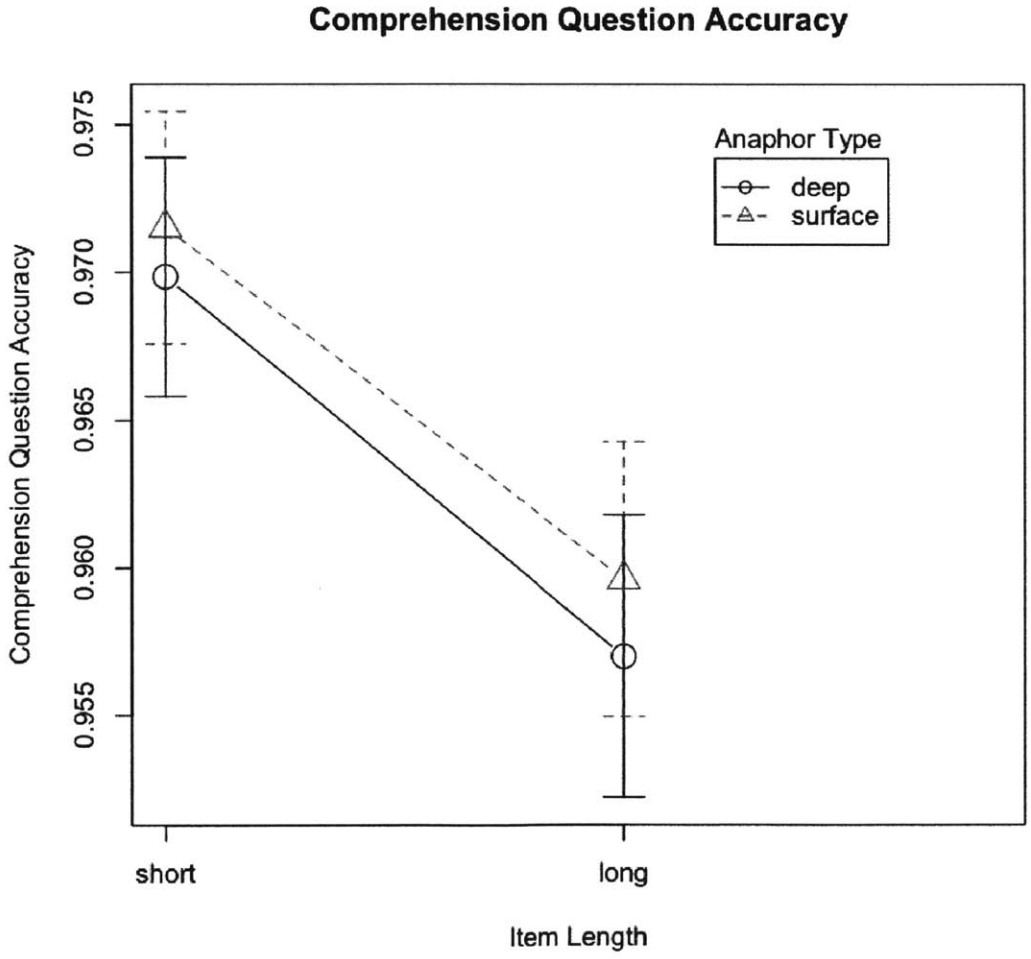

Figure 2: Comprehension question accuracy collapsed across coherence condition for the distance naturalness judgment study.

A 2 x 2 x 3 (deep/surface anaphor x short/long length x elaborative/cause-effect/temporal coherence relation) analysis was performed on the comprehension question accuracy data using a mixed linear model (lmer function in R) with subjects and items as random effects and experimental conditions of interest as fixed effects (e.g. Van Dongen et al 2004, Jaeger 2008). There was a significant main effect of length ($F(1, 7143) = 10.0$; $p < 0.01$), with more errors occurring in response to long items than short items (mean accuracy for long items = 96.3%, mean accuracy for short items = 97.1%)[2]. Comprehension question accuracy, distinguishing between coherence conditions, is shown below in Figure 3.

[2] A mixed effects logistic regression on the comprehension question accuracy data also showed a significant main effect of length ($z=2.237$, $p<0.05$) with no significant main effect of anaphor type or significant interaction.

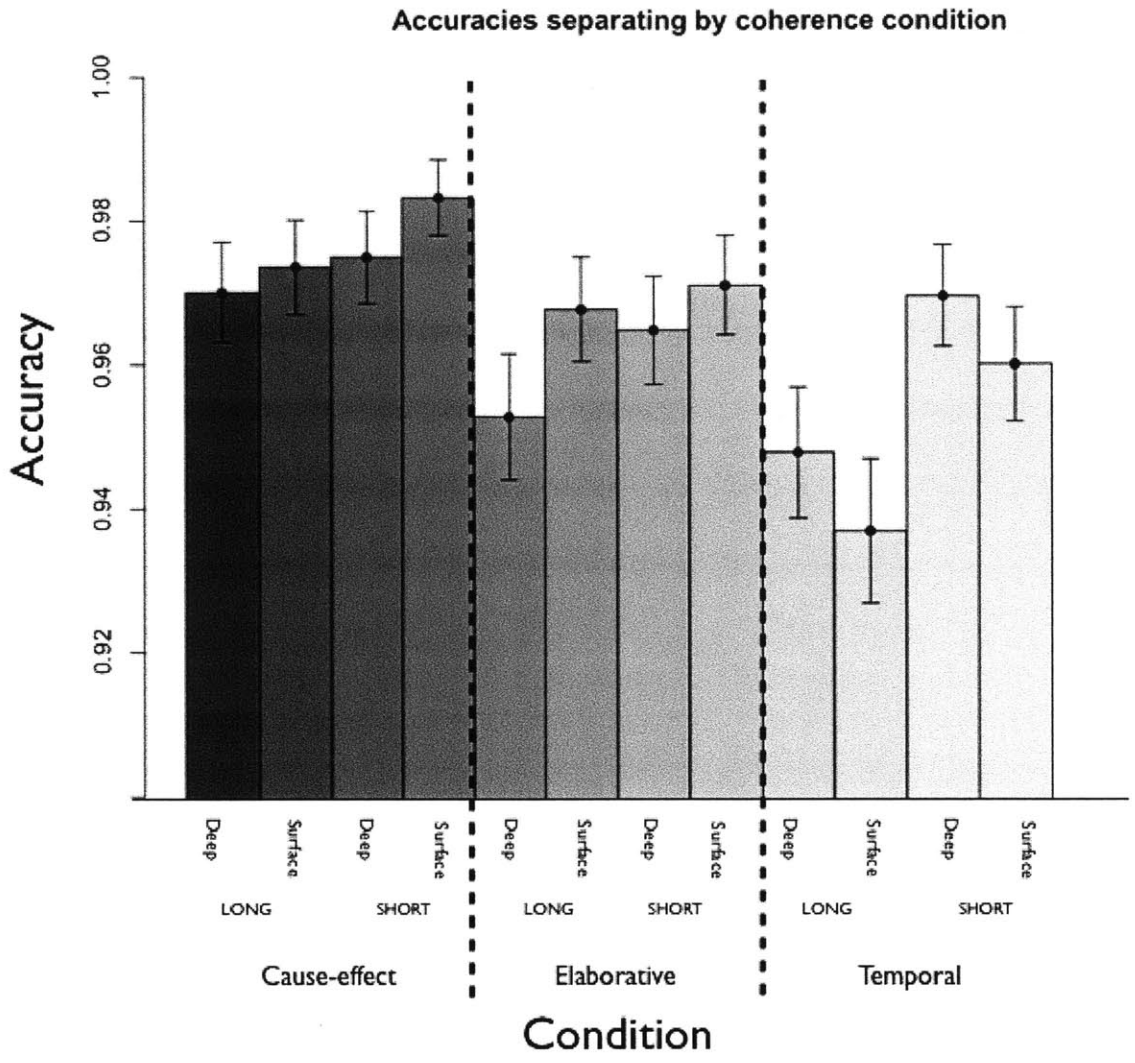

Figure 3: Comprehension question accuracy, distinguishing by coherence condition, for the distance naturalness judgment study.

The main effects of anaphor type and coherence condition and all interactions between conditions were not significant (all $p > 0.05$).

For the naturalness rating data, only items with correct answers to the comprehension questions were included in the analysis. A 2 x 2 x 3 (deep/surface anaphor x short/long length x elaborative/cause-effect/temporal coherence relation) analysis was performed on the naturalness rating data using a mixed linear model (lmer function in R) with subjects and items as random

effects and experimental conditions of interest as fixed effects (e.g. Van Dongen et al 2004, Jaeger 2008). Naturalness ratings collapsing across coherence condition are shown below in Figure 4.

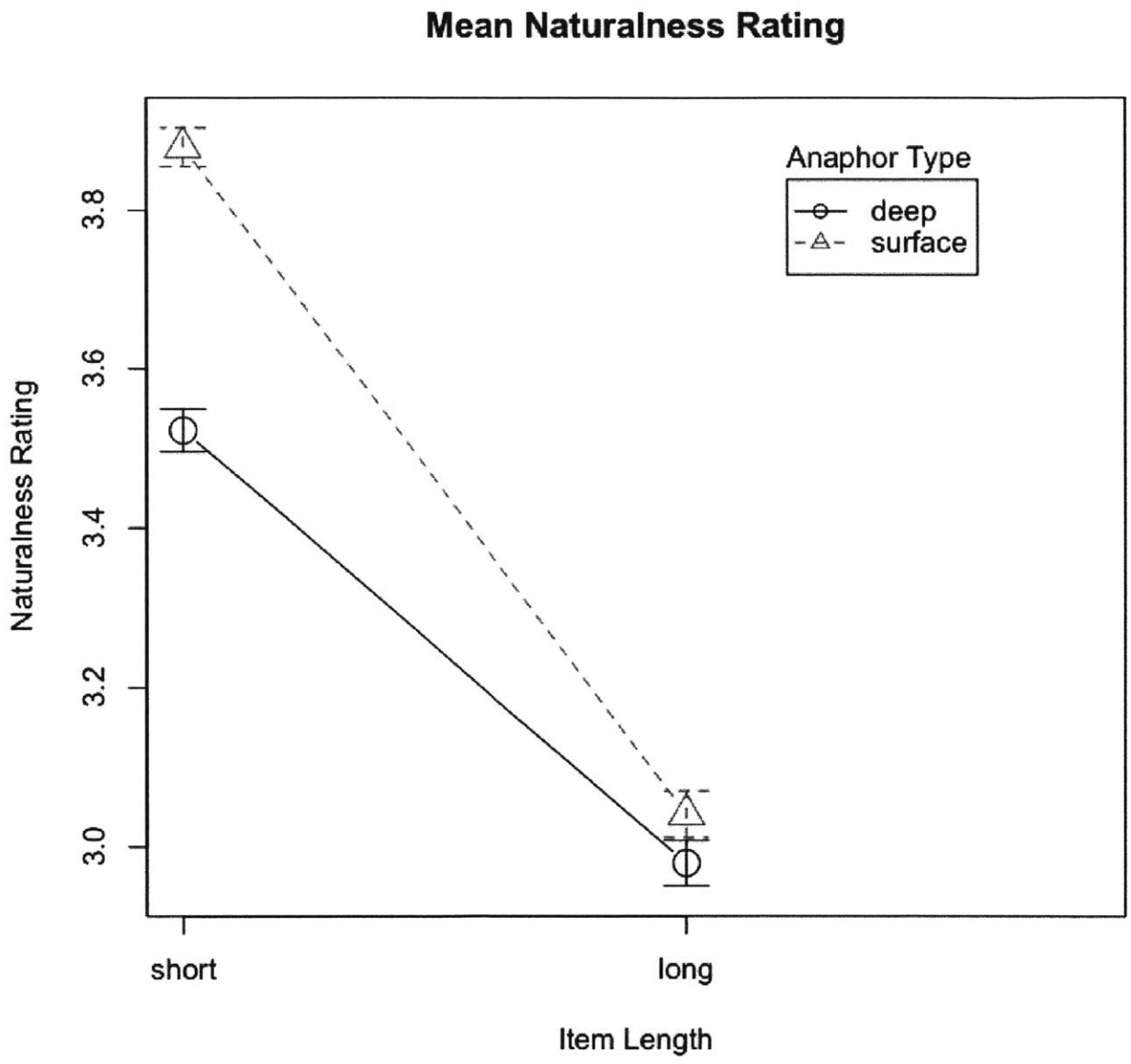

Figure 4: Naturalness ratings collapsed across coherence condition for the distance naturalness judgment study.

There was a significant main effect of anaphor type ($F(1,7104) = 90.2$; $p < 0.0001$), such that items with surface anaphors were rated higher (more natural) than items with deep anaphors

(mean rating for surface anaphors = 3.46 out of 5, mean rating for deep anaphors = 3.25 out of 5). There was a significant main effect of length ($F(1, 7104) = 1005.1$; $p < 0.001$), such that short items were rated higher than long items (mean rating for short items = 3.70, mean rating for long items = 3.01). Ratings divided by coherence condition are shown below in Figure 5.

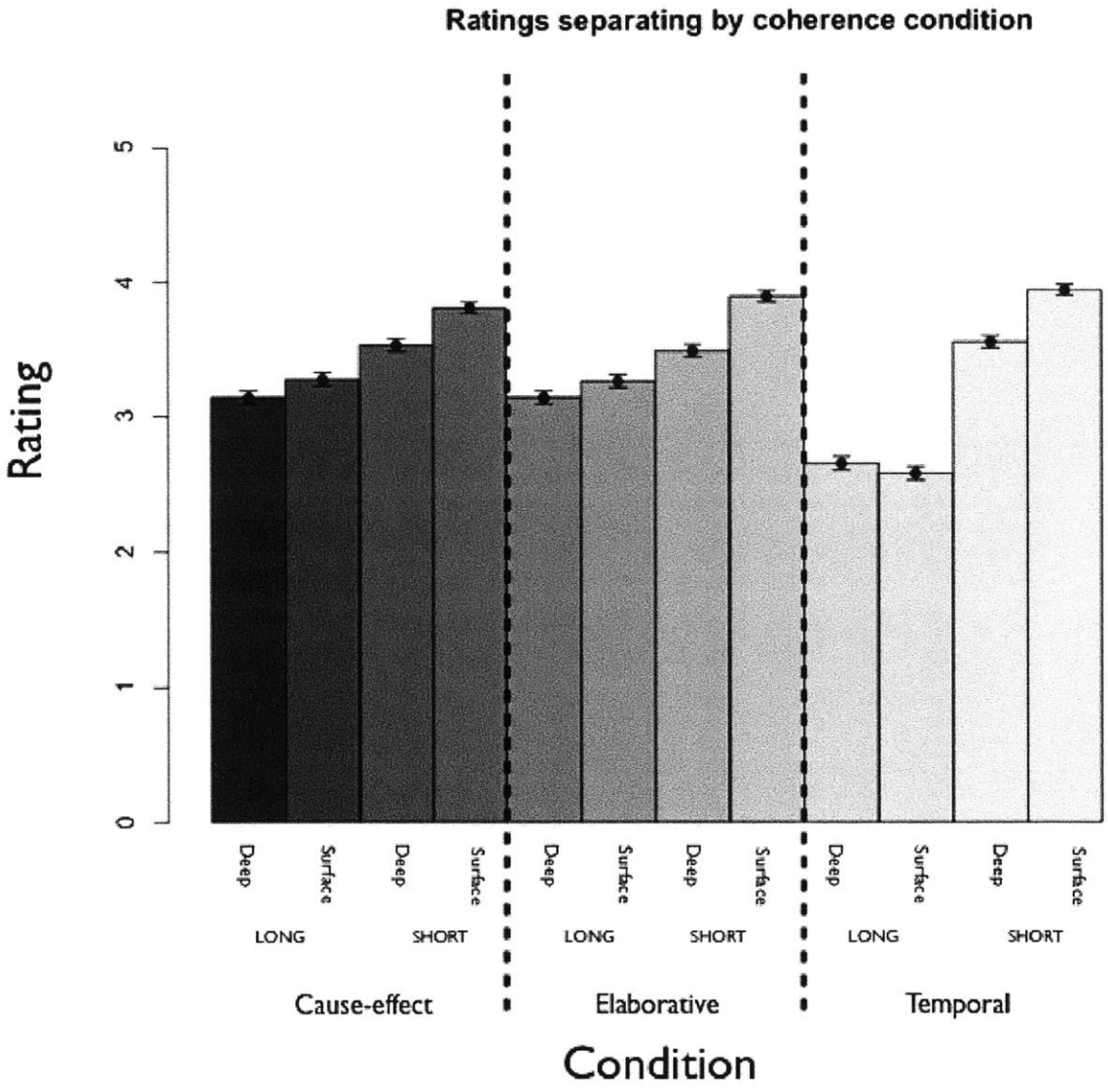

Figure 5: Naturalness ratings, distinguishing by coherence condition, for the distance naturalness judgment study.

There was also a significant main effect of coherence relation ($F(2, 7104) = 10.6$; $p < 0.001$), such that items with temporal coherence relations were rated lower than those with cause-effect and elaborative relations (mean rating for temporal items = 3.18, mean rating for cause-effect items = 3.44, mean rating for elaborative items = 3.45).

Of interest was a significant interaction between the anaphor type and length ($F(1, 7104) = 47.7$; $p < 0.001$). Post-hoc t-tests indicated that the differences between short deep and short surface, short deep and long deep, and short surface and long surface were all significant ($p < 0.001$), while the difference between long deep and long surface conditions was not significant ($p > 0.05$). A post-hoc t-test showed that the difference between long surface and short surface was significantly greater than the difference between long deep and short deep ($p<0.001$). This pattern was consistent within all three coherence relations.

2.3.4. Discussion

The processing model predicts that the presence of an intervening sentence between the anaphor and the antecedent should make the surface anaphor less natural than when there is no intervening sentence, as the surface information will be degraded by the increased distance between the antecedent and the anaphor. The results obtained here are consistent with that prediction, as the acceptability of surface anaphors decreases significantly with the introduction of an intervening sentence.

However, the results here indicate that, although smaller than the surface difference, the difference between deep short and deep long items is also significant, which was not predicted by Sag and Hankamer's processing model. This could be because the addition of an intervening sentence makes the conceptual representation of the discourse more complex and thus makes the

processing of the deep anaphor more difficult when the conceptual representation must be accessed after an intervening sentence. This would indicate that the processing model is correct with respect to the type of representation contacted by each type of anaphor, but that Sag and Hankamer did not take into account that the deep anaphor resolution processes could be affected by an intervening sentence. It could also be possible that both surface and deep anaphors use linguistic level information that is affected by the introduction of an intervening sentence, which would indicate that the processing model is inaccurate and both deep and surface anaphors use linguistic level information. There is no way to distinguish between these two possibilities using results from the naturalness rating data, as the data do not elucidate underlying mechanisms of processing. However, the fact that the surface anaphors exhibited a significantly larger decrease in acceptability with the introduction of an intervening sentence indicates that the processes contributing to deep and surface anaphora resolution are different in some way, which is consistent with the first explanation but not necessarily confirmative of Sag and Hankamer's original theory.

Another aspect of the results that was not predicted by Sag and Hankamer is that surface anaphors were rated as significantly more natural than deep anaphors in the short condition. The significant interaction between anaphor type and distance resulted in the surface long and deep long conditions having statistically equivalent naturalness ratings, even though the drop in acceptability with the introduction of the intervening sentence was greater for the surface anaphors than the deep anaphors. This effect of anaphor type in the short condition could arise because the surface anaphor used here is used more frequently than the deep anaphor and is therefore more familiar and easy to process, or because the mechanism by which the surface

anaphor is processed is cognitively easier overall than the mechanism by which the deep anaphor is processed. Given the unpredicted effect of the surface anaphors being rated as more natural than the deep anaphors in the short condition, the greater drop in naturalness from long to short for surface anaphors is consistent with Sag and Hankamer's predictions.

Looking within coherence relations, this pattern was consistent within all three coherence relations, indicating that surface and deep anaphors are not affected by length disparately due to coherence condition. However, all three coherence relations were not equally natural; the temporal coherence conditions were rated significantly less natural than the cause-effect and elaborative conditions, due to a much lower rating for the long conditions (both deep and surface). This could be because the temporal conditions have different, slightly less natural temporal orders than the cause-effect and elaborative conditions. In all long items, the anaphor always referred to the first sentence, not the second sentence. In cause-effect and elaborative conditions, the temporal order of the sentences is generally Sentence 2 --> Sentence 1 --> Sentence 3 or Sentence 1 and 2 --> Sentence 3, while in the temporal conditions it is generally Sentence 1 --> Sentence 2 --> Sentence 3 such as in the examples below.

29.	Cause-effect:	2	Donald installed new windows in his apartment.
		1	The cold wind was causing a draft.
		3	Later that month, Harry did/did it too.
	Elaborative:	1&2	Molly wrote a letter to the senator.
		1&2	It was twelve pages long.
		3	The next day, Nicole did/did it too.
	Temporal:	1	Virginia cut all her hair off last year.
		2	Then it grew long again.
		3	Last week, Kathleen did/did it too.

It is possible that it is more natural for the anaphor to refer to the sentence that directly precedes the antecedent temporally, such as in the cause-effect and elaborative conditions, than when it refers to a sentence that is temporally more distance, such as in the temporal condition.

2.4. Experiment 1b: Distance reading time study

2.4.1. Introduction

Although Sag and Hankamer only make predictions about acceptability of surface and deep anaphors with and without intervening sentences, predictions can also be made about the effects of intervening sentences on the reading times of surface and deep anaphors. If the differences in acceptability that Sag and Hankamer claim are present result from difficulties in processing, these difficulties should be evident in reading time data as well. Following Sag and Hankamer's claim that surface anaphors with intervening sentences should be unacceptable and surface anaphors without intervening sentences as well as deep anaphors with or without intervening sentences should be acceptable, we would expect that there should be a reading time penalty for surface long conditions as compared to surface short conditions, whereas deep long and deep short conditions should have approximately equal reading times.

However, the results from the naturalness judgment experiment did not reflect the predictions of Sag and Hankamer. The naturalness rating results showed an effect of anaphor type, such that surface anaphors are read more quickly than deep anaphors in the short condition. If the effect of anaphor type in the naturalness judgment experiment arose from a difficulty in processing or in familiarity due to differences in frequency, we would expect this to be reflected in reading time results. We also saw a significant difference in naturalness between deep long and deep short conditions, which was not predicted by Sag and Hankamer's model. This

difference could also be reflected in reading time results, with deep long conditions having longer reading times than deep short conditions. As described in the naturalness rating section, this could indicate that intervening sentences make the deep anaphors more difficult to process due to an increase in complexity of the discourse model, or that deep and surface anaphors are processed using similar information that is degraded by intervening material.

2.4.2. Methods

Participants 24 college-aged subjects were recruited from the Boston University community and paid for their involvement in the study. The subjects were native speakers of English from the USA who were naive to the purposes of the experiment. Informed consent was obtained prior to participation.

Materials 120 items in the format described in the Distance Materials section were used in the reading time study. Subjects saw only one condition from each item, to eliminate familiarity effects. Subjects therefore saw 120 sets of sentences, or 30 examples per each of the 4 conditions. These 120 sets of sentences were drawn from 4 distinct lists created from the 120 items. Half of the items were followed by comprehension questions. The items were interspersed with 120 filler items of the format described in the Distance Materials section. Full lists of the items and fillers used in the reading time experiment are shown in Appendices A and B respectively.

Procedures In the reading time experiments, the materials were presented on a screen and reading times were measured using the Linger program (Rohde 2003). Subjects were instructed

to press the space bar when they were done reading the words on the screen in order to progress to the next sentence or group of words.

The antecedent and intervening sentences appeared in whole sentence form. The anaphoric sentence was then presented phrase-by-phrase, with dashed lines indicating the words that are not currently being presented. Each anaphoric sentence was divided into two phrases, with the second phrase including the anaphor, such as [The next day] [Bob did too]. After the sentences were presented, a comprehension question was presented in whole-sentence form after 50% of the items. The subject typed "F" for 'yes' or "J" for 'no' depending on what the answer to the comprehension question was. Subjects were reminded which keys to press at the bottom of the screen with the instructions "'F' for yes. 'J' for no." each time a comprehension question was presented. Items with comprehension questions were randomly mixed in with items without comprehension questions to reduce expectations and strategies.

2.4.3. Results

Data for 24 subjects were included in the analysis. Each subject had an average accuracy of greater than 67%. The average accuracy for the comprehension questions was 88.5%.

Only items where the true/false repeated statement was answered correctly were included in the reading time analysis. Data points where the reading time was more than 3 standard deviations from the mean were removed from the analysis, which affected 1.2% of the data. Data points where the response time to the true/false repeated statement was more than 3 standard deviations from the mean were also removed from the analysis, which affected 2% of the data.

The items were divided into three or four regions for the reading time analysis, depending on whether the target item included an intervening sentence between the antecedent and anaphor or not. These regions were numbered as follows:

30. [John raked the leaves in the back yard.]
1

[The oak and maple leaves covered the ground.]
2 (only present in long conditions)

[Later,] [Bill did it too.]
3 4

The critical region was region 4, which included the anaphoric phrase. As in the particle shift experiment, all statistical tests were carried out with mixed linear models (lmer function in R), using subjects and items as random effects and experimental conditions of interest as fixed effects (e.g. Van Dongen et al 2004, Jaeger 2008).

Raw reading time data for each region are shown below in Table 2. Residual reading times were analyzed to account for differences in word length and for differences in participants' mean reading rates. To determine residual reading times from raw reading times, a linear regression equation was computed for each subject which predicted reading times as a function of word length using all items and fillers (e.g. Ferreira and Clifton 1986; Trueswell, Tanenhaus, and Garnsey 1994). The residual reading time for each region was then calculated by subtracting the predicted reading time as calculated by the subject's regression equation from the subject's measured raw reading time. Conceptually, a negative value of the residual reading time indicates that the reading time was faster than predicted, and a positive value of the residual reading time indicates that the reading time was slower than predicted. A less negative value of the residual reading time indicates a slower reading time than a more negative value of the residual reading

time. Statistical results were consistent between raw reading times and residual reading times, but only residual reading times will be discussed. Data for the residual reading times in each region is shown below in Figure 6.

	Region			
Condition	**1**	**2**	**3**	**4**
Deep Short	1647.6 (34.2)		639.1 (12.4)	783.2 (16.8)
Surface Short	1667.1 (34.4)		616.9 (11.0)	724.1 (15.5)
Deep Long	1750.2 (34.8)	1345.6 (24.8)	604.5 (10.9)	761.2 (16.9)
Surface Long	1704.6 (33.6)	1325.8 (24.3)	585.5 (8.6)	716.8 (15.5)

Table 2: Raw reading times per region in the distance reading time experiment. Times shown are in milliseconds. Numbers in parentheses indicate standard error in milliseconds. The regions are explained above.

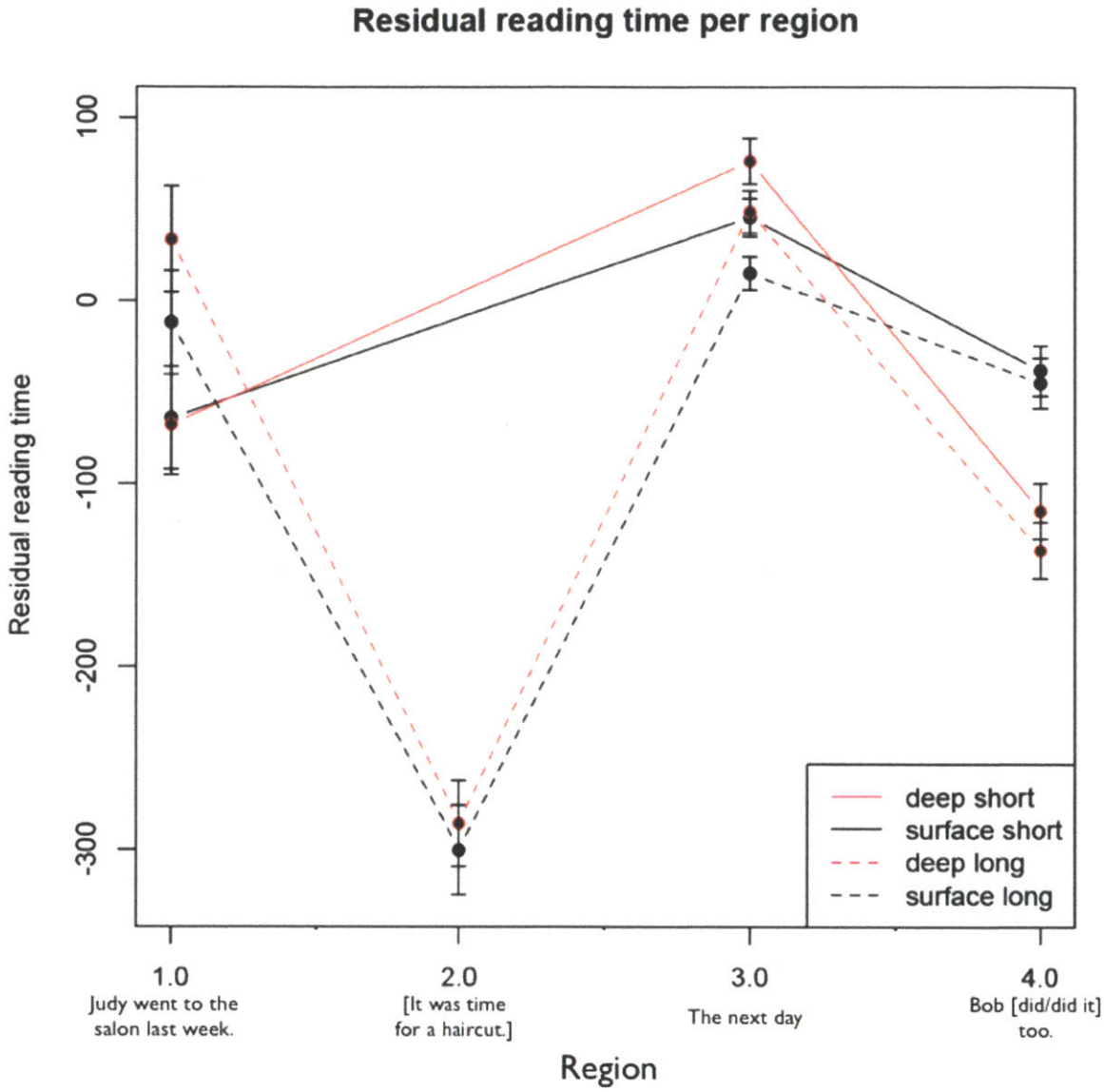

Figure 6: Residual reading time per region for the distance reading time experiment. Regions are defined above in Example 29.

Within each region, 2x2 (distance x anaphor type) analyses were conducted using mixed linear models as described above. When interactions were significant, planned comparisons (t-tests) between long and short conditions were conducted within anaphor types on residual

reading times for each region individually (surface long versus surface short and deep long versus deep short).

In the first region, which was the antecedent sentence, there was a significant main effect of distance ($F(1,2765)=8.30$, $p<0.01$), such that short items were read more quickly than long items (mean short = -65.96 msec, mean long = 10.84 msec). The main effect of anaphor type and the interaction between length and anaphor type were insignificant (both $p > 0.1$).

In the second region, which was the intervening sentence (only present in the long conditions), a pair-wise comparison (t-test) of residual reading time by anaphor type (deep versus surface) revealed no significant difference between reading times for deep versus surface anaphor conditions ($F(1, 1428) = 0.36$; $p = 0.55$). This is predicted as the items do not differ across anaphor type in this region or prior to this region.

In the third region, which was the introductory word/phrase leading into the third (anaphoric) sentence, there was a significant main effect of distance ($F(1,2875)=8.67$, $p>0.01$), such that short conditions were read more slowly than long conditions (mean short = 60.92 msec, mean long = 31.83 msec). There was also a significant main effect of anaphor type ($F(1,2875=10.37$, $p < 0.01$), such that items with deep anaphors were read more slowly than items with surface anaphors (mean deep = 62.42 msec, mean surface = 30.36 msec). The interaction between anaphor type and length was insignificant ($p > 0.5$).

In the fourth region, which was the critical region that included the anaphor, there was a significant main effect of anaphor type ($F(1,2875)=35.11$, $p<0.001$), such that the residual reading times for surface anaphors were greater than the residual reading times for deep anaphors

(mean surface residual = -41.46 msec, mean deep residual = -125.67 msec)[3]. The main effect of distance and the interaction between distance and anaphor type were insignificant.

The comprehension question response times are shown below in Figure 7. For the comprehension question response time, there was a significant main effect of length ($F(1,1458)=17.34$, $p < 0.001$), such that the response times for long items were greater than the response times for short items (mean long = 1862.7 msec, mean short = 1731.5 msec). There was also a significant main effect of anaphor condition ($F(1,1458)=4.64$, $p < 0.05$), such that response times for items with surface anaphors were greater than the response times for items with deep anaphors (mean surface RT = 1825.9 msec, mean deep RT = 1768.7 msec). The interaction between length and anaphor type was also significant ($F(1,1458)=10.48$, $p < 0.01$). Planned comparisons within anaphor type were conducted between long and short conditions (surface long vs. surface short and deep long vs. deep short). Surface anaphors differed significantly between long and short conditions ($F(1,715)=26.85$, $p<0.001$) such that response times were faster for surface short conditions than for surface long conditions (mean surface short = 1716.8 msec, mean surface long = 1937.8 msec). The difference between long and short conditions was not significant for deep anaphors ($F(1,743)=1.13$, $p=0.29$, mean deep short = 1746.2 msec, mean deep long = 1792.9 msec).

A post-hoc analysis was performed to determine differences in comprehension question response times between questions that referred to the antecedent sentence, intervening sentence, and anaphoric sentence. Comprehension questions were the same for each version of a given item, allowing comparisons between comprehension question responses. For comprehension

[3] The effect of anaphor type was still significant in Region 4 when the reading times from Region 3 were factored out ($F(1,2874)=35.84$, $p<0.001$), indicating that this main effect in Region 4 is not a spillover effect from Region 3.

questions that referred to the antecedent sentence, there was no difference in response time between conditions (all $p > 0.5$). For comprehension questions that referred to the intervening sentence, there was an effect of anaphor type, such that surface conditions were answered more slowly than deep conditions ($F(1,248)=5.99$, $p<0.05$). For comprehension questions that referred to the anaphoric sentence, there was a main effect of length ($F(1,604)=18.6$, $p<0.001$), such that long conditions were answered more slowly than short conditions, a main effect of anaphor type ($F(1,604)=4.36$, $p<0.05$), such that surface conditions were answered more slowly than deep conditions, and a significant interaction between length and anaphor type ($F(1,604)=13.17$, $p<0.001$), such that surface conditions differed significantly between long and short while deep conditions did not differ significantly between long and short.

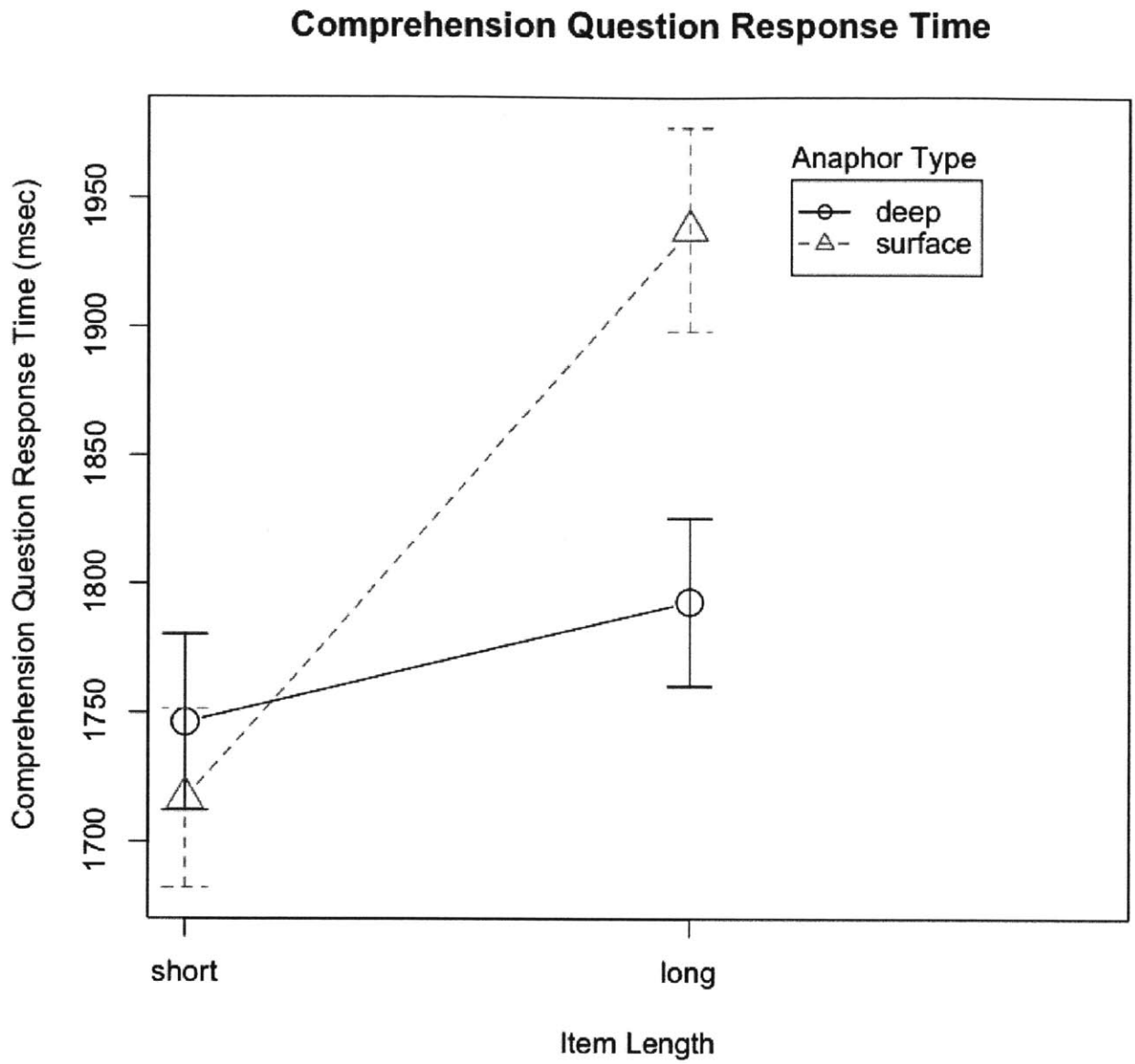

Figure 7: Comprehension question response times for the distance reading time experiment.

The comprehension question accuracies are shown below in Figure 8. For the comprehension question accuracy, there was a significant effect of length ($F(1,1458)=7.05$, $p < 0.01$), such that subjects were more accurate on short items than on long items (mean short accuracy = 90.34%, mean long accuracy = 86.24%). The main effect of anaphor type and the interaction between length and anaphor were both insignificant (both $p > 0.05$).

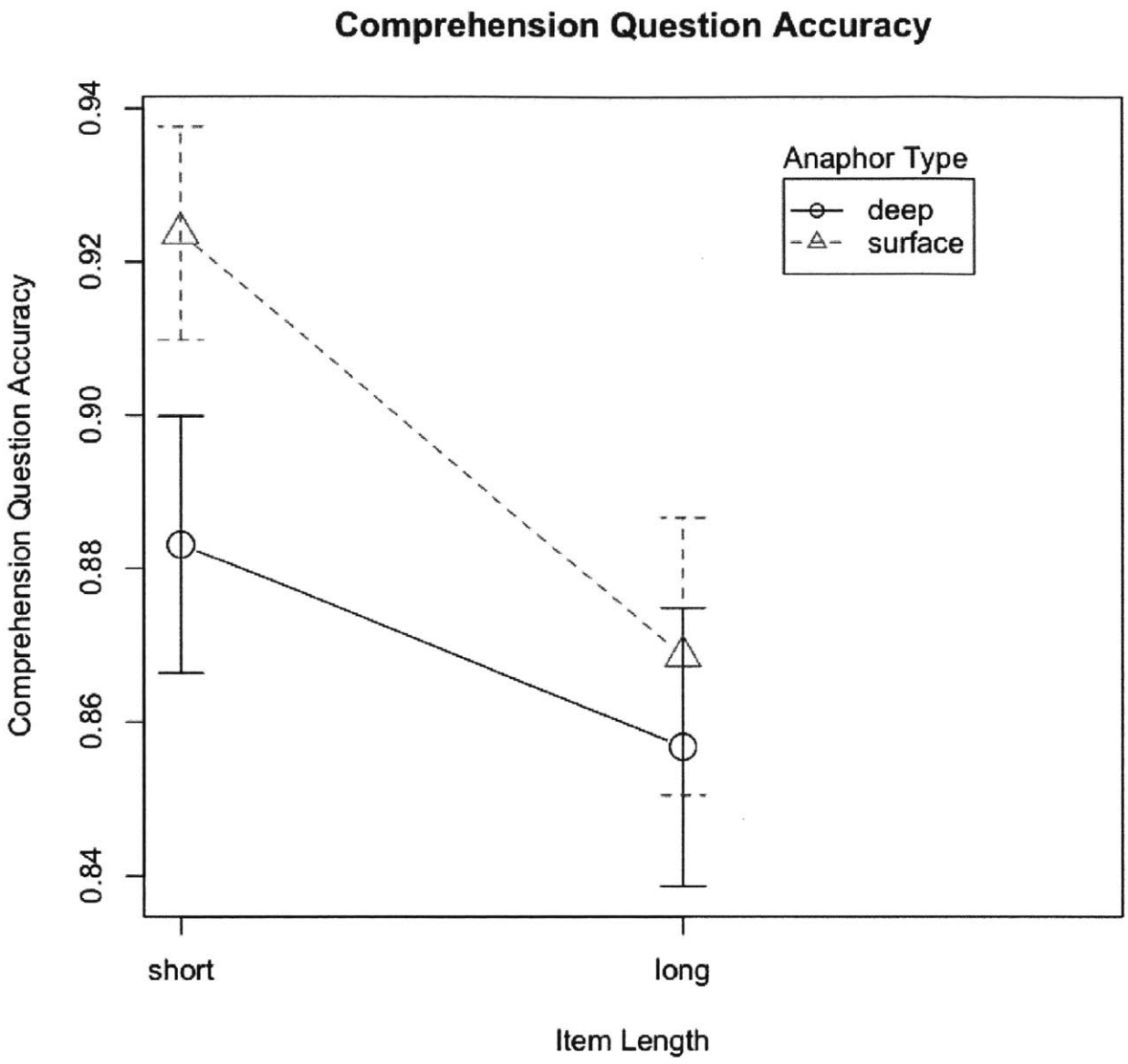

Figure 8: Comprehension question accuracy for the distance reading time experiment.

2.4.4. Discussion

A main effect of item length was seen on reading times in the antecedent sentence, such that reading times for long conditions were slower than reading times for short conditions. Although this region itself does not differ between long and short items, this difference may reflect an anticipatory effect, as the number of displayed blanked-out lines differs between the

short and long conditions. For example, in the short conditions subjects would see a display like this:

31. John raked the leaves in the back yard.

---.

Whereas for the long conditions subjects would see a display like this:

32. John raked the leaves in the back yard.

---.

---.

The subjects may predict or learn that it will be more difficult to remember the first sentence in the long conditions than in the short conditions, so they may spend more time reading it in the long conditions.

Similarly, there was a main effect of anaphor type in Region 3, which was the introductory region of the anaphoric sentence, such that deep conditions were read more slowly than surface conditions. As the materials did not differ at this point between deep and surface conditions, this effect may be an anticipatory effect reflecting the visible difference in length of the deep and surface conditions, since the deep conditions were 3 characters longer than the surface conditions.

In the critical anaphoric region, there was an effect of anaphor type, such that surface anaphors were read more slowly than deep anaphors. This finding is consistent with the idea that surface and deep anaphors are processed differently. This result could reflect a process in which surface anaphors access an additional level of representation (such as a syntactic representation) before relating the meaning of the antecedent to the meaning of the anaphoric sentence.

Although the predicted effect of distance on surface anaphors was not evident in the reading times for the anaphoric sentence itself, this effect was apparent after the fact in the response time to comprehension questions. Comprehension question response time for surface anaphors increased for long conditions as compared to short conditions, while there were no differences between short and long conditions for deep anaphors. The selective difficulty in surface long conditions is consistent with a processing theory where items with surface anaphors are more difficult to process when intervening sentences degrade the linguistic level information in the antecedent, while items with deep anaphors are not affected because the anaphor does not depend as heavily on linguistic level information.

The question remains as to why distance effects for surface anaphors in the reading time experiment were seen only in comprehension question response time and not in anaphoric region reading times. The absence of an effect on reading times could be due to the somewhat short length of the intervening sentences. Sachs (1974) used intervening sentences of varying syllabic lengths to observe how long it takes surface level word information to decay, and found that exact word information was very salient with 0 syllables (corresponding to 1 second), decayed substantially with 20 syllables (4 seconds), and was completely forgotten after 40 syllables (8.5 seconds). The intervening sentences used in this study were around 10 syllables long, and subjects read these sentences in an average of 1.3 seconds. These sentences may not have been long enough for the surface information from the antecedent sentences to decay substantially by the time the anaphoric sentence is reached, resulting in no reading time penalties for the surface anaphors over the deep anaphors in this region. To investigate whether this is contributing to the difficulty in surface long conditions not showing up in anaphoric reading times, a followup

experiment would have to be conducted using longer intervening sentences or intervening sentences of varying lengths.

It is also possible that length effects are not seen in the surface anaphoric region because the surface anaphor does not reactivate the antecedent's surface information. If the surface anaphor doesn't reactivate the surface information of the antecedent, this could be because the surface anaphor doesn't use or access any surface information, so it is not reactivated. Another way that the anaphor could access its antecedent without reactivating the surface form of the antecedent is through the use of a pointer mechanism rather than a copying mechanism to access surface information of the antecedent in memory. This mechanism would result in surface information being accessed and used without being copied into the anaphoric site, and therefore the surface information would not be reactivated at the anaphoric site. This idea is discussed in further detail in the Overall Discussion of Results section.

Length effects in surface anaphors did show up in the question response time data, such that surface long conditions were answered slower than surface short conditions, with no significant differences between deep short and deep long conditions. A post-hoc analysis showed that this difference arose when the comprehension question referred to the anaphoric sentence, but not when the comprehension question referred to the antecedent sentence. Responding to the comprehension question requires retrieving information from the sentence to which the comprehension question refers. In surface conditions, length did not affect the response time to questions referring to the antecedent but did affect the response time to questions referring to the anaphor, indicating that length adversely affected the representation of information in the

anaphor but not in the antecedent. These results are consistent with a processing difficulty due to length that is specific to the sentence containing the surface anaphor.

Overall, the results from the distance reading time experiment do not support Sag and Hankamer's processing model, in that there was no difference in reading time for surface anaphors between short and long conditions. The only difference in reading times at this region were differences between the overall reading times for surface and deep anaphors, which may reflect a difference in overall processing between surface and deep anaphors but does not elucidate any mechanism by which this may be the case. However, the comprehension question response time data showed that surface long conditions were answered significantly slower than surface short conditions when the comprehension question referred to the anaphoric sentence, with no significant differences between the deep short and long conditions, indicating that increased distance negatively affects surface anaphors specifically.

2.5. Experiment 1c: Distance fMRI study

2.5.1. Introduction

Although several imaging experiments have been conducted on noun-phrase anaphora (e.g. Santi and Grodzinsky 2007a, Santi and Grodzinsky 2007b, Almor et al 2007, Nieuwland et al 2007), no experiments as of yet have examined verb-phrase anaphora. There have also been no previous experiments conducted on differences in brain activity between surface and deep anaphora. The results of experiments 1a and 1b showed differences in naturalness ratings, reading times, and question answering times between deep and surface anaphora that at times interacted with the distance between the anaphor and the antecedent. Differences in BOLD

signal magnitude and areas of activity in response to processing deep versus surface anaphors and long versus short items would provide further evidence for a difference in how they are processed.

If imaging results are consistent with Sag and Hankamer's model, we would expect to see differences in activity within surface anaphora conditions between short and long conditions, with long conditions having more brain activity due to more difficulty in processing. Sag and Hankamer's model does not predict that we should see differences between deep short and deep long conditions, as they claim that these conditions should be equally acceptable. However, if our imaging results are consistent with our naturalness judgment experiment, we would expect to see differences in BOLD signal between deep short and deep long conditions, as these conditions had significantly different naturalness ratings. If deep and surface anaphors are processed using different mechanisms, we would expect significant contrasts of surface > deep or deep > surface, whereas if they processed using the same mechanism, we would not expect that these contrasts would be significant. If they were processed differently, we would also expect that the contrasts of surface long > surface short and deep long > deep short would be localized in different areas of the brain, whereas if they are processed using the same mechanisms, we would expect the contrasts to be located in similar areas of the brain.

Finally, as the distance reading time experiment showed differences between surface short and surface long conditions only in comprehension question response times and not in reading times at the anaphoric region, it would be of interest to compare imaging results to the reading time results. Imaging data could also potentially show differences in the surface long > surface short contrast in the anaphoric sentence that were not seen in the reading time data.

2.5.2. Methods

Participants 12 right-handed college-aged volunteers who had not participated in the reading time experiment were recruited from the Boston University community and paid for their participation. These subjects were native speakers of English from the USA who were naive to the purposes of the experiment, and informed consent was obtained prior to participation.

Materials 30 items of the format described in the Distance Materials section were used in the distance fMRI experiment. Each subject saw all 4 conditions of each item for a total of 120 items. Conditions from the same item were divided across runs to avoid familiarity effects, so that subjects did not see more than one condition from a particular item in the same run. Every item was followed by a comprehension question. The items were interspersed with 120 fillers. Complete lists of target items and fillers used in the fMRI experiment can be found in Appendices A and B respectively.

Procedures fMRI stimuli were presented using a rapid-presentation event-related (RPER) design. Each trial began with a 300 msec fixation block (a centered '+'), followed by a 100 msec blank screen block. These blocks were followed by presentation of the first sentence for 3 seconds, then presentation of the second/intervening sentence (in the long condition) or third/ anaphoric sentence (in the short condition) for another 3 seconds, which was then followed by the third/anaphoric sentence in the long condition for another 3 seconds. These sentences were followed by a comprehension question which appeared for 3 seconds, during which the subject responded "yes" or "no" by pressing a button with either the index or middle finger of the left hand. In the long conditions, the comprehension questions were followed by an additional 1-second blank screen block so that the lengths of these trials were multiples of the TR. All

conditions were then followed by a 600 msec blank screen block, so that short conditions were 10 seconds long and long conditions were 14 seconds long.

A variable fixation block was then presented, with a duration of 0, 2, 4, or 6 seconds. The length of the variable fixation trial and the trial type presentation order were determined using optseq2, a program developed to optimize the estimation of the hemodynamic response and reduce subject habituation and expectation (Burock et al, 1998; Dale, 1999; Dale and Buckner, 1997).

Including variable fixation trials, the functional part of the distance experiment lasted approximately 55 minutes. The distance experiment was divided into 6 runs of 9.17 minutes (550 seconds, or 275 time points) each. Subjects were allowed a short rest period between runs. Visual stimuli were projected onto a screen at the back of the scanner using an LCD projector, and they were viewed by subjects using a mirror attached to the head coil. Subjects recorded their responses using an MRI-compatible button box held in the left hand, with a button under the middle finger indicating a true comprehension question and a button under the index finger indicating a false comprehension question. An Apple PowerBook G4 computer running PsyScope software (Cohen et al, 1993) was used to present the stimuli and to record responses and reaction times.

MR Imaging Parameters All images were acquired using a 3.0 T whole-body Siemens Trio scanner (Siemens Medical Solutions USA Inc., Malvern, PA, USA). Following a brief localizer and auto-align scan paradigm, one high-resolution anatomical image was acquired using a T1-MPRAGE sequence (TR=2530 msec, TE=3.39 msec, flip angle = 7°). The anatomical image

volume consisted of 128 sagittal slices with an effective thickness of 1.33 mm and an in-plane resolution of 1.0 mm X 1.0 mm (256 X 256 matrix size, 256 mm FOV).

Functional volumes were acquired using a T2*-weighted gradient-echo pulse sequence that is sensitive to blood oxygenation level-dependent (BOLD) signal (TR = 2 sec, TE = 30 msec, flip angle = 90°). Functional volumes consisted of 30 transverse slices aligned along the anterior/posterior commissure (AC-PC) plane as determined by a registration volume. The transverse slices had an effective thickness of 3.0 mm with a gap of 0.9 mm between slices. The in-plane resolution was 3.125 mm X 3.125 mm (64 X 64 matrix, 200 mm FOV). By definition, the 30 slices in each volume were acquired in one TR (2 sec), with a new volume collected in each TR. Each run in the distance experiment consisted of 275 volumes, resulting in 8250 images collected per run. 8 seconds (4 TRs) of RF pulse activations were presented prior to each run during which no visual stimuli were presented and no images were acquired, in order to maximize signal strength during the functional run.

fMRI Data Analysis The data from the distance experiment were analyzed to determine differences in BOLD signal responses to deep and surface anaphora. Each sentence in the set of 3 (for short items) or 4 (for long items) was analyzed separately. The difference between neural correlates of sentences containing deep versus surface anaphors (sentence 2 in short conditions, sentence 3 in long conditions) was examined. The effect of increasing length on each anaphor type was investigated, as in the reading time experiment, by comparing BOLD responses to the anaphoric sentence between the deep short and deep long conditions and between the surface short and surface long conditions.

All preprocessing and statistical analysis of the fMRI data was performed using the SPM8 software package. Preprocessing consisted of slice time correction, motion correction, intensity normalization, registration and resampling in standard MNI space, and spatial smoothing with a Gaussian kernel of 6 mm full width half maximum (FWHM). The functional images were then analyzed using a random effects model of an event-related paradigm, with the hemodynamic response function (HRF) modeled as a Gamma function. Functional images were high-pass filtered at a cutoff of 128 s to remove scanner drift effects. Outlier time points in the fMRI time series due to head motion or spikes in global mean intensity were then removed using the Artifact Detection Toolbox (ART) for SPM8 (developed by Shay Mozes and Susan Whitfield-Gabrieli - http://web.mit.edu/swg/software.htm, see also Mazaika et al 2007, Mazaika et al 2009, Mazaika et al 2005). For the group analysis, paired t-tests were performed on individual contrast images to determine whether individual voxel differences were reliably different from zero. The resulting statistical parametric maps were FDR-corrected for multiple comparisons at $p < 0.05$ with an additional cluster size threshold of 50 voxels. Local maxima for significant clusters and the brain regions encompassed by each cluster were determined using the SPM8 toolboxes aal (Tzourio-Mazoyer 2002) and xjView (developed by Cui, Li, and Song - http://www.alivelearn.net/xjview8/).

2.5.3. Results

Data for 12 subjects were included in the analysis. Eleven of these subjects also completed the particle shift fMRI experiment, with half participating in the distance experiment first and half participating in the particle shift experiment first. One subject was replaced due to

excessive head motion and questionable handedness. All subjects had accuracies higher than 83%, with an overall mean accuracy of 90.1%. Mean accuracy on target items alone was 89.7%.

Behavioral Results As in the reading time experiments and the particle shift fMRI experiment, all statistical tests were carried out with mixed linear models (lmer in R), using subjects and items as random effects and experimental conditions of interest as fixed effects (e.g. Van Dongen et al 2004, Jaeger 2008). 2x2 (short/long items, surface/deep anaphors) analyses were conducted on the comprehension sentence response times and accuracies. If the interaction was found to be significant, planned comparisons (t-tests) between short and long items were conducted on the comprehension question response times and accuracies for both surface and deep anaphors. Comprehension question response times are shown below in Figure 9 and comprehension question accuracies are shown in Figure 10.

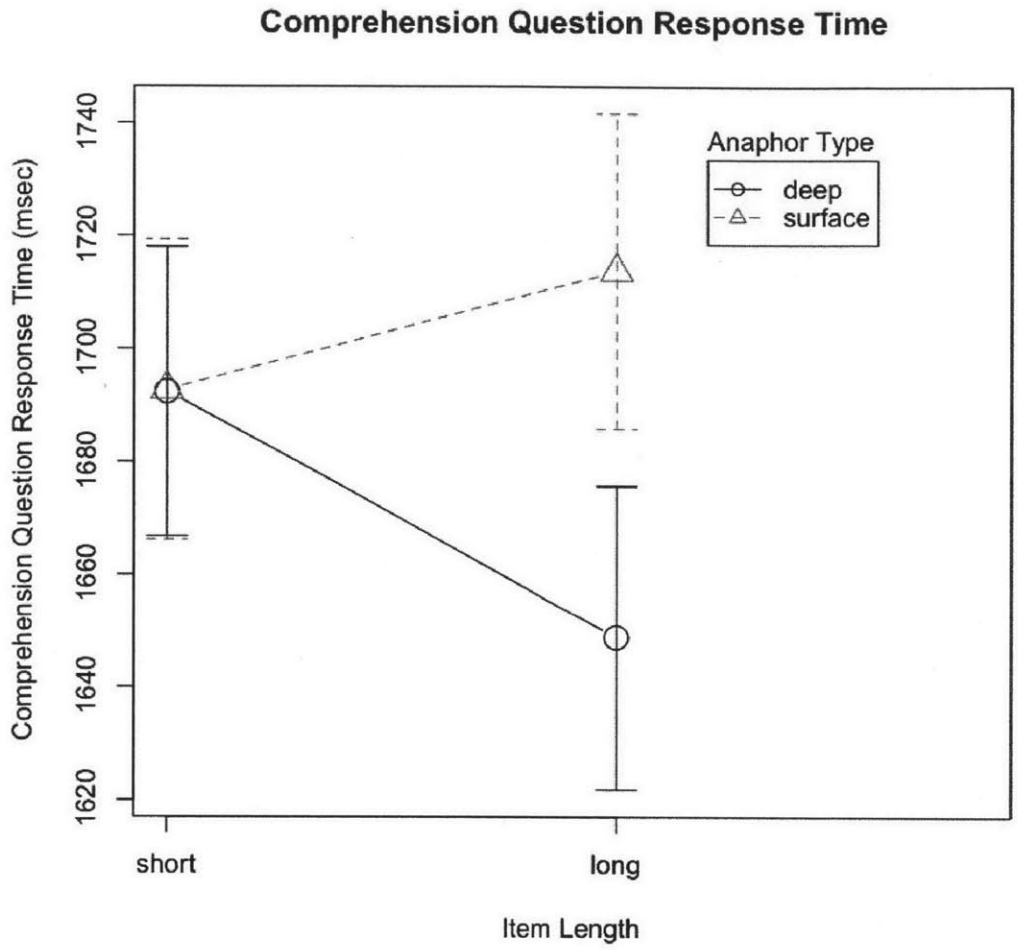

Figure 9: Comprehension question response times for the distance fMRI experiment.

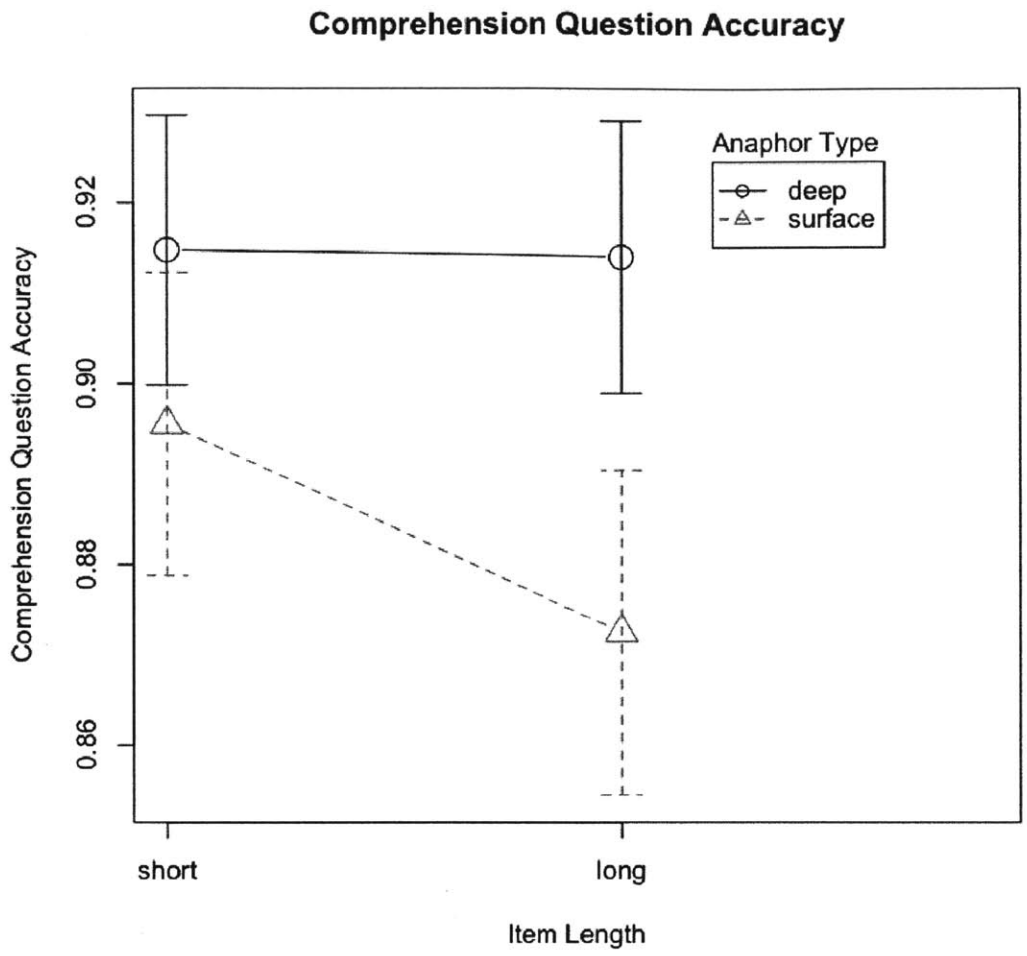

Figure 10: Comprehension question accuracies for the distance fMRI experiment.

For the comprehension question response times, there was a significant main effect of anaphor type ($F(1,1377=9.47$, $p < 0.01$), such that response times for surface anaphors were slower than response times for deep anaphors (mean surface = 1703 msec, mean deep = 1671 msec). The effect of distance and the interaction between distance and anaphor type were not significant (both $p > 0.1$).

For the comprehension question accuracies, there was a significant main effect of anaphor type ($F(1,1377)=7.2$, $p < 0.01$), such that subjects were more accurate on items with

deep anaphors than items with surface anaphors (mean deep accuracy = 91.4%, mean surface accuracy = 88.4%). The effect of distance and the interaction between distance and anaphor type were not significant (both $p > 0.1$).

fMRI Results For the contrasts of Deep > Surface and Surface > Deep for the anaphoric sentence only, no clusters of activation passed the threshold for FDR correction at $p < 0.05$.

For the contrast of Surface Long > Surface Short for the anaphoric sentence only, 5 clusters of activation passed significance for FDR-correction for multiple comparisons at $p < 0.05$ and a cluster size of 50 voxels. Data on the regions of activation are shown in Table 3, and images showing the activation clusters for this contrast are shown in Figure 11.

In the left hemisphere, one cluster (cluster #1 in Table 3) with its peak located in the left precentral gyrus extended into the left middle frontal and left inferior frontal areas, specifically the Pars Opercularis (BA 44). Another cluster in the left hemisphere (cluster #4 in Table 3) had a peak located in the left superior frontal gyrus, extending into the left middle frontal, left superior frontal, and left medial superior frontal areas. A final cluster in the left hemisphere (cluster #5 in Table 3) was located in the middle occipital gyrus.

In the right hemisphere, one cluster (cluster #2 in Table 3) had a peak in the right inferior frontal gyrus, specifically the Pars Opercularis (BA 44), extending into the Pars Triangularis (BA 45) and the right precentral and right middle frontal regions. A second cluster (cluster #3 in table 3) was located in the right supplemental motor area, and extended into the right precentral, left supplemental motor, right superior frontal, right postcentral, and right cingulum areas.

No clusters passed the threshold for FDR correction at $p < 0.05$ for the Deep Long > Deep Short contrast.

#	Location of local maximum	Regions encompassed by cluster	t_{max} (Z_{max})	peak-level p value (FDR-corrected at $p < 0.05$)	cluster size (in voxels)	Peak MNI Coordinates		
						X	Y	Z
1	L Precentral Gyrus	L Precentral, L Middle Frontal, L Inferior Frontal	5.57 (4.82)	0.026	358	-34	-6	48
2	R Inferior Frontal Gyrus	R Inferior Frontal, R Precentral, R Middle Frontal	5.53 (4.79)	0.026	452	36	8	32
3	R Supplemental Motor Area	R Supp. Motor, R Precentral, L Supp. Motor, R Superior Frontal, R Postcentral, R Cingulum	5.52 (4.78)	0.026	425	8	0	56
4	L Superior Frontal Gyrus	L Middle Frontal, L Superior Frontal, L Medial Superior Frontal	4.62 (4.15)	0.027	183	-20	50	12
5	L Middle Occipital Gyrus	L Middle Occipital	4.44 (4.01)	0.030	77	-30	-84	32

Table 3: Localization of activation in the distance experiment contrast "Surface Long > Surface Short" in the anaphoric sentence only. MNI Coordinates, p-, t-, and Z-values are reported for the voxel with maximum activity in each cluster.

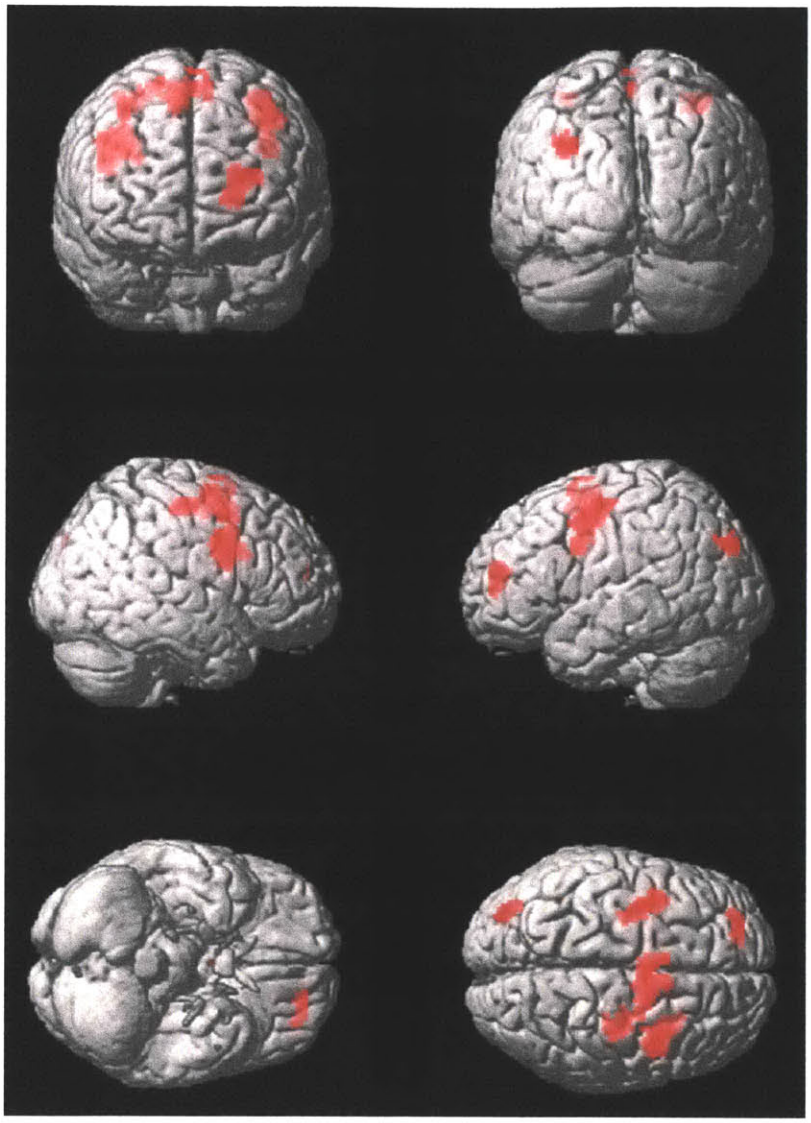

Figure 11: Visual representation of the areas of activation in the distance experiment contrast "Surface Long > Surface Short" in the anaphoric sentence only.

2.5.4. Discussion

In the behavioral data, there was a main effect of anaphor type in the comprehension question response time and accuracy, such that items with surface anaphors were answered more slowly and less accurately than items with deep anaphors. Deep and surface items were answered approximately equally quickly and accurately in the short conditions, but questions following surface long conditions were answered slower and less accurately than following deep long conditions (although the interactions were not significant). This result is consistent with the results from the distance reading time experiment, in which the comprehension question response time data showed similar patterns. Although not directly predicted by Sag and Hankamer's processing model, these results are consistent with the idea that surface long conditions are more difficult to process than surface short conditions whereas deep long conditions are not affected by distance in the same ways. Unfortunately the distance fMRI experiment was not set up in a way that made it possible to observe reading times for the anaphoric sentence (or any other sentences), so this result could not be compared to the distance reading time results.

No regions were found to be significantly active in the contrasts of surface > deep or deep > surface. This is inconsistent with the reading time results, which showed a significant difference in reading times between surface and deep anaphors (such that surface anaphors were read more slowly than deep anaphors).

Several regions were significantly active in the contrast of surface long > surface short, while no regions were significantly active in the contrast of deep long > deep short. This is consistent with a processing model that predicts that an intervening sentence will result in

increased activity in items with surface anaphors, as it will be more difficult to access the linguistic form of the antecedent, but will not affect items with deep anaphors due to a method of accessing the antecedent that is dependent on the discourse information and is not affected by distance.

One cluster of activation (cluster #1 in Table 3), which had a peak located in the left precentral gyrus and extended into the left middle frontal and left inferior frontal regions, and a second cluster (cluster #3 in Table 3) with its peak located in the right supplemental motor area which extended into the right precentral, left supplemental motor, right superior frontal, right postcentral, and right cingulum regions, are located in areas that have previously been related to the speech motor network. Neuroimaging studies have consistently implicated bilateral precentral and supplemental motor areas in tasks involving the production of speech (e.g. Dhanjal et al 2008, Fiez and Petersen 1998, Riecker et al 2005, Turkeltaub et al 2002). Additionally, the posterior region of Broca's area is thought to be involved in speech production (e.g. Fiez and Petersen 1998, Guenther 2006, Riecker et al 2005, Turkeltaub et al 2002), as are the anterior cingulate (e.g. Fiez and Petersen 1998, Guenther 2006) and areas of the postcentral gyrus (e.g. Fiez and Petersen 1998). However, subjects were instructed to read the stimuli silently, and head motion data indicated that subjects were not moving in a manner consistent with speech or articulation, indicating that the activations in these areas were not due to articulatory motion or subjects producing speech during the task.

The left-hemisphere areas that are generally related to speech motor activity, in which we saw activity (cluster #1 in Table 3), have also been shown to be active in subvocal rehearsal (no speech or articulatory movement) of verbal material in language and working memory tasks. By

comparing memory tasks to tasks requiring subvocal rehearsal, by comparing memory tasks that involved varying levels of subvocal rehearsal, or by comparing subvocal rehearsal tasks to controls involving articulatory suppression, many studies have consistently shown that the left pars opercularis (BA 44), precentral/premotor, and supplementary motor areas are involved in subvocal rehearsal of verbal material (e.g. Awh et al 1996, Chen and Desmond 2005, Gruber 2001, Koelsch et al 2009, McGettigan et al 2011, Narayanan et al 2005, Paulesu et al 1993, Smith et al 1998, Smith and Jonides 1998, Stowe et al 2005, Wager and Smith 2003, Woodward et al 2006). As the tasks in this experiment did not involve articulatory suppression, it is possible that these areas of activation reflect subjects' use of subvocal rehearsal strategies to maintain verbal information in memory prior to responding to the comprehension question. This strategy would have been utilized more in surface long versus surface short conditions, as long conditions require more information to be held in memory for a longer period of time. The absence of these areas in the deep long > deep short contrast may indicate that length did not affect the difficulty of deep conditions as much as the surface conditions, and subjects therefore did not have to utilize rehearsal strategies more in the deep long conditions compared to the deep short conditions.

Several areas of activation seen in this study have also previously been related to aspects of working memory besides subvocal rehearsal. Narayanan et al 2005 separated fMRI activations in response to the encoding, maintenance, and retrieval phases of a verbal working memory task, and found bilateral activation of BA 44 and BA 45 (included in clusters #1 and #2 in our results), the anterior cingulate (included in cluster #3 in our results), and the left middle occipital gyrus (included in cluster #5 in our results) in the encoding phase, and left hemisphere

activation of BA 44 and 45 in the retrieval phase of the task. A meta-analysis of working memory studies additionally associated the right inferior frontal gyrus (cluster #2 in our results) with executive control in working memory, whereas storage of verbal material was associated with activity in the left inferior frontal gyrus (Wager and Smith 2003). Additionally, bilateral middle frontal and superior frontal areas (included in clusters #2 and #4 in our results) have been implicated in explicit retrieval of information from memory (Buckner and Koutstaal 1998) and also in storage of verbal information and manipulation of information in working memory (Wager and Smith 2003). The task in this experiment involved storing information from the antecedent and intervening sentence in memory and retrieving this information in order to process the anaphoric sentence and the comprehension question, as well as storing information about the anaphoric sentence in memory in order to process the following comprehension question. It is therefore reasonable that the activations seen in response to this task have previously been shown to be involved in working memory functions. These activations could specifically reflect brain functions associated with retrieving information about the antecedent to process the anaphor, and encoding of this information in memory in anticipation of the upcoming comprehension question. These processes would be more difficult and would involve storage and retrieval of more information (due to the addition of the intervening sentence) in the surface long conditions as compared to the surface short conditions, which may be why we see these activations in this contrast. These areas are not seen in the deep long > deep short contrast, which could indicate that increasing length does not tax verbal working memory processes for deep anaphora as much as it does for surface anaphora.

Other areas of activation seen in our results, namely the bilateral inferior frontal, middle frontal, and superior frontal areas, have previously been related to language processing and syntactic complexity. The left inferior frontal gyrus, or Broca's area (BA 44, 45, and sometimes 47) is frequently activated in language studies that involve the comparison of syntactically complex as compared to simple sentences (e.g. object- vs. subject-relative), which could reflect increased syntactic demands and/or increased memory demands (Just et al 1996, Kaan and Swaab 2002, Keller et al 2001). Kaan and Swaab 2002 also related activation in the left middle frontal and superior frontal gyrus (cluster #4 in our analysis) to the same comparison of syntactically complex vs. simple sentences. The activation in these areas in our results could therefore reflect the increased syntactic or memory demands of processing a surface anaphor whose antecedent is separated by an intervening sentence as compared to a surface anaphor that directly follows its antecedent, especially if surface anaphors access a syntactic representation of their antecedents as predicted by Sag and Hankamer. The absence of these areas in the deep long > deep short contrast could suggest different levels of processing in deep versus surface anaphora, specifically that surface anaphors use syntactic information which becomes more difficult to access with additional length, whereas deep anaphors do not use this information.

The precise location of left inferior frontal gyrus activation in our results is of interest. Although Broca's area as a whole has been frequently implicated in many kinds of language tasks, as described above, several research groups have subdivided Broca's area into subsections that perform different language-related tasks. The activation seen in this study is in the dorsal posterior area of pars opercularis (BA 44), which has been previously associated with syntactic processing, as opposed to the ventral/anterior areas, which have been associated with semantic

processing (e.g. Friederici et al 2000, Newman et al 2003, Dapretto and Bookheimer 1999). Bookheimer 2002 proposed a further functional subdivision of Broca's area into three subsections, with the most anterior/ventral parts of BA 47 and 45 functioning in semantic processing, more central parts of BA 45 and 44 functioning in syntactic processing, and the most posterior/dorsal parts of BA 44 extending into BA 6 functioning in phonological processing (although they state that these areas overlap considerably). Vigneau et al 2006 performed a large-scale meta-analysis of 129 studies involving neuroimaging of language, and found results that fall somewhere between these two schools of thought; they propose that the dorsal/posterior region of Broca's area is involved in phonological and syntactic language functions (with no clear regional subdivision between these functions), while the anterior/ventral region is involved in semantic functions. The activation seen in the pars opercularis in our results is therefore consistent with regions of Broca's area that have been associated with syntactic and/or phonological processes previously. This could indicate that subjects are using syntactic or phonological information in relating the surface anaphors to their antecedents.

The inferior frontal, middle frontal, and superior frontal areas seen activated in our results have also been related to studies of discourse processing and inferences or connections across sentences. Hagoort 2005 suggested that Broca's area is responsible for unification in language, which he defined as the integration of semantic, syntactic, or phonological information as it becomes available into developing representations of syntactic, discourse, or intonational phrasing structures. In our materials, subjects must incorporate information about the antecedent and the intervening sentence while processing the anaphoric sentence, which could involve semantic and/or syntactic unification processes. Ferstl and von Cramon 2001 compared sentence

pairs that were either coherent, such as "Sometimes a big truck drives by the house. The dishes start to rattle." or incoherent, such as "Sometimes a big truck drives by the house. The car doesn't start." They found that the left superior frontal gyrus (cluster #4 in our results) was found to be active in the comparison of coherent to incoherent sentence pairs, and concluded that this area (along with the posterior cingulate and inferior precuneus, which were not activated in our results) is involved in inferential processing or the connection of adjacent related sentences. In a more thorough investigation of the effects of relationships between sentences, Kuperberg et al 2006 compared neural responses to three-sentence scenarios where the third sentence in the scenario was highly causally related, intermediately related, or unrelated to the first two sentences. The intermediately related sentences were found to generate causal inferences, while no causal inferences were generated in response to the highly related or unrelated scenarios. Comparing the intermediately causally related scenarios to the highly related and unrelated scenarios yielded activation in the bilateral inferior frontal gyrus (BA 45 and 47, not in 44 as seen in our left-hemisphere cluster), left middle frontal gyrus, and bilateral superior frontal gyrus and medial superior frontal gyrus (clusters #1, #2, and #4 in our results), along with inferior parietal and temporal areas not found in our results. These three studies indicate involvement of the inferior frontal, middle frontal, and superior frontal regions in cross-sentence connections and inferences. The distance task in our experiment involved connecting information from the antecedent, intervening, and anaphoric sentences to relate the antecedent to the anaphor correctly, and the activation that we found in these frontal areas could be related to these processes. The fact that we see activations in these areas in the surface long > surface short contrast but not in the deep long > deep short contrast could indicate that the processes involved in connecting a

surface anaphor to its antecedent become more difficult with the introduction of an intervening sentence, while the processes involved in connecting a deep anaphor with its antecedent are less affected by distance, which would be consistent with behavioral results that showed no distance effects for deep anaphors.

It is also possible to compare the results from this imaging study to results from imaging studies of noun phrase anaphora. Santi and Grodzinsky (2007b) compared reflexive binding, which has been proposed to be a syntactic (or "surface") function, to movement of a noun phrase and found activation in the right middle frontal gyrus and left inferior precentral regions, along with activation in temporal areas. Santi and Grodzinsky (2007a) also studied reflexive binding and syntactic movement, and found that binding resulted in activation in the right middle frontal gyrus and left inferior frontal gyrus along with the medial cingulate and temporal areas, while syntactic movement resulted in activation in the left inferior frontal gyrus along with the left superior temporal gyrus and left inferior precentral sulcus. Nieuwland et al (2007) compared activations between ambiguous pronouns and coherent pronouns, and found that ambiguous pronouns > coherent pronouns activated bilateral medial frontal, right superior frontal, bilateral medial parietal, and bilateral lateral parietal areas, which they claimed reflected areas involved in selecting the correct antecedent from the discourse. Coherent pronouns > ambiguous pronouns activated bilateral inferior frontal areas, which they claim represents successful binding of the antecedent to the anaphor. Finally, Almor et al (2007) compared repeated names (which are considered by Cloitrew and Bever 1988 to be "surface" anaphors) to pronouns (which are considered to be "deep" anaphors), and the contrast of repeated names > pronouns yielded activation in the middle and inferior temporal gyri as well as in the inferior parietal sulcus. Areas

common to these studies of noun phrase anaphora and our results are the left inferior frontal, right superior frontal, and right middle frontal regions. These regions may be responsible for interpreting or processing anaphora regardless of the type of antecedent (noun phrase versus verb phrase). The fact that we find these activations in the surface long > surface short contrast and not the deep long > deep short contrast may indicate that surface and deep anaphors are processed differently. Specifically, the interpretation or processing of the surface anaphor could be made more difficult with the introduction of an intervening sentence, while the processes involved in interpreting or processing the deep anaphor could be less affected by distance.

In summary, the fMRI results showed differences between the surface long > surface short contrast but not the deep long > deep short contrast, which is consistent with a processing model that predicts increased difficulty with distance for surface but not deep anaphors. The specific locations of activation are consistent with the increased language processing and working memory demands associated with accessing a more distant antecedent, with some locations of activation indicating the possible presence of subvocal rehearsal strategies.

3. The Effect of Word Order on Deep and Surface Anaphora (Particle Shift)

3.1. Previous studies: parallelism

Several studies have varied syntactic parallelism between the antecedent and the anaphor in order to test Sag and Hankamer's claim that parallelism is necessary for surface anaphors but not deep anaphors. In Murphy (1985) Experiment 3, an experiment was conducted using both verb-phrase ellipsis (VPE) and "do it" anaphora, where parallelism was varied by passivizing the antecedent sentences and "in some cases changing the wording slightly to improve naturalness", and distance was also varied as in Experiments 1 and 2. No examples of the stimuli are provided, so it is not known whether the passive antecedent sentences were long or short, which can have an effect on results due to topic shifts potentially being introduced in long materials (Mauner et al 1995). The results showed a main effect of distance, in that reading of both surface and deep anaphors slowed with increased distance (significant over items, $F(1,23) = 11.60$, $p < 0.005$; marginally significant over subjects, $F(1,32) = 2.99$, $p < 0.10$). There was no significant interaction of distance and anaphor type. There was a significant interaction of parallelism and distance, such that nonparallel antecedents slowed reading times for the anaphors in the "close" distance condition but not in the "far" distance condition (reliable over subjects, $F(1,31) = 4.90$, $p < 0.05$, but not over items). The important interaction of parallelism and anaphor type was not significant, indicating that parallelism did not differentially affect surface and deep anaphors, and the three-way interaction between parallelism, anaphor type, and distance was also not significant, indicating that these results do not support the prediction that parallelism will affect surface anaphors more than deep anaphors when the distance between the

antecedent and anaphor is increased. These results instead indicate that both surface and deep anaphors are affected by the linguistic form of the antecedent when the antecedent is close, and neither is affected when separated from the antecedent.

Tanenhaus and Carlson (1990) performed three experiments varying parallelism of deep and surface anaphors while keeping distance constant. The first experiment used a plausibility judgment task with two-line items, where the first line of an item was the antecedent and the second line was the anaphor. The parallelism of the antecedent was varied by using either an active (parallel) or passive (nonparallel) voice, as the anaphor was always active. An example of an item provided in the paper is shown below:

33. 1a. Someone had to take out the garbage. (active - parallel)

1b. The garbage had to be taken out. (passive - nonparallel)

2a. But Bill refused to. (surface)

2b. But Bill refused to do it. (deep)

The subjects made a timed plausibility judgment of the materials, where half of the filler sentences were constructed to be implausible such as the following items:

34. After the exam Bill decided to have a beer or two. Sam didn't either.

35. Yesterday, the sports star announced his retirement. Sam denied it, too.

The judgment response time data showed no significant interaction between anaphor type and parallelism, as the response latency for both deep and surface anaphors were equally affected by nonparallel antecedents. However, the percentage of items judged to make sense (judgment data) did show an interaction between anaphor type and parallelism: positive judgments decreased reliably for surface anaphors from 89% for parallel antecedents to 70% for nonparallel antecedents but not for deep anaphor conditions (94% for parallel antecedents, 91% for nonparallel antecedents).

In a second experiment, Tanenhaus and Carlson (1990) used verbal-nominal pairs instead of active-passive pairs to vary parallelism, as shown below:

36. 1a. It would do you good to jog into town. (verbal - parallel)

1b. A jog into town would do you good. (nominal - nonparallel)

2a. Please let me know if you decide to. (surface)

1b. Please let me know if you decide to do it. (deep)

The authors achieved similar results as in the active/passive pair experiment, with a significant interaction between anaphor type and parallelism for the judgment data (surface: 89% for parallel antecedents, 71% for nonparallel antecedents; deep: 86% for both parallel and nonparallel antecedents) but no significant interaction in the judgment response latency.

In a third experiment, Tanenhaus and Carlson used a null complement for the deep anaphor instead of "do it" to explore whether the results in Experiments 1 and 2 were affected by

the phonological presence of the deep anaphor. Both passive and nominal antecedents were used to make the antecedent nonparallel. An example of an item with a passive nonparallel antecedent is as follows:

37. 1a. Someone has to take out the garbage. (active - parallel)

1b. The garbage has to be taken out. (passive - nonparallel)

2a. But Bill refused to. (surface)

2b. But Bill refused. (deep)

Once again, the results showed a significant interaction in the judgment data between anaphor type and parallelism (surface anaphors: 95% for parallel antecedents, 77% for nonparallel antecedents; deep anaphors: 92% for parallel antecedents, 89% for nonparallel antecedents) and no interaction in the latency data (response times increased for deep and surface anaphors after nonparallel antecedents). This indicates that the results found in Experiment 1 and 2 were not due to the phonological presence of the deep anaphor, as similar results were found when the phonological presence of the deep anaphor is removed.

In all three experiments in Tanenhaus and Carlson (1990), deep and surface anaphors were affected differentially by the parallelism of the antecedent in terms of plausibility judgments, but not in terms of the latency of these plausibility judgments. However, the plausibility of passive antecedents followed by active surface anaphors can vary widely from very unacceptable to mostly acceptable depending on the length and structure of the antecedent

and anaphor (as noted in Belanger 2004). In this study, materials were not controlled for length and structure between different conditions. Belanger (2004) suggested that the lack of an interaction in the judgment latency data could be due to the lack of controlled length and structure in the materials across conditions, and that controlling the length and structure would directly affect the plausibility judgment data in these experiments.

Mauner et al (1995) addressed the lack of control for length and antecedent structure in Tanenhaus and Carlson (1990) by conducting two additional studies that controlled the structure of the passive antecedent. They noted that Tanenhaus and Carlson (1990) used both long passives with agent by-phrases such as:

38. 1a. John broke the antique vase which belonged to Mrs. Jones. (active - parallel)

1b. The valuable antique vase which belonged to Mrs. Jones was broken by John. (passive - nonparallel)

2a. She was furious that he did it. (deep)

2b. She was furious that he did. (surface)

and short passives without agent by-phrases such as:

39. 1a. Tom hadn't finished the report yet. (active - parallel)

1b. The report hadn't been finished yet. (passive - nonparallel)

2a. Tom said that he hadn't been able to do it. (deep)

2b. Tom said that he hadn't been able to. (surface)

Mauner et al (1995) pointed out that the long passives sound less natural than the short passives, and conducted a post-hoc item analysis of Tanenhaus and Carlson's (1990) results which showed that differential effect of parallelism on deep and surface anaphors was greater in the short passives than in the long passives. They also suggested that the inclusion of both types may have masked any presence of an interaction between anaphor type and parallelism in the judgment latency data.

Mauner et al (1995)'s first follow-up study was similar to Tanenhaus and Carlson's (1990) study but included only short passives without agent by-phrases such as:

40. 1a. Someone needs to feed the kitten. (active - parallel)

1b. The kitten needs to be fed. (passive - nonparallel)

2a. Joey forgot to again. (surface)

2b. Joey forgot to do it again. (deep)

As in Tanenhaus and Carlson (1990), they used a plausibility judgment task and measured percent of positive judgments ("makes sense") and response latency. Using only short passives, they found a significant interaction in the judgment data similar to that found in Tanenhaus and Carlson (1990), where positive judgments were lower for nonparallel than parallel antecedents

for surface anaphors (94% with parallel antecedent, 82% with nonparallel antecedent) but were similar across antecedent type with deep anaphors (96% for both parallel and nonparallel antecedents). Unlike Tanenhaus and Carlson (1990), Mauner et al also found an interaction between parallelism and anaphor type in the response latency data. This interaction mirrored that of the judgment data, with surface anaphors judged more slowly for nonparallel antecedents than parallel antecedents (2979 msec for parallel antecedents, 3531 msec for nonparallel antecedents), and similar response times for deep anaphors across antecedent condition (3075 msec for parallel antecedents, 3149 msec for nonparallel antecedents).

In a second experiment, Mauner et al used only long passives (which were shorter than Tanenhaus and Carlson 1990's long passives) with agent by-phrases such as:

41. 1a. Someone needs to feed the kitten. (active - parallel)

1b. The kitten needs to be fed by someone. (passive - nonparallel)

2a. Joey forgot to again. (surface)

2b. Joey forgot to do it again. (deep)

The phrase "by someone" was always used as the agent by-phrase so that the long passive sentences would be comparable to the interpretation of both the active sentences and the short passive sentences in Experiment 1. Using the long passives, the interactions between parallelism and anaphor type were eliminated for both the judgment data and the latency data. In both the judgment data and the latency data, only the main effect of parallelism was significant, such that

both surface and deep anaphors were judged as less likely to make sense after a nonparallel antecedent, and judgment latency increased for both surface and deep anaphors after nonparallel antecedents. Mauner et al claim that the long passives are more awkward and unnatural than short passives, which influenced the effects of parallelism in these materials. They argued that because the interaction effects were eliminated using the long passive materials, the lack of an interaction in the response time data in Tanenhaus and Carlson (1990) could reflect the varied acceptability and awkwardness of the materials used in that study. Their concluding argument is that the evidence from short passives supports Sag and Hankamer's claim that syntactic parallelism is necessary for surface anaphors but not for deep anaphors, but that pragmatic and stylistic factors can influence the effects of parallelism, such as in the long passives.

3.2. Previous studies: particle shift

Tanenhaus et al (1985) looked at effects of deep and surface anaphora without introducing additional distance and without varying parallelism with passivization or nominalization. In these materials, the antecedent sentence included a particle+verb construction, followed by a second sentence containing either a deep ("did it") or surface (VPE) anaphor. A third "verification" sentence then followed, which the subject had to categorize as true or false. Importantly, this verification sentence included the same particle+verb as the first sentence. The order of the particle/direct object construction was varied to either match or mismatch the order of the particle/direct object construction in the verification sentence, as in this example:

42. 1a. Jenny asked Ann's boyfriend out yesterday. (match)

1b. Jenny asked out Ann's boyfriend yesterday. (mismatch)

2a. Ann was furious that she did. (surface)

2b. Ann was furious that she did it. (deep)

3. Jenny asked Ann's boyfriend out. (verification sentence)

Tanenhaus et al predicted that if surface anaphors access some surface level of processing, the surface form of the antecedent sentence would be more salient during the processing of the surface anaphor than during processing of the deep anaphor due to a reactivation of the surface form. If this is the case, Tanenhaus et al predicted that when the particle/direct object order of the antecedent and verification sentences were mismatched, the verification sentence would be more difficult to read following the surface anaphor than following the deep anaphor, as the surface form of the mismatching antecedent would be more salient in the case of the surface anaphor. In the case of the deep anaphor, Tanenhaus et al predicted that the deep anaphor would access a non-linguistic level that is insensitive to the particle/direct object order. In this case the surface form of the antecedent sentence should decay normally by the time the verification sentence is presented and there should be no difference in verification sentence reading times between the mismatch and match conditions.

One advantage of using the particle shift paradigm instead of the parallelism paradigms used by other groups is that grammaticality and awkwardness are not factors in the interpretation of the materials, as both versions of the particle/direct object order are equally grammatically

acceptable (e.g. Svenonius 1996), and surface and deep anaphors are equally acceptable after either antecedent. An advantage of the particle shift paradigm over paradigms using increased distance is that added distance could potentially increase the complexity of the conceptual form of an antecedent as well as making the surface form less available, which could inadvertently complicate the processing of deep anaphors as well as surface anaphors.

As predicted, the verification response time after a surface anaphor increased by 627 msec in the mismatch condition as compared to the match condition (2190 msec for the match condition, 2817 msec for the mismatch condition), while the response time after a deep anaphor decreased by 31 msec in the mismatch condition as compared to the match condition (2320 msec for the match condition, 2289 msec for the mismatch condition). However, no statistics were provided, nor were details regarding subjects or procedure. Additionally, no control condition was provided as a baseline to compare to the deep/surface anaphor conditions, making it impossible to know how the anaphor conditions compared to conditions with "normal" decay.

Belanger (2004) attempted to reproduce Tanenhaus et al (1985)'s particle shift experiment with control sentences. Belanger used stimuli similar to those in Tanenhaus et al (1985), but added three control conditions (null, adjective, and neutral) that were controls for the deep and surface anaphors. An example is as follows:

43. 1a. The travel agent messed up Frank's booking. (matching antecedent)

1b. The travel agent messed Frank's booking up. (mismatching antecedent)

2a. He was very disappointed that she did. (surface anaphor)

2b. He was very disappointed that she did it. (deep anaphor)

2c. He was very disappointed. (adjective control)

2d. He decided to spend the holidays at home. (neutral control)

2e. [no second sentence] (null control)

3. The travel agent messed up Frank's booking. (verification sentence)

The null control was designed to show the maximum difference between the match and mismatch conditions, as no intervening material was placed between the antecedent and the verification sentence, allowing the surface form of the antecedent to remain very salient. The adjective control was designed to show the effect of match/mismatch on the verification sentence directly before introducing the anaphor. This control was noted as being potentially problematic, as these sentences involve an intuitive completion or ellipsis, as in "He was very disappointed (that the travel agent messed up his booking)". Finally, the neutral control was designed to show the "normal" effect of decay of the antecedent word order over a sentence that had neither a surface nor a deep anaphor. The neutral control sentences included a pronoun that referred to the antecedent, but no other anaphors. One potential problem with the neutral control conditions was that the neutral sentences introduced additional information that was not included in the

anaphoric conditions, which could have caused more interference with the verification sentence than in the anaphoric conditions (but should not have affected match/mismatch conditions disparately). Another potential problem was that in the first experiment, the neutral sentences were not controlled for number of clause boundaries; the sentences in the surface and deep conditions were always complex (containing subordinate clauses) while the neutral sentences were sometimes simple clauses and sometimes complex. This issue was addressed in a second experiment.

The results of this first experiment showed a main effect of match condition, where verification sentences in the match condition were read faster than in the mismatch condition. The results also showed a marginally significant effect of sentence type, with verification sentences read marginally faster in the surface than in the deep condition (by items, $p<0.06$; by subjects, $p<0.09$) and marginally faster in the surface than in the null condition (by subjects, $p<0.08$), with no other significant differences between conditions. However, the important interaction between match and type was insignificant, indicating that the mismatch between the antecedent and verification sentence was not differentially affected by surface and deep anaphors.

In a second experiment, Belanger (2004) varied the position of the particle in both the antecedent and the verification sentence, whereas it had only been shifted in the antecedent in Experiment 1. Belanger also controlled the neutral sentences for number of clause boundaries. Finally, the fillers were adjusted and yes-no questions were added to 10% of the items to ensure that subjects were paying attention to all parts of the sentence. All other aspects of the materials were comparable to those in Experiment 1. In Experiment 2, the main effects of shift and match

were significant as in Experiment 1. Whereas Experiment 1 showed no main effect of sentence type, Experiment 2 showed a highly significant effect of sentence type, which arose because the null condition was slower than the other four sentence types. Belanger suggested that the null condition could have incurred some sort of reading time penalty associated with the presentation of two very similar or identical sentences following one another.

In Experiment 2, the important interaction between match and sentence type was significant by subjects but only approached significance by items. This interaction arose due to all sentences types being affected by mismatch except the deep anaphor condition. Even the neutral condition showed a slowing of verification time in the mismatch condition, indicating that surface form is more salient following "normal" decay than following the deep anaphor. Belanger suggested that these findings indicate that the surface form is actually suppressed in the deep anaphor condition, although it could also indicate that the surface form is simply bypassed in this condition. To complicate these findings, an analysis of the mean differences between the match and mismatch conditions within sentence type showed that only the surface and deep anaphor mean differences differed significantly from each other, with no other sentence conditions differing significantly from each other (notably, the mean differences between deep and neutral conditions did not significantly differ, and neither surface nor deep anaphors differed significantly from any of the control conditions).

The findings from both Experiment 1 and Experiment 2 of Belanger (2004) may be complicated by the problems with materials mentioned previously. Notably, the new information introduced by the neutral control condition could have contributed to additional difficulties with the verification sentence as compared to the surface and deep anaphor conditions, which did not

introduce comparable new information. The adjective control condition was also problematic due to its potential implied ellipsis, as noted by Belanger. Additionally, the null control condition was also potentially more difficult than intended, as no studies have been done on the difficulty of processing a sentence that is almost identical to the previous sentence (this condition was in fact demonstrated to be more difficult than all other conditions in Experiment 2). Another potential problem with Belanger's study is that only 4 items per condition were used in Experiment 1, and only 5 items per condition were used in Experiment 2, leading to a potential lack of sufficient data with which to draw statistically relevant conclusions.

3.3. Particle shift materials

For the particle shift experiments, each item consisted of 12 different versions of a set of sentences, where a set of sentences included an introductory antecedent sentence, an anaphoric sentence, and then a verification sentence. Within each item, conditions were varied systematically in terms of anaphor type (either neutral, surface, or deep) and match type (the particle location in the verification sentence was either a match or a mismatch to the particle location in the antecedent sentence). Additionally, each [anaphor type x match type] condition had both a shifted and an unshifted version of the antecedent sentence, over which analysis was collapsed.

The introductory sentence was always in the form of "[Subject] [verb] [particle in unshifted condition] [Object] [particle in shifted condition] [Prepositional phrase]", such as "Rebecca dropped off Jonathan at the pool" for the unshifted condition or "Rebecca dropped Jonathan off at the pool" for the shifted condition. The anaphoric sentence was always in the form of "[S/he (referring to object)] was [adjective of feeling] that [s/he (referring to subject)]

[anaphor referring to verb]", such as "He was excited that she drove/did/did it". For the surface anaphor conditions, the anaphor was in the form of verb-phrase ellipsis, or simply "did". For the deep anaphor conditions, the anaphor was in the form of "did it". For the neutral anaphor conditions, the anaphor was a very simple rephrasing of the verb phrase. The verification sentence was always an exact repetition of the antecedent sentence, except that the particle was moved in the mismatch condition. An example of one item for the particle shift experiment is shown below in Table 4.

anaphor type	match type	shift type	Antecedent Sentence	Anaphoric Sentence	Verification Sentence
neutral	match	unshifted	Rebecca dropped off Jonathan at the pool.	He was excited that she drove.	Rebecca dropped off Jonathan at the pool.
neutral	match	shifted	Rebecca dropped Jonathan off at the pool.	He was excited that she drove.	Rebecca dropped Jonathan off at the pool.
neutral	mismatch	unshifted	Rebecca dropped off Jonathan at the pool.	He was excited that she drove.	Rebecca dropped Jonathan off at the pool.
neutral	mismatch	shifted	Rebecca dropped Jonathan off at the pool.	He was excited that she drove.	Rebecca dropped off Jonathan at the pool.
surface	match	unshifted	Rebecca dropped off Jonathan at the pool.	He was excited that she did.	Rebecca dropped off Jonathan at the pool.
surface	match	shifted	Rebecca dropped Jonathan off at the pool.	He was excited that she did.	Rebecca dropped Jonathan off at the pool.
surface	mismatch	unshifted	Rebecca dropped off Jonathan at the pool.	He was excited that she did.	Rebecca dropped Jonathan off at the pool.
surface	mismatch	shifted	Rebecca dropped Jonathan off at the pool.	He was excited that she did.	Rebecca dropped off Jonathan at the pool.
deep	match	unshifted	Rebecca dropped off Jonathan at the pool.	He was excited that she did it.	Rebecca dropped off Jonathan at the pool.
deep	match	shifted	Rebecca dropped Jonathan off at the pool.	He was excited that she did it.	Rebecca dropped Jonathan off at the pool.
deep	mismatch	unshifted	Rebecca dropped off Jonathan at the pool.	He was excited that she did it.	Rebecca dropped Jonathan off at the pool.
deep	mismatch	shifted	Rebecca dropped Jonathan off at the pool.	He was excited that she did it.	Rebecca dropped off Jonathan at the pool.

Table 4: Example of one item for the particle shift experiment.

In addition to varying these variables of interest, the materials were also varied in other ways. The antecedent sentence always had one male and one female as the subject and object, and the gender of the subject and object was counter-balanced across items. The animacy of the object was also counter-balanced across items, so that half of the items had animate objects (such as "Jonathan") and half of the items had inanimate objects which are the possessions of gendered persons (such as "Bob's car"). The items were also counter-balanced in terms of using proper names versus professions in the object and subject positions in order to keep items interesting. Each particle+verb construction was used in four different items (48 distinct verb+particle

constructions are used), and the gender and proper name/profession categorization were counter-balanced across these four items.

The target items were interspersed with 90 fillers, designed to reduce the development of strategies and to ensure that subjects read all sentences in the set, as some fillers referred to the anaphoric sentence rather than the initial sentence. All fillers were in the same format as the items. The fillers varied whether the verification sentence was true or false, and whether the verification sentence referred to the antecedent sentence or the anaphoric sentence, as the target items contained only verification sentences that were true and referred only to the antecedent sentence.

30 of the fillers contained verification sentences that were 'true' verifications of the anaphoric sentence instead of the antecedent sentence in the set, such as

44. 1. The secretary filled Bob's form out during the meeting.
    2. He was thankful that she did
    3. He was thankful that she did. (True)

Half of the verification sentences in these 30 fillers were exact restatements of the anaphoric sentence (as in the example above), and half were reworded slightly (with the name substituted for "he" or "she", for example), such as

45. 1. The fireman let down Ruth with the rope.

2. She was grateful that he did it.

3. Ruth was grateful that the fireman did it. (True)

The 60 remaining fillers were sets in which the verification sentences were either 'false' verifications of the antecedent sentence (30 fillers), such as

46. 1. Rebecca turned on the client's computer at the table.

2. He was irritated that she did.

3. Rebecca turned the client's computer on at the desk. (False)

or 'false' matches to the anaphoric sentence (30 fillers), such as

47. 1. Andrea mixed up the client's column in the paper.

2. He was angry that she did.

3. He was glad that she did. (False)

The fillers that contain 'false' verifications of the antecedent sentence were balanced as to whether the particle location in the antecedent sentence and verification sentence matched or mismatched.

To summarize, there were 30 fillers in which the verification sentence was true and referenced the anaphoric sentence instead of the antecedent sentence, 30 fillers in which the

verification sentence was false and referenced the anaphoric sentence instead of the antecedent sentence, and 30 fillers in which the verification sentence was false and referenced the antecedent sentence. All 90 fillers were balanced for anaphor type, particle location, gender, object animacy, and proper name vs. profession of the subject/object.

3.4. Experiment 2a: Particle shift reading time study

3.4.1. Introduction

The particle shift paradigm is one way to test Sag and Hankamer's claim that surface anaphors involve reactivation of surface syntactic form. Sag and Hankamer claim that surface anaphors are processed by accessing surface syntactic information in the antecedent, while deep anaphors are processed by accessing discourse information that does not consider surface syntactic information. If this is true, the surface anaphor in the particle shift paradigm should reactivate the surface syntactic information in the particle shift construction, making the following verification sentence more difficult to process in the surface mismatch condition than in the surface match condition. In this case, we would expect that the surface mismatch verification sentence should be read more slowly than the surface match verification sentence. If the deep anaphor does not access surface syntactic information, as claimed by Sag and Hankamer, this information should decay normally over the presentation of the anaphoric sentence instead of being reactivated, and the verification sentence should be equally easy to process in the deep mismatch and deep match conditions. In this case, we would expect that the verification sentence reading time would be equal in the deep match and deep mismatch conditions. Since we are using both shifted and unshifted versions of the antecedent and

verification sentence particle verb constructions, any potential small variation in acceptability between shifted and unshifted constructions should be eliminated.

Unlike studies that vary parallelism between the antecedent and the anaphor, altering the particle verb order does not change the grammaticality of the sentence. An advantage of the particle shift paradigm over paradigms using increased distance is that added distance could potentially increase the complexity of the conceptual form of an antecedent (by introducing new discourse information) as well as making the surface form less available, which could inadvertently complicate the processing of deep anaphors as well as surface anaphors.

3.4.2. Methods

Participants 24 college-aged volunteers from the Boston University community were recruited who had not participated in the distance reading time study or either of the fMRI studies. These volunteers were native speakers of English from the USA who were naive to the purposes of the experiment. Informed consent was obtained prior to participation and subjects were paid for their involvement in the study.

Materials 192 items of the format described above in the Particle Shift Materials section were used in the reading time study. Subjects saw only one condition from each item to avoid familiarity effects. Subjects therefore saw 192 distinct items, or 32 items per condition. These sets were drawn from 12 distinct sets of sentences. The target items were interspersed with 90 fillers of the format described in the Particle Shift Materials section. Full lists of the target items and fillers used in the particle shift reading time experiment can be found in Appendices C and D respectively.

Procedures Materials were presented on a screen and reading times were measured using the Linger program (Rohde 2003). Subjects were instructed to press the space bar when they were done reading the words on the screen in order to progress to the next sentence or group of words.

All sentences were presented in a phrase-by-phrase paradigm in order to examine reading times for the verb+particle constructions in the antecedent sentence and verification sentence and the anaphoric phrase in the anaphor sentence. Masked presentation was used, so that only the phrase being presented was visible and all phrases before or after that phrase were replaced with one dashed line per character (including spaces). Phrases were defined as in the following example:

48. [Judy] [turned down Anthony] [in the meeting.]
[He was angry] [that she did.]
?? [Judy] [turned Anthony down] [in the meeting.] ??

The subject pressed the space bar to automatically advance to the next phrase. After the verification sentence was presented, the subject was asked to answer whether the verification sentence was true or false by pressing "F" for 'true' or "J" for 'false', with a reminder of which keys to press presented at the bottom of the screen.

3.4.3. Results

Data for 24 subjects were included in the analysis. One subject had to be replaced due to an average reading time that was below 2.5 standard deviations from the mean. Each subject had

an average accuracy of greater than 67%. The average accuracy for the true/false repeated statements was 93.4%.

For the reading time data, residual reading times were analyzed to account for differences in word length and for differences in participants' mean reading rates. To determine residual reading times from raw reading times, a linear regression equation was computed for each subject which predicted reading times as a function of word length using all items and fillers (e.g. Ferreira and Clifton 1986; Trueswell, Tanenhaus, and Garnsey 1994). The residual reading time for each region was then calculated by subtracting the predicted reading time as calculated by the subject's regression equation from the subject's measured raw reading time. Conceptually, a negative value of the residual reading time indicates that the reading time was faster than predicted, and a positive value of the residual reading time indicates that the reading time was slower than predicted. A less negative value of the residual reading time indicates a slower reading time than a more negative value of the residual reading time. Statistical results were consistent between raw reading times and residual reading times, but only residual reading times will be discussed.

Only items where the true/false repeated statement was answered correctly were included in the reading time analysis. Data points where the reading time was more than 3 standard deviations from the mean were removed from the analysis, which affected 1.7% of the data. Data points where the response time to the true/false repeated statement was more than 3 standard deviations from the mean were also removed from the analysis, which affected 1.2% of the data.

The items were divided into eight regions for the reading time analysis. These regions were numbered as follows:

49. [Judy] [turned down Anthony] [in the meeting.]
1 2 3

[He was angry] [that she did.]
4 5

?? [Judy] [turned Anthony down] [in the meeting.] ??
6 7 8

The critical region was region 7, which included the repeated verb+particle construction. All statistical tests were carried out with mixed linear models (lmer function in R), using subjects and items as random effects and experimental conditions of interest as fixed effects (e.g. Van Dongen et al 2004, Jaeger 2008). The raw reading times per region are shown below in Table 5 and the residual reading times per region are shown below in Figure 12.

	Region							
Condition	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Surface Match	611.7 (10.4)	1080.6 (19.3)	748.9 (13.2)	626.4 (10.1)	560.9 (10.4)	506.2 (6.8)	675.9 (10.0)	576.6 (8.8)
Surface Mismatch	612.8 (9.9)	1059.8 (18.7)	757.2 (13.0)	617.6 (9.0)	546.3 (10.0)	503.3 (6.6)	718.6 (12.2)	598.5 (10.0)
Deep Match	607.4 (9.7)	1089.0 (18.5)	756.1 (13.0)	633.2 (10.1)	587.8 (11.4)	513.3 (7.0)	684.4 (10.6)	592.5 (9.5)
Deep Mismatch	615.3 (10.0)	1081.6 (19.0)	757.1 (13.6)	625.1 (10.1)	546.9 (8.7)	506.9 (7.1)	710.3 (11.4)	598.7 (9.8)

Table 5: Raw reading times per region in the particle shift reading time experiment. Times shown are in milliseconds. Numbers in parentheses indicate standard error in milliseconds. The regions are presented in Example 31 above.

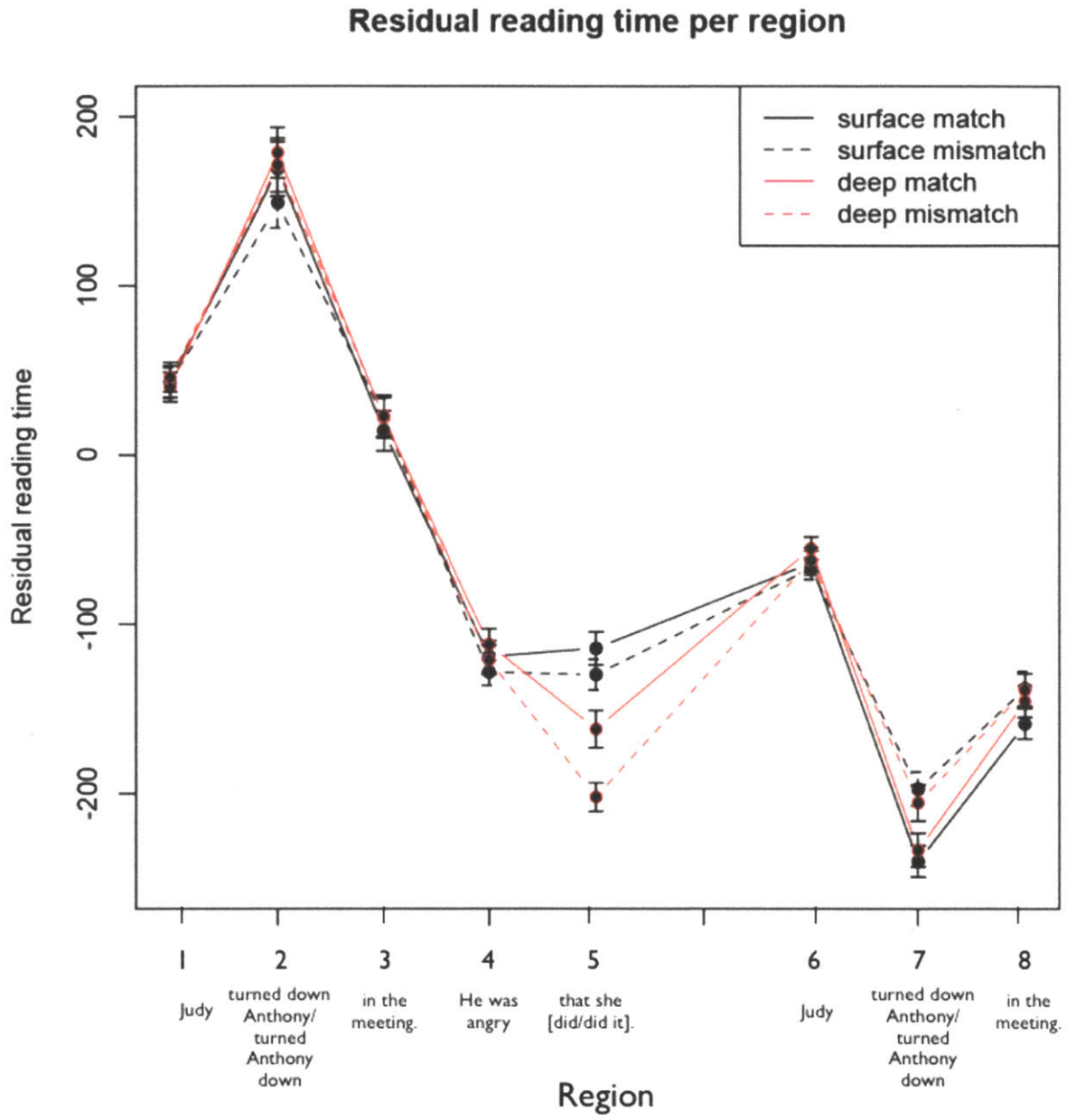

Figure 12: Residual reading time per region for the particle shift reading time experiment. Regions are defined in Example 31 above.

In region 2, the first particle+verb construction region, a pair-wise comparison of residual reading time by shift condition (shifted - "turned down Anthony" vs. unshifted - "turned Anthony down") revealed no significant difference between reading times for shifted versus unshifted

conditions ($F(1, 2918) = 0.0963$; $p = 0.76$). This indicates that the initial order of the verb +particle construction did not affect the reading time for that region.

In region 5, a 2 x 2 (shifted/unshifted first particle+verb construction, deep/surface anaphor) analysis was conducted on the residual reading time using a mixed linear model with subjects and items as random effects and experimental conditions of interest as fixed effects (e.g. Van Dongen et al 2004). There was a significant main effect of anaphor type ($F(1, 3027) = 45.2$; $p < 0.001$), such that deep anaphors were read more quickly than surface anaphors (mean residual reading time for deep anaphors = -182.2 msec, and for surface anaphors = -122.2 msec). Notably there was no significant effect of shift condition nor was there a significant interaction between shift condition and anaphor type on the anaphoric reading time (both $p > 0.5$), indicating that the order of the initial verb+particle construction did not affect reading times for the anaphoric region.

2 x 2 (match/mismatch conditions, deep/surface anaphor) analyses were then conducted on residual reading times for each region individually. When significant interactions were found, planned comparisons between match and mismatch conditions within anaphor types were conducted (surface mismatch versus surface match and deep mismatch versus deep match). All statistics were calculated using mixed linear models with subjects and items as random effects and experimental conditions of interest as fixed effects. For the 1st-4th regions, the main effects of anaphor type and match type and the interaction between anaphor and match were not significant (all $p > 0.1$). This is expected as the materials do not differ at all across either match condition or anaphor type in these regions.

In the 5th region, which is the region containing the anaphor, there was a significant main effect of match ($F(1,3027)=9.4$, $p < 0.01$), such that match conditions were read more slowly than mismatch conditions (mean match = -138.32 msec, mean mismatch = -166.01 msec). This is not an expected result, as the items do not differ at this point between match and mismatch conditions. This cannot be explained by prediction effects, as the upcoming sentences look identical between match and mismatch conditions. There was also a significant main effect of anaphor type ($F(1,3027)=45.4$, $p <0.001$), such that surface anaphors were read more slowly than deep anaphors (mean deep = -182.18 msec, mean surface = -122.13 msec). The interaction between match and anaphor type was not significant ($p > 0.1$).

In the 6th region, which includes the subject of the repeated sentence, the main effects of anaphor type and match type and the interaction between anaphor and match were all insignificant (all $p > 0.1$).

In the 7th region, which was the critical region including the repeated verb+particle construction, there was a significant main effect of match ($F(1,3052)=15.2$, $p > 0.001$), such that mismatch conditions were read significantly more slowly than match conditions (mean match = -236.74 msec, mean mismatch = -201.44 msec). The main effect of anaphor type and the critical interaction between match and anaphor conditions were insignificant ($p > 0.1$). The residual reading time data for only this region are shown below in Figure 13.

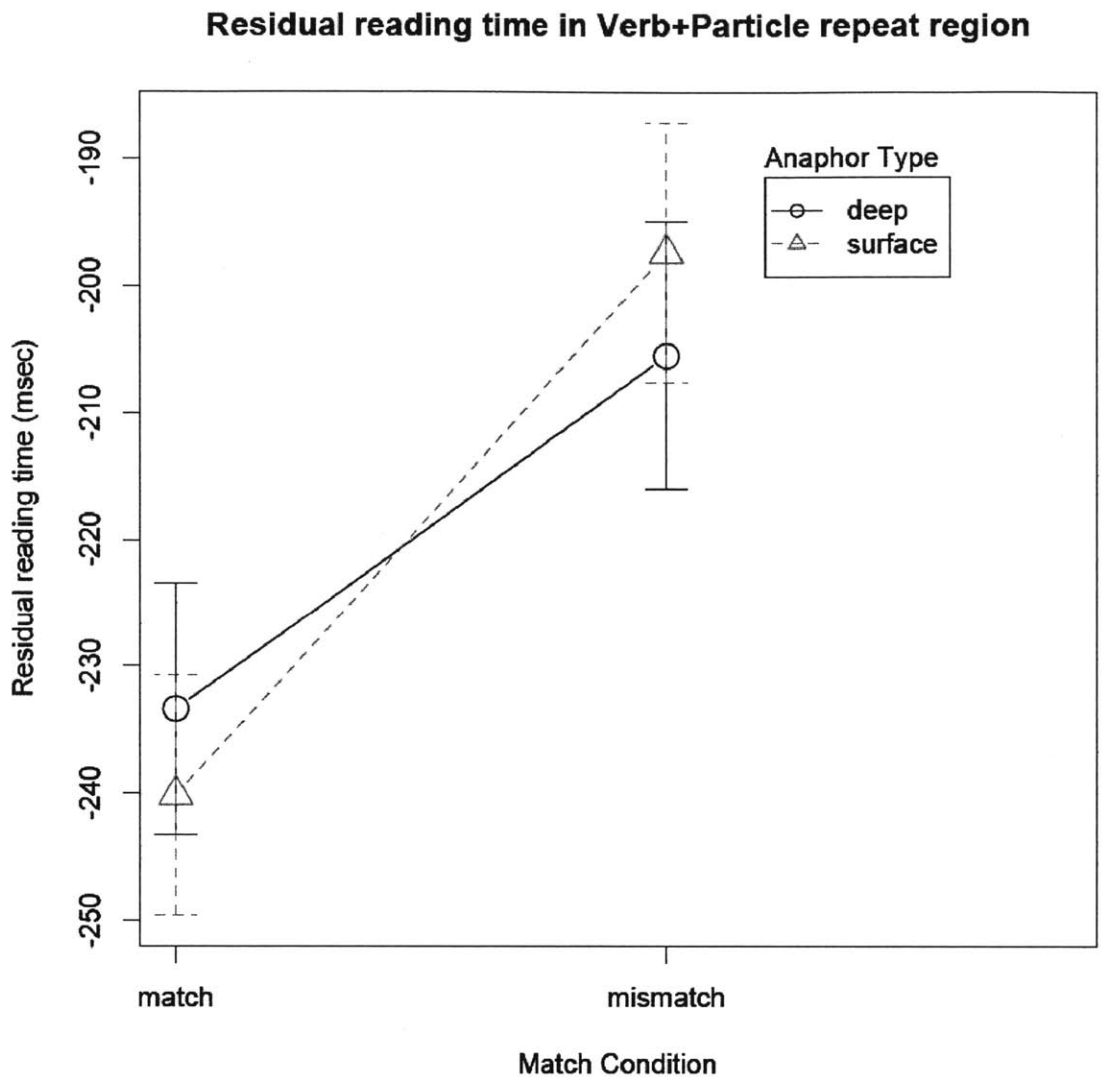

Figure 13: Residual reading time for region 7, the verb+particle repeat region, for the particle shift reading time experiment.

In the 8th region, which was the end of the true/false verification sentence, the main effects of anaphor type and match type and the interaction between anaphor and match were all insignificant (all $p > 0.1$).

For the question accuracy and question response time, the main effects of anaphor type and match type and the interaction between anaphor and match were all insignificant (all $p >$

0.1). The question accuracy and question response time are shown below in Figures 14 and 15 respectively.

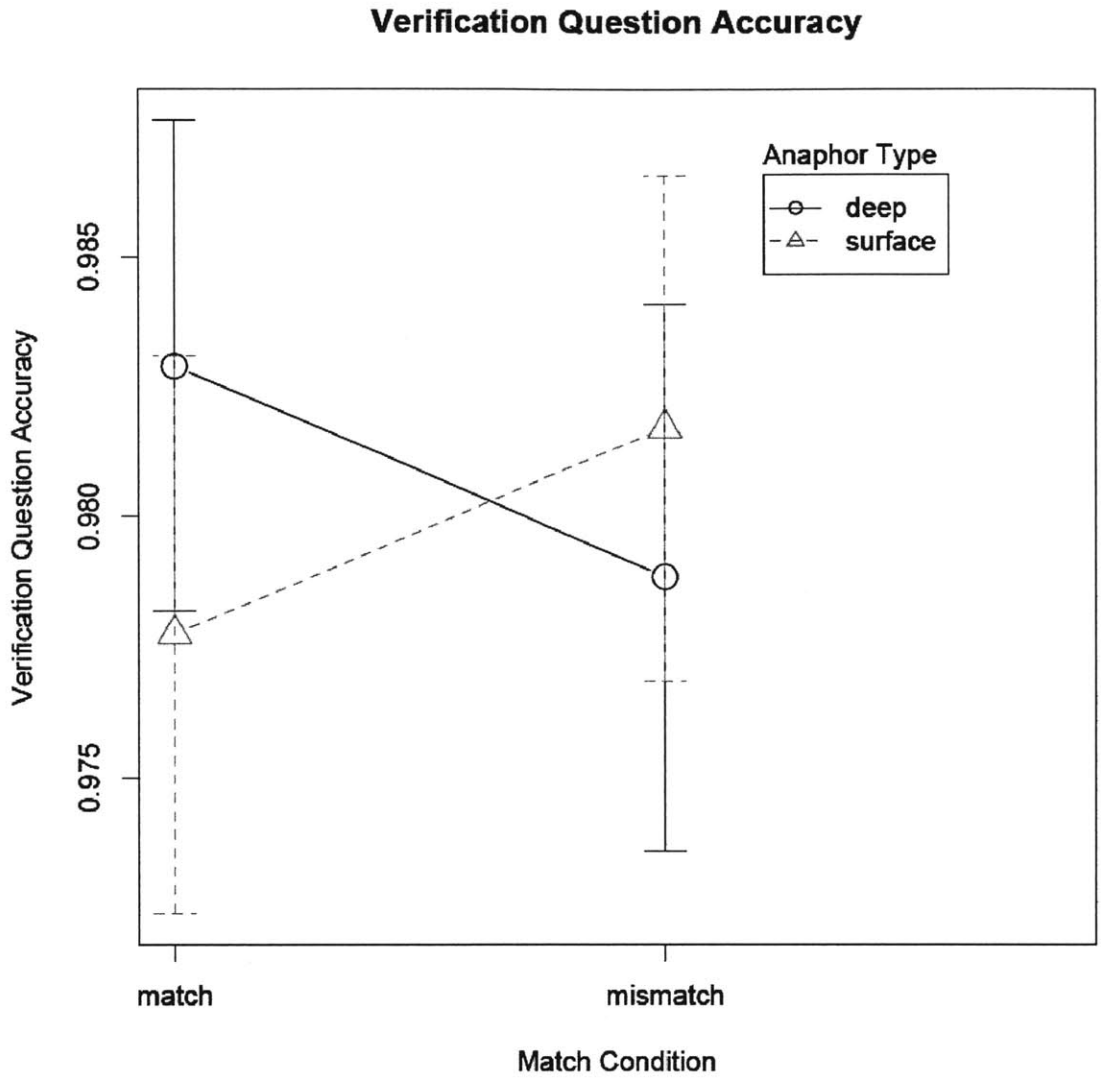

Figure 14: Verification question accuracy for the particle shift reading time experiment.

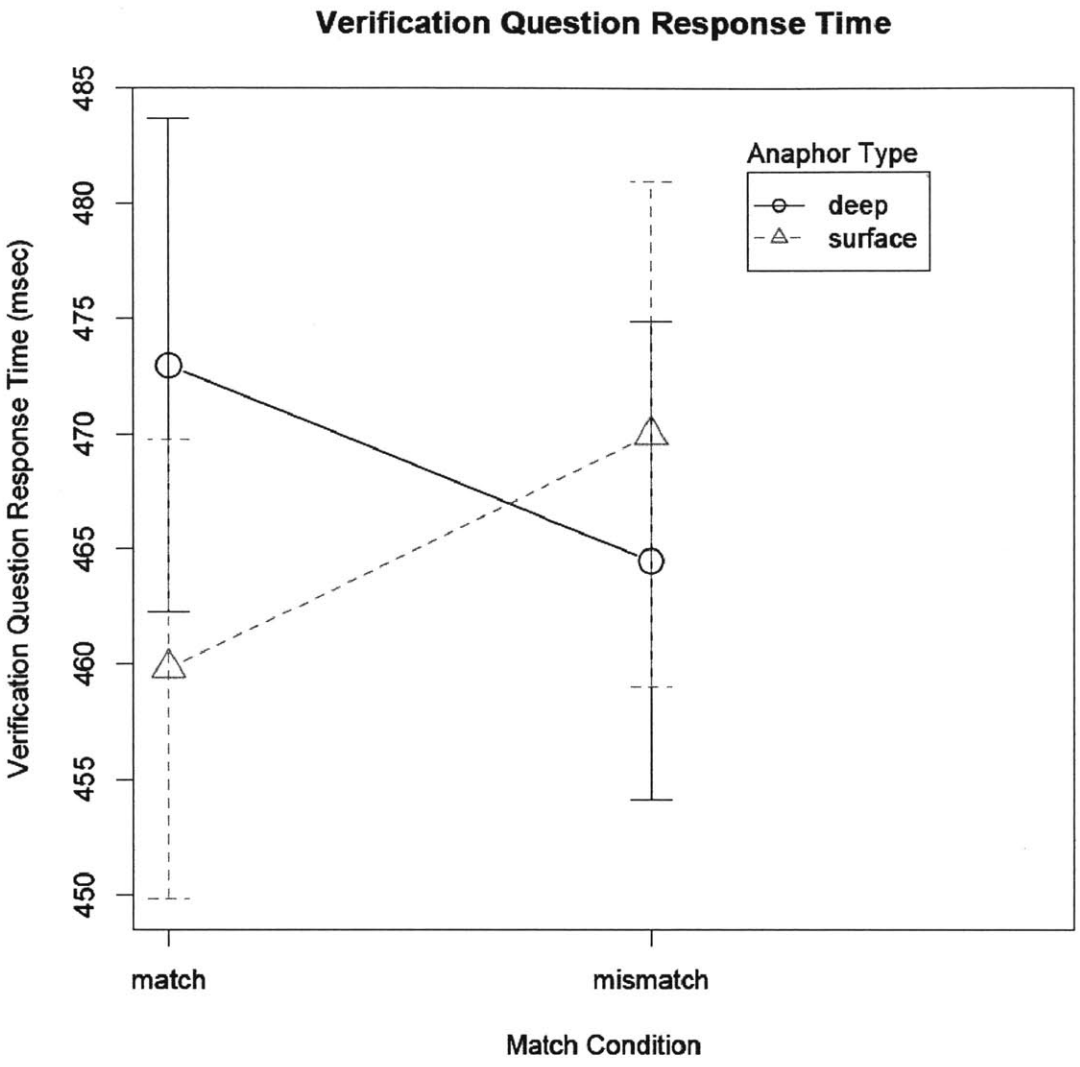

Figure 15: Verification question response times for the particle shift reading time experiment.

3.4.4. Discussion

In the region containing the anaphor, reading times were slower for surface anaphors than for deep anaphors. This result was also shown in the distance reading time results, and indicates an overall processing difference between surface and deep anaphors. The slower reading times for surface anaphors could reflect an additional level of processing involved in accessing the

surface anaphors' antecedents, such as the access of a syntactic level of information, before relating the meaning of the antecedent to the meaning of the anaphor.

In the critical region containing the verb+particle repetition, there was a main effect of match condition, such that items with both surface and deep anaphors had significantly slower RTs when the verb+particle construction did not match the initial verb+particle construction than when it did match. This main effect of match condition indicates that sentences with mismatching repeated verb+particle constructions are more difficult to process than sentences with matching repeated constructions, regardless of what kind of anaphor intervenes between the two verb+particle construction sentences. Sag and Hankamer's model would have predicted a significant interaction, in which mismatch would have affected items with surface anaphors more than items with deep anaphors. This critical interaction was not found to be significant. Therefore this finding does not support Sag and Hankamer's claim that surface anaphors reactivate surface syntactic form while deep anaphors do not, which would have resulted in a difference in reading time for surface conditions but not for deep conditions.

The diversion of our results from Sag and Hankamer's claims can be potentially explained by two different phenomena. First, it is possible that neither anaphor type reactivates information about exact word order, and mismatch effects arise because the word order of the antecedent is still salient at the point of the verification sentence regardless of the type of intervening sentence, resulting in a priming effect of the antecedent form on the verification sentence. To investigate this possibility, a follow-up experiment could to be conducted using longer anaphoric sentences or anaphoric sentences of varying lengths. A second possible explanation for these results is that the processing model is inaccurate and both anaphor types

(not just surface anaphors) reactivate the surface form of the first particle+verb construction, resulting in slow-downs in mismatch conditions for both anaphor types. A follow-up experiment using much longer anaphoric sentences or anaphoric sentences of varying lengths could help distinguish between these two possibilities: if longer anaphoric sentences resulted in no mismatch effects for either surface or deep conditions, the first phenomenon would be likely, whereas if longer anaphoric sentences still resulted in mismatch effects for both surface and deep conditions the second phenomenon would be likely.

3.5. Experiment 2b: Particle Shift fMRI study

3.5.1. Introduction

In parallel with the distance fMRI study, an fMRI study was conducted using the particle shift materials. This provided an opportunity to compare neural responses to deep and surface anaphora in a second set of materials. Since we saw significant differences in reading times for surface and deep anaphors in the particle shift reading time experiment, we expected to see these differences reflected in BOLD signal as well. Another goal of the particle shift fMRI experiment was to explore differences within surface and deep anaphors in response to mismatching versus matching verification sentences. In the particle shift reading time experiment, we saw a main effect of match condition, such that mismatch conditions were read significantly slower than match conditions for both surface and deep anaphors, but no effect of anaphor type and no interaction between match condition and anaphor type. This result alone indicates either that there is no difference between the processing of deep and surface anaphors, or that any difference in processing is obscured by the salience and priming of the antecedent sentence's

word order leading to a mismatch effect regardless of what kind of processing occurs at the anaphoric region. Because of the main effect of match type found in the reading time experiment, we expected to see corresponding BOLD effects in the contrast between mismatch and match.

Even though we did not see a difference between how surface and deep anaphors process mismatch conditions as compared to match conditions in the reading time experiments, it is possible that there is an underlying difference in how the brain processes these two anaphor types that is not reflected in reading times. If we were to compare mismatch>match contrasts between surface and deep anaphors and find different areas of activation for these contrasts in surface materials versus deep materials, this would be a strong indication that these two anaphor types are processed differently even though this difference was not reflected in the reading time data. If the mismatch>match contrast resulted in similar areas of activation for surface and deep anaphors, we could conclude that Sag and Hankamer's model is unsupported and that similar processing mechanisms are used for both surface and deep anaphors.

3.5.2. Methods

Participants 12 college-aged volunteers from the Boston University who had not participated in the reading time studies were recruited and paid for their participation. Subjects were right-handed native speakers of English from the USA and were naive to the purposes of the experiment. Informed consent was obtained prior to participation.

Materials 30 items of the format described in the Particle Shift Materials section were used in the fMRI experiment. Each subject saw all 6 variations of each item, or 180 target items total.

Variations of the items were distributed across runs so that no two variations from the same item were shown in the same run, in order to avoid familiarity effects. The target items were interspersed with 90 fillers of the format described in the Particle Shift Materials section. Complete lists of the target items and fillers used in the particle shift fMRI experiment can be found in Appendices C and D, respectively.

Procedures fMRI stimuli were presented using a rapid-presentation event-related (RPER) design. Each trial began with a 300 msec fixation block (a centered '+'), followed by a 100 msec blank screen block. The first sentence was presented following the fixation/blank block for 3 seconds, the second sentence then followed for 3 seconds, and the third (verification) sentence then appeared for another 3 seconds. During the presentation of the third/verification sentence, the subject's task was to make a response as to whether the third sentence was true or not by pressing a button with either the index or middle finger of the left hand. The presentation of the third sentence in this experiment was followed by a 600 msec blank screen block, so that each trial was 10 seconds long in total.

A variable fixation block was then presented, with a duration of 0, 2, 4, or 6 seconds. The length of the variable fixation trial and the trial type presentation order were determined using optseq2, a program developed to optimize the estimation of the hemodynamic response and reduce subject habituation and expectation (Burock et al, 1998; Dale, 1999; Dale and Buckner, 1997).

Including variable fixation trials, the functional part of the particle shift experiment lasted approximately 50 minutes. The particle shift experiment was divided into 6 runs of 8.33 minutes (500 seconds, or 250 time points) each. Subjects were allowed a short rest period between runs.

Visual stimuli were projected onto a screen at the back of the scanner using an LCD projector, and they were viewed by subjects using a mirror attached to the head coil. Subjects recorded their responses using an MRI-compatible button box held in the left hand, with a button under the middle finger indicating a true verification sentence and a button under the index finger indicating an false verification sentence. An Apple PowerBook G4 computer running PsyScope software (Cohen et al, 1993) was used to present the stimuli and to record responses and reaction times.

MR Imaging Parameters All images were acquired using a 3.0 T whole-body Siemens Trio scanner (Siemens Medical Solutions USA Inc., Malvern, PA, USA). Following a brief localizer and auto-align scan paradigm, one high-resolution anatomical image was acquired using a T1-MPRAGE sequence (TR=2530 msec, TE=3.39 msec, flip angle = 7°). The anatomical image volume consisted of 128 sagittal slices with an effective thickness of 1.33 mm and an in-plane resolution of 1.0 mm X 1.0 mm (256 X 256 matrix size, 256 mm FOV).

Functional volumes were acquired using a T2*-weighted gradient-echo pulse sequence that is sensitive to blood oxygenation level-dependent (BOLD) signal (TR = 2 sec, TE = 30 msec, flip angle = 90°). Functional volumes consisted of 30 transverse slices aligned along the anterior/posterior commissure (AC-PC) plane as determined by a registration volume. The transverse slices had an effective thickness of 3.0 mm with a gap of 0.9 mm between slices. The in-plane resolution was 3.125 mm X 3.125 mm (64 X 64 matrix, 200 mm FOV). By definition, the 30 slices in each volume were acquired in one TR (2 sec), with a new volume collected in each TR. Each run in the particle shift experiment consisted of 250 volumes, resulting in 7500 images collected per run. 8 seconds (4 TRs) of RF pulse activations were presented prior to each

run during which no visual stimuli were presented and no images were acquired, in order to maximize signal strength during the functional run.

fMRI Data Analysis In the particle shift experiment, the BOLD responses to each sentence (antecedent, anaphoric, or verification sentences) were examined individually. The first contrast of interest was to compare sentences containing deep anaphors to sentences containing surface anaphors (the second sentence in the set of three). The second contrast of interest was to compare responses to matching versus mismatching trial types within deep and surface anaphor trial types (the third sentence in the set of three). Within deep anaphor trials, we expected that there would be no difference between matching and non-matching third sentences if the processing model is accurate and surface information is not salient after processing a deep anaphor. Within surface anaphor trials, we expected that there would be an area or areas where more activation was shown in mismatching trials than in matching trials, and that this area would be similar or identical to areas shown to exhibit more activity in response to surface anaphors than deep anaphors. This area (or these areas) could reflect a location where surface information is processed, as the processing model predicts that surface information will be salient after the surface anaphor, making it more difficult or requiring more effort to process the shifted verb +particle construction after the surface anaphor.

All preprocessing and statistical analysis of the fMRI data was performed using the SPM8 software package. Preprocessing consisted of slice time correction, motion correction, intensity normalization, registration and resampling in standard MNI space, and spatial smoothing with a Gaussian kernel of 6 mm full width half maximum (FWHM). The functional images were then analyzed using a random effects model of an event-related paradigm, with the

hemodynamic response function (HRF) modeled as a Gamma function. Functional images were high-pass filtered at a cutoff of 128 s to remove scanner drift effects. Outlier time points in the fMRI time series due to head motion or spikes in global mean intensity were then removed using the Artifact Detection Toolbox (ART) for SPM8[4] (Mazaika et al 2007, Mazaika et al 2009, Mazaika et al 2005). For the group analysis, paired t-tests were performed on individual contrast images to determine whether individual voxel differences were reliably different from zero. The resulting statistical parametric maps were FDR-corrected for multiple comparisons at $p < 0.05$ with an additional cluster size threshold of 50 voxels. Local maxima for significant clusters and the brain regions encompassed by each cluster were determined using the SPM8 toolboxes aal (Tzourio-Mazoyer 2002) and xjView[5].

3.5.3. Results

Data for 12 subjects were included in the final analysis. 11 of these subjects also completed the distance fMRI experiment. One subject was replaced due to excessive head motion and questionable handedness. All subjects had accuracies higher than 89%, with an overall mean accuracy of 94.3%. Mean accuracy for target items alone was 99.1%.

Behavioral Results As in the reading time experiments, all statistical tests were carried out with mixed linear models (lmer function in R), using subjects and items as random effects and experimental conditions of interest as fixed effects (e.g. Van Dongen et al 2004, Jaeger

---

[4] developed by Shay Mozes and Susan Whitfield-Gabrieli - http://web.mit.edu/swg/software.htm

[5] developed by Cui, Li, and Song - http://www.alivelearn.net/xjview8/

2008). 2x2 (mismatch/match verb+particle constructions, surface/deep anaphors) analyses were conducted on the verification sentence response times and accuracies. If the interaction was found to be significant, planned comparisons between items with mismatching and matching verb+particle constructions were conducted on the verification sentence response times and accuracies for both surface and deep anaphors. Verification sentence response times are shown below in Figure 16 and verification sentence accuracies are shown in Figure 17.

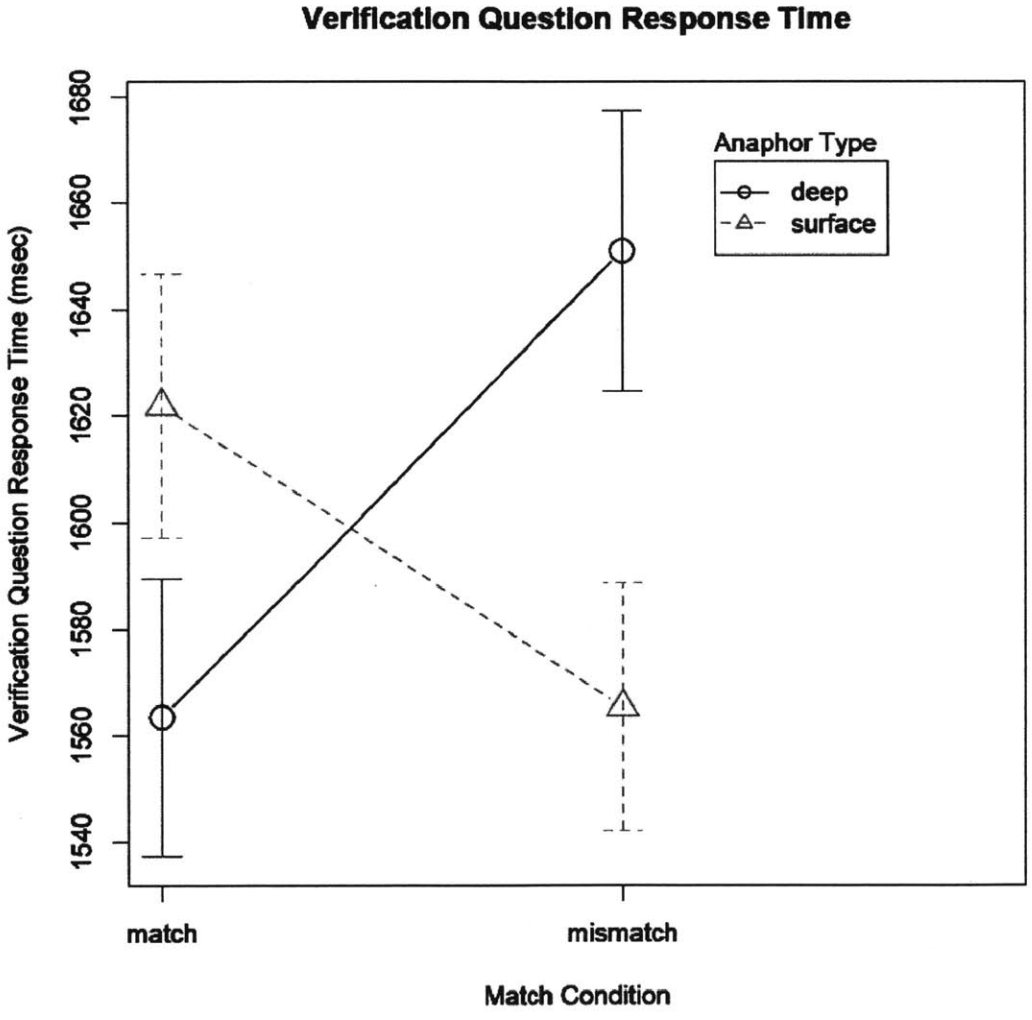

Figure 16: Verification question response times for the particle shift fMRI experiment.

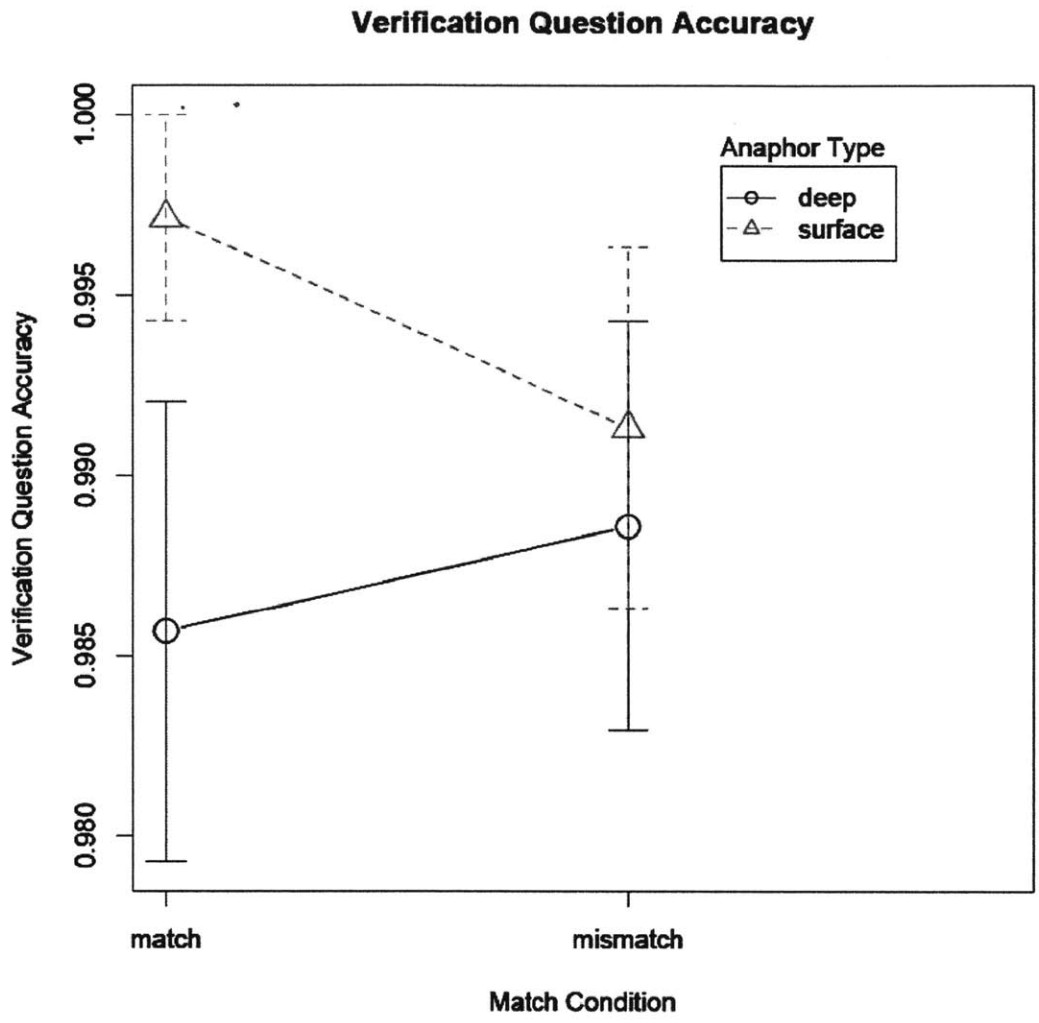

Figure 17: Verification question accuracies for the particle shift fMRI experiment.

For the verification sentence response times, the main effects of match and anaphor type were both insignificant (both $p > 0.05$). The interaction between match and anaphor type was significant ($F(1,1391)=5.09$, $p < 0.024$), but comparisons between match and mismatch conditions within anaphor type showed that response times did not differ significantly between mismatch and match conditions for surface anaphors ($F(1,693)=0.492$, $p=0.48$) or for deep anaphors ($F(1,698)=3.22$, $p=0.073$). Instead, the interaction arose because the difference

between surface mismatch and deep mismatch was significant ($F(1,695)=7.02$, $p < 0.01$), with no other differences reaching significance.

For the verification sentence accuracy, the main effects of match and anaphor type and the interaction between match and anaphor were not significant (all $p > 0.1$).

In the fMRI study, reading times were not acquired for individual regions of the sentences, so it was not possible to compare reading times in the VP repetition region to the results obtained in the reading time study.

fMRI Results No clusters passed the threshold for significance using FDR correction at $p < 0.05$ and a cluster size of 50 voxels for the contrasts of Surface Mismatch > Surface Match, Deep Mismatch > Deep Match, Surface > Deep, or Deep > Surface.

3.5.4. Discussion

As there was no significant difference between BOLD measures when comparing mismatch and match conditions or deep and surface conditions, it is difficult to draw any conclusions about this experiment. Because the reading time study showed significant differences between mismatch and match conditions for both surface and deep anaphors as well as significant differences in reading times between surface and deep anaphors, we expected to see significant differences between mismatch and match conditions for surface and deep anaphors and between surface and deep anaphors in the fMRI study. As significant BOLD results were seen at an uncorrected level of $p < 0.001$ and cluster size > 25 voxels for the contrast of deep mismatch > deep match, it is possible that there were simply not enough subjects to

obtain similar results at an FDR-corrected level of $p < 0.05$. However, since the same group of subjects demonstrated significant FDR-corrected BOLD signal differences in the distance experiment, it is also possible that the null results demonstrate that there is actually no discernible difference between how the brain processes mismatch and match conditions.

A third possibility for the existence of the null results is that subjects developed strategies which allowed them to complete the task without fully processing the materials. As the accuracy for the target items was nearly perfect (99.1%) while the accuracy for non-target items was significantly lower (90.4%), it is possible that subjects developed unnatural strategies which allowed them to recognize and answer the target items correctly without fully processing them, which may have eliminated any differences in processing that would have been evident in BOLD and behavioral results. The disparity between the performance on target items and non-target items may be explained by the fact that the ratio of items (targets and fillers) with a correct answer of "true" to items with a correct answer of "false" was 210:60 (180 target items with correct answer of "true", 30 filler items with correct answer of "true", 60 filler items with correct answer of "false"). Additionally, all verification sentences in the 180 target items referred to the antecedent sentence, had an answer of "true", and were either identical to the antecedent sentence or differed only slightly, while only 30 fillers referred to the first sentence and had an answer of "false", and many of these fillers were reworded from the antecedent significantly to make the verification answer "false". Subjects may therefore have developed a strategy of responding "true" when the verification sentence referred to or closely resembled the antecedent sentence, without having to thoroughly read and process the verification sentence.

Another reason that subjects may have been able to complete the task without thoroughly processing the materials is that subjects saw all 6 versions of 30 items in the fMRI experiment. Since the target items were repeated several times (with only differences in the anaphoric sentence, which was not important in responding to the verification sentence), subjects may have learned to recognize the target items. Since the answer to all target verification questions was "true", subjects may have learned that if they recognized an item, the answer to the verification sentence was always "true". In this case, they would only have to fully process the filler items in order to do well on the task, which they would not be familiar with, as there was only one version of each filler item.

4. Overall Discussion of Results

The reading time results from the distance and particle shift experiments showed that there is a reading time difference between surface and deep anaphora. In both experiments, surface anaphors were read more slowly than deep anaphors. This could indicate an additional processing step in relating surface anaphors to their antecedents, such as contacting a syntactic representation of the antecedent.

The results from the distance materials are consistent with the prediction that increased distance between the antecedent and the anaphor adversely affects surface anaphors more than deep anaphors. Increased distance decreased the acceptability of the surface anaphors more than it decreased the acceptability of deep anaphors. Increased distance resulted in increased comprehension question response times for surface anaphors but had no effect on deep anaphors. Finally, increased distance resulted in increased BOLD signal in several language-related areas of the brain for surface anaphors but had no effect on deep anaphor BOLD signal. These results indicate that surface anaphors are adversely affected by distance, which may be due to a decay in the representation of the antecedent during the intervening sentence. The information that decays over distance has previously been shown to be related to syntax and/or exact word order information (see Chapter 1: Introduction), indicating that the surface anaphor may contact syntactic and/or word order information. Deep anaphors were not affected by distance, indicating that the information they access is not degraded with distance, such as non-linguistic discourse or semantic information. Sachs 1974 found that the semantic content of a sentence was still significantly salient even after the maximum tested length of intervening material (80

syllables/16.5 seconds), suggesting that semantic information can easily be maintained over short intervening sentences such as those used in the distance experiments here.

The results from the particle shift materials are not consistent with the prediction that changing word order has greater effects on items with surface anaphors than deep anaphors. In the reading time experiment, the reading time at the verb+particle repeat region increased in the mismatch condition for surface anaphors and deep anaphors, and no effects of match condition were seen in the verification question response time or accuracy for either surface or deep anaphors. In the fMRI experiment, neither items with surface or deep anaphors differed in BOLD signal, response time, or accuracy between match and mismatch conditions. Two possible explanations follow from this pattern of results. First, it is possible that both surface and deep anaphors access information about the exact order of the particle+verb construction, so this information is reactivated in both cases and results in the mismatch effect seen for both surface and deep conditions. As deep anaphors were not affected by distance in the distance experiments, however, it seems unlikely that deep anaphors rely on exact word order information. A second possible explanation is that neither surface nor deep anaphors access information about the exact order of the particle+verb construction. For example, surface anaphors could be accessing the antecedent at a level where some information about syntax, but not the exact surface word order, is important, while deep anaphors could be accessing the antecedent at a level where neither syntactic nor surface word order is important.

In the particle shift experiment, if both deep and surface anaphors access the antecedent at a level of representation that is not sensitive to the exact order of the particle+verb construction, it would be useful to know the level of processing at which this representation is

thought to occur. Particle-verb constructions with shifted and unshifted particles differ in certain respects. For instance, in English, only shifted constructions allow object pronouns (*John picked up it) and control of an embedded infinitive (*We lifted up $Gary_i$ PRO_i to see over the wall). It has been proposed that particle+verb constructions where the particle is and is not shifted are represented differently at a phonological or prosodic level (Kremers 2009), and at the level of syntactic structure and logical form (Svenonius 1994, Svenonius 1996, Elenbaas 2007). It is possible that verb phrase ellipses are interpreted by a process that utilizes one or more of these levels of representation.

The results from these experiments could also be consistent with a processing model in which the syntactic form of the antecedent is not copied into the location of the anaphor, but is instead accessed using a pointer or index to the antecedent's preexisting representation in memory. The idea of anaphoric resolution via a pointer mechanism rather than a copying mechanism has been proposed and supported by several previous studies. Martin and McElree (2008 and 2009) used verb-phrase ellipsis with antecedents of varying length and complexity and argued that their finding of lower accuracy and unchanged speed-accuracy dynamics with increasingly complex or long antecedents is consistent with a system in which antecedents are accessed by anaphors with a pointer mechanism rather than a copying mechanism. They argue that a copying mechanism would be slower when copying more complex or lengthy antecedents into the anaphoric site, whereas the time to process a pointer mechanism from the anaphoric site to the preexisting antecedent information would not be affected by the length or complexity of the antecedent. Lower accuracy was associated with longer and more complex antecedents, which Martin and McElree attributed to reduced quality of the retrieved antecedent. Frazier and

Clifton (2001) proposed a similar idea in which ellipses access the syntactic information contained in their antecedents through a "cost-free copy-α" operation (where α refers to the antecedent), which is implemented by a pointer from the ellipsis site to the left-most corner of the antecedent's syntactic representation in memory. Frazier and Clifton (2005) suggested a "structure-sharing" mechanism for ellipsis, in which the site of the ellipsis shares the structure of the antecedent, a process which could be carried out using a pointer mechanism.

The results of the distance reading time experiment are consistent with Martin and McElree (2008), as increasing the distance of the antecedent from the surface anaphor did not result in slower RTs at the anaphoric region, but did result in degraded comprehension question performance. This suggests the presence of a pointer to the antecedent representation, where increased distance reduces the quality of the antecedent's representation and affects comprehension, rather than a copying mechanism of the antecedent information onto the anaphoric site which would affect reading time. The particle shift reading time results could also be explained by the presence of a pointer mechanism rather than a copying mechanism. Our results were not consistent with the antecedent word order information being copied into the surface anaphoric region or made more salient at that point. If the antecedent's word order information was instead accessed by a pointer to the antecedent's representation in memory, the anaphor could have accessed this information without actually copying it into the anaphoric region.

Both deep and surface anaphors could potentially access their antecedents using a pointer mechanism, but with the pointer directed towards different levels of information for surface or deep anaphors, such as semantic or discourse information for deep anaphors and syntactic

information for surface anaphors. Martin and McElree (2008) claim that their proposed pointer mechanism for anaphoric resolution is equally compatible with accounts in which anaphoric resolution depends on syntactic information (such as Frazier and Clifton 2001) and with accounts in which anaphoric resolution depends on discourse information (such as Garnham 2001), the only difference being that the pointer would point either to a syntactic representation or a discourse representation.

To summarize, our results are consistent with a processing model in which surface anaphors access a level of representation that decays with increased distance between the antecedent and anaphor, such as syntactic information, while deep anaphors access a level of representation that remains salient with increased distance, such as semantic or discourse information. Surface and deep anaphors did not differ in their sensitivity to exact word order in the particle shift experiments. Additionally, surface and deep anaphors may contact their antecedents by using a pointer to the antecedent's representation in memory, not by copying the antecedent information into the anaphoric site.

5. Conclusions

In experiments examining the effects of increasing the distance between the antecedent and the anaphor, surface anaphors were affected by increased distance more than deep anaphors in terms of naturalness ratings, comprehension question response time, and BOLD signal activations. In experiments examining the effects of word order, mismatching word order affected items with surface and deep anaphors equally in terms of reading time, and neither surface nor deep anaphors were affected in terms of BOLD signal activations. Additionally, overall differences in reading times were seen between surface and deep anaphors in distance and word order experiments.

These results suggest that surface and deep anaphora are processed by accessing their antecedents in different ways and at different levels. The results obtained here are consistent with a model in which surface anaphors are resolved by using a pointer which points to the antecedent's preexisting structure in memory, rather than copying this structure into the anaphoric site. The antecedent form which is accessed by the pointer includes information which is sensitive to decay over intervening material, such as syntactic information. Deep anaphora were affected less by the introduction of an intervening sentence, indicating that they access their antecedents at a level that is less subject to decay over intervening material, such as a non-linguistic discourse-related level. It is possible that deep anaphora also access their antecedents using a pointer to the antecedent's representation, except this representation includes discourse or semantic information instead of syntactic information.

Imaging results supported these claims for surface anaphora, as the distance contrast for surface anaphora resulted in areas of activation that have previously been related to phonological

and syntactic functions. No imaging results were significant for deep anaphora or for the comparison of deep to surface anaphora in either the distance or particle shift experiments, making it impossible to draw conclusions about how deep anaphors are processed in the brain or about any differences between surface and deep anaphora in the brain.

In summary, these studies provide evidence that surface and deep anaphora are processed in different ways, with surface anaphora depending more on syntactic-level information and deep anaphora depending more on non-linguistic discourse-related information. These findings therefore support Hankamer and Sag's original observation that deep and surface anaphora are distinct categories of anaphora, which are processed in different ways and at different levels of representation. However, these results are inconsistent with the exact level of processing proposed by Sag and Hankamer, in that Sag and Hankamer proposed that surface anaphors are processed by accessing the surface syntactic structure of the antecedent, while our results showed that surface anaphors are more dependent on information that decays over distance (such as syntactic information) than deep anaphors, but do not differ from deep anaphors in terms of accessing exact surface word information.

6. References

Almor A. (1999). Noun-phrase anaphora and focus: The information load hypothesis. Psychological Review 105: 748-765.

Almor A, Smith DV, Bonilha L, Fridriksson J, and Rorden C. (2007). What is in a name? Spatial brain circuits are used to track discourse references. NeuroReport 18: 1215-1219.

Ariel M. (1990). Accessing noun-phrase antecedents. London, UK: Routledge.

Arregui A, Clifton C, Frazier L and Moulton K. (2006). Processing elided VPs with flawed antecedents. Journal of Memory and Language 55: 232-246.

Awh E, Jonides J, Smith E, Schumacher E, Koeppe R, and Katz S. (1996). Dissociation of storage and rehearsal in working memory: PET evidence. Psychological Science 7: 25-31.

Belanger S. (2004). Processing of deep and surface anaphors: a replication and extension of Tanenhaus, Carlson & Seidenberg (1985). Toronto Working Papers in Linguistics 22: 1-40.

Bookheimer S. Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. Annual Review of Neuroscience 25: 151-188.

Buckner R and Koutstaal W. Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. Proceedings of the National Academy of Sciences 95: 891-898.

Burock MA, Buckner RL, Woldorff MG, Rosen BR and Dale AM. (1998). Randomized event- related experimental designs allow for extremely rapid presentation rates using functional MRI. Neuroreport, 9, 3735-3739.

Chen S and Desmond J. (2005). Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks. NeuroImage 24: 332-338.

Clifton C and Frazier L. (2010). Imperfect ellipsis: Antecedents beyond syntax? Syntax 13: 279-297.

Cloitrew M and Bever T. (1988). Linguistic anaphors, levels of representation, and discourse. Language and Cognitive Processes 3: 293-322.

Cohen JD, MacWhinney B, Flatt M, and Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. Behavioral Research Methods, Instruments, and Computers, 25, 257-271.

Dale AM. (1999). Optimal experimental design for event-related fMRI. Human Brain Mapping, 8, 109-114.

Dale AM and Buckner RL. (1997). Selective averaging of rapidly presented individual trials using fMRI. Human Brain Mapping, 5, 329-340.

Dalrymple M, Sheiber S, and Pereira FCN. (1991). Ellipsis and higher-order unification. Linguistics and Philosophy 14: 399-452.

Dapretto M and Bookheimer S. (1999). Form and Content. Neuron 24: 427-432.

Dell G, McKoon G, and Ratcliff R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. Journal of Verbal Learning and Verbal Behavior 22: 121-132.

Dhanjal N, Handunnetthi L, Patel M, and Wise R. (2008). Perceptual systems controlling speech production. Journal of Neuroscience 28: 9969-9975.

Elenbaas M. (2007). The verb-particle combination in Present-Day English. In The Sychronic and Diachronic Syntax of the English Verb-Particle Combination. Doctoral Thesis, Netherlands Graduate School of Linguistics, #40178.

Ferreira F and Clifton C. (1986). The independence of syntactic processing. Journal of Memory and Language, 25, 348-368.

Ferstl E and von Cramon D. (2001). The role of coherence and cohesion in text comprehension: An event-related fMRI study. Cognitive Brain Research 11: 325-340.

Fiengo R. (1974). Semantic conditions on surface structure. PhD thesis, MIT, Cambridge, MA.

Fiengo R and May R. (1994). Indices and identity. Cambridge, MA: MIT Press.

Fiez J and Petersen S. (1998). Neuroimaging studies of word reading. Proceedings of the National Academy of Sciences 95: 914-921.

Frazier L and Clifton C. (2001). Parsing coordinates and ellipsis: Copy α. Syntax 4: 1-22.

Frazier L and Clifton C. (2005). The syntax-discourse divide: Processing ellipsis. Syntax 8: 121-174.

Frazier L and Clifton C. (2006). Ellipsis and discourse coherence. Linguistics and Philosophy 29: 315-346.

Friederici A. (2002). Towards a neural basis of auditory sentence processing. Trends in Cognitive Sciences 6: 78-84.

Friederici A, Opitz B, Cramon D. (2000). Segregating semantic and syntactic aspects of processing in the human brain: an fMRI investigation of different word types. Cerebral Cortex 10: 698-705.

Garnham A. (2001). Mental models and the interpretation of anaphora. Hove, East Sussex, UK: Psychology Press.

Garnham A, Oakhill J, and Cain K. (1997). The interpretation of anaphoric noun phrases: Time course, and effects of overspecificity. The Quarterly Journal of Experimental Psychology 50A: 149-162.

Garrod S and Sanford A. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding. In M. A. Gernsbacher (Ed.), Handbook of psycholinguistics, pp675-698. San Diego, CA: Academic Press.

Gernsbacher M. (1989). Mechanisms that improve referential access. Cognition 32: 99-156.

Gibson E and Federenko E. (2010). Weak quantitative standards in linguistics research. Trends in Cognitive Sciences 14: 233-234.

Gibson E and Federenko E. (In Press). The need for quantitative methods in syntax and semantics research. Language and Cognitive Processes.

Gordon P, Grosz B, and Gilliom L. (1993). Pronouns, names, and the centering of attention in discourse. Cognitive Science 17: 311-347.

Gordon P and Chan D. (1995). Pronouns, passives, and discourse coherence. Journal of Memory and Language 34: 216-231.

Gordon P, Scearce K. (1995). Pronominalization and discourse coherence, discourse structure and pronoun interpretation. Memory and Cognition 23: 313-323.

Grinder J and Postal P. (1971). Missing antecedents. Linguist Inquiry 2: 269-312.

Grosz B, Joshi A, and Weinstein S. (1983). Providing a unified account of definite noun phrases in discourse. In Proc. 21st Annual Meeting of the ACL, pp 44-50, Association of Computational Linguistics.

Gruber O. (2001). Effects of domain-specific interference on brain activation associated with verbal working memory task performance. Cerebral Cortex 11: 1047-1055.

Guenther F. (2006). Cortical interactions underlying the production of speech sounds. Journal of Communication Disorders 39: 350-365.

Hagoort P. (2005). On Broca, brain, and binding: A new framework. Trends in Cognitive Sciences 9: 416-423.

Hammer A, Jansma B, Lamers M, and Munte T. (2008). Interplay of meaning, syntax, and working memory during pronoun resolution investigated by ERPs. Brain Research 1230: 177-191.

Hankamer J and Sag T. (1976). Deep and surface anaphors. Linguistic Inquiry 7: 391-426.

Hardt D. (1992). VP ellipsis and contextual interpretation. In Proceedings of the International Conference on Computational Linguistics (COLING-92): 303-309.

Hardt D. (1993). Verb phrase ellipsis: Form, meaning, and processing. PhD Thesis, University of Pennsylvania.

Hardt D and Romero M. (2004). Ellipsis and discourse structure. Journal of Semantics, 21: 375-414.

Hestvik A, Nordby H, and Karlsen G. (2005). Antecedent reactivation by surface and deep anaphora in Norwegian. Scandinavian Journal of Psychology 46: 229-238.

Jaeger TF. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language 59: 434-446.

Jackendoff RS. (1972). Semantic interpretation in generative grammar. MIT Press, Cambridge, MA.

Just M, Carpenter P, Keller T, Eddy W, and Thulborn K. Brain activation modulated by sentence comprehension. Science 274: 114-116.

Kaan E and Swaab T. The brain circuitry of syntactic comprehension. Trends in Cognitive Sciences 6: 350-356.

Kaiser E, Runner J, Sussman R, and Tanenhaus M. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. Cognition 112: 55-80

Kehler A. (1993). A discourse copying algorithm for ellipsis and anaphora resolution. Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics . Utrecht, the Netherlands, 203-212.

Kehler A. (2000). Coherence and the resolution of ellipsis. Linguistics and Philosophy 23: 533-575.

Keller T, Carpenter P, and Just M. The neural bases of sentence comprehension: A fMRI examination of syntactic and lexical processing. Cerebral Cortex 11: 223-237.

Kertz L. (2008). Focus structure and acceptability in verb phrase ellipsis. Proceedings of the 27th West Coast Conference on Formal Linguistics. 283-291.

Koelsch S, Schulze K, Sammler D, Fritz T, Muller K, and Gruber O. (2009). Functional architecture of verbal and tonal working memory: An fMRI study. Human Brain Mapping 30: 859-873.

Kremers J. (2009). A prosodic approach to particle verbs. Linguistische Berichte 2009.

Kuperberg G, Lakshmanan B, Caplan D, and Holcomb P. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. NeuroImage 33: 343-361.

Lasnik H. (2001). When can you save a structure by destroying it? Proceedings of NELS 31: 301-320.

Martin A and McElree B. (2008). A content-addressable pointer mechanism underlies comprehension of verb phrase ellipsis. Journal of Memory and Language 58: 879-906.

Martin A and McElree B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. Journal of Experimental Psychology 35: 1231-1239.

Mauner G, Tanenhaus M and Calrson G. (1995). A note on parallelism effects in processing deep and surface verb-phrase anaphora. Language and Cognitive Processes 10: 1-12.

Mazaika P, Whitfield S, and Cooper J. (2005). Detection and repair of transient artifacts in fMRI data. Paper presented at the Annual Meeting of the Organization for Human Brain Mapping.

Mazaika P, Whitfield-Gabrieli S, and Reiss A. (2007). Artifact repair for fMRI data from high motion clinical subjects. Paper presented at the Annual Meeting of the Organization for Human Brain Mapping.

Mazaika P, Hoeft F, Glover G, and Reiss A. (2009). Methods and software for fMRI analysis for clinical subjects. Paper presented at the Annual Meeting of the Organization for Human Brain Mapping.

McGettigan C, Warren J, Eisner F, Marshall C, Shanmugalingam P, and Scott S. (2011). Neural correlates of sublexical processing in phonological working memory. Journal of Cognitive Neuroscience 23: 961-977.

McKoon G and Ratcliff R. (1980). The comprehension processes and memory structures involved in anaphoric reference. Journal of Verbal Learning and Verbal Behavior 19: 668-682.

Munro R, Bethard S, Kuperman V, Lai VT, Melnick R, Potts C, Schnoebelen T, and Tily H. (2010). Crowdsourcing and language studies: the new generation of linguistic data. Proceedings of NAACL 2010.

Murphy G. (1985). Processes of understanding anaphora. Journal of Memory and Language 24: 290-303.

Murphy G. (1990). Interpretation of verb phrase anaphora: influences of task and syntactic context. The Quarterly Journal of Experimental Psychology 424: 675-692.

Myers J and O'Brien E. (1998). Accessing the discourse representation during reading. Discourse Processes 26: 131-157.

Narayanan N, Prabhakaran V, Bunge S, Christoff K, Fine E, and Gabrieli J. (2005). The role of the prefrontal cortex in the maintenance of verbal working memory: An event-related fMRI analysis. Neuropsychology 19: 223-232.

Newman S, Just M, Keller T, and Carpenter P. (2003). Differential effects of syntactic and semantic processing on the subregions of Broca's area. Cognitive Brain Research 16: 297-307.

Nicol J and Swinney D. (1989). The role of structure in coreference assignment during sentence comprehension. Journal of Psycholinguistic Research 18: 5-19.

Nieuwland MS, Petersson KM, and Van Berkum JJA. (2007). On sense and reference: Examining the functional neuroanatomy of referential processing. NeuroImage 37: 993-1004.

Nieuwland M and Van Berkum J. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. Brain and language 106: 119-131.

O'Brien E, Raney G, Albrecht J, and Rayner K. (1997). Processes involved in the resolution of explicit anaphors. Discourse Processes 23: 1-24.

Paulesu E, Frith C, and Frackowisk R. (1993). The neural correlates of the verbal component of working memory. Nature 362: 342-345.

Poirier J, Shapiro LP, Love T, and Grodzinsky Y. (2009). The on-line processing of verb-phrase ellipsis in aphasia. Journal of Psycholinguistic Research 38: 237-253.

Poppenk J, Walia G, McIntosh AR, Joanisse MF, Klein D, and Kohler S. (2008). Why is the meaning of a sentence better remembered than its form? An fMRI study on the role of novelty-encoding processes. Hippocampus 18: 909-918.

Postal P. (1970). On coreferential complement subject deletion. Linguistic Inquiry 1: 439-500.

Postal P. (1972). Some further limitations of interpretive theories of anaphora. Linguistic Inquiry 3: 349-371.

Riecker A, Mathiak K, Wildgruber D, Erb M, Hertrich I, Grodd W, and Ackermann H. (2005). Neurology 64: 700-706.

Rohde D. (2003). Linger, version 2.94. *http://tedlab.mit.edu/~dr/linger/.*

Ross JR. (1967). Constraints on variables in syntax. PhD Thesis, MIT, Cambridge, MA.

Ross JR. (1969). On the cyclic nature of pronominalization. In D. Reibel and S. Schane, eds., Modern Studies in English, Prentice-Hall, Englewood Cliffs, New Jersey.

Sachs J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics 2: 437-442.

Sachs J. (1974). Memory in reading and listening to discourse. Memory & Cognition 2: 95-100.

Sag T and Hankamer J. (1984). Toward a theory of anaphoric processing. Linguistics & Philosophy 7: 325-345.

Sanford A and Garrod S. (1989). What, when, and how? Questions of immediacy in anaphoric reference resolution. Language and Cognitive Processes 4: SI235-SI262.

Santi A and Grodzinky Y. (2007a). Taxing working memory with syntax: Bihemispheric modulations. Human Brain Mapping 28: 1089-1097.

Santi A and Grodzinsky Y. (2007b). Working memory and syntax interact in Broca's area. NeuroImage 37: 8-17.

Shopen T. (1972). A generative theory of ellipsis. PhD Thesis, UCLA, Los Angeles, CA.

Smith E and Jonides J. (1998). Neuroimaging analyses of human working memory. Proceedings of the National Academy of Sciences 95: 12061-12068.

Smith E, Jonides J, Marshuetz C, and Koeppe R. (1998). Components of verbal working memory: Evidence from neuroimaging. Proceedings of the National Academy of Sciences 95: 876-882.

Stowe L, Haverkort M, and Zwarts F. (2005). Rethinking the neurological basis of language. Lingua 115: 997-1042.

Svenonius P. (1994). Verb-Particle Constructions. In Dependent Nexus: Subordinate Predication Structures in English and the Scandinavian Languages. PhD Dissertation, University of California Santa Cruz.

Svenonius P. (1996). The optionality of particle shift. Working Papers in Scandinavian Syntax 57: 47-75.

Tanenhaus M, Carlson G and Seidenberg M. (1985). Do listeners compute linguistic representations? In DR Dowty, L Karttunen and AM Zwicky (eds.), Natural language parsing: theoretical, psychological and computational perspectives. Cambridge, Cambridge University Press.

Tanenhaus M and Carlson G. (1990). Comprehension of deep and surface verb phrase anaphors. Language and Cognitive Processes 5: 257-280.

Tartter V. (1998). Language and its normal processing. Thousand Oaks, CA: Sage Publications.

Trueswell JC, Tanenhaus MK, and Garnsey SM. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. Journal of Memory and Language, 33, 285-318.

Turkeltaub P, Eden G, Jones K, and Zeffiro T. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. NeuroImage 16: 765-680.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, and Joliot M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15: 273-289.

Van Dongen H, Olofsen E, Dinges D, and Maislin G. (2004). Mixed-Model Regression Analysis and Dealing with Interindividual Differences. In: Johnson ML, Brand L, eds. Numerical Computer Methods (part E). San Diego: Academic Press:139-171.

Vigneau M, Beaucousin V, Herve PY, Duffau H, Crivello F, Houde O, Mazoyer B, and Tzourio-Mayozer N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. NeuroImage 30: 1414-1432.

Wager T and Smith E. (2003). Neuroimaging studies of working memory: A meta-analysis. Cognitive, Affective, and Behavioral Neuroscience 3: 255-274.

Wasow T. (1975). Anaphoric pronouns and bound variables. Language 51: 368-383.

Woodward T, Cairo T, Ruff C, Takane Y, Hunter M, and Ngan E. (2006). Functional connectivity reveals load dependent neural systems underlying encoding and maintenance in verbal working memory. Neuroscience 139: 317-325.

7. Appendices

Appendix A: Materials for distance experiments. All materials were used in the naturalness rating and reading time experiments. Items in bold type and all fillers were used in the fMRI experiment. Y/N after the item indicates the correct answer to the comprehension question.

1	elaborative	John raked the leaves in the back yard. (The oak and maple leaves covered the ground.) Later, Bill (did/did it) too. Did John rake the leaves? Y
2	cause-effect	Lauren got a flu shot last Monday. (She didn't want to get sick and miss school.) On Tuesday, Katie (did/did it) too. Did Lauren get a shot on Monday? Y
3	temporal	Ben practiced the piano all day. (After that, the power went out.) That evening, Julie (did/did it) too. Did Ben practice the piano? Y
4	elaborative	Nancy read a book for fun. (The book was about Islam.) That month, Mark (did/did it) too. Did Nancy read for fun? Y
5	cause-effect	Sean bought a new car on Labor Day. (His old car was totalled in an accident.) The next day, Peter (did/did it) too. Did Sean buy a car on Labor Day? Y
6	**temporal**	**Sam brought cupcakes to the party.** **(Then they fell on the floor.)** **Luckily, Ken (did/did it) too.** **Did Ken bring cupcakes too? Y**
7	elaborative	Molly wrote a letter to the senator. (It was twelve pages long.) The next day, Nicole (did/did it) too. Did Nicole write a letter to the senator? Y
8	**cause-effect**	**Terry met with a tutor after school.** **(His grades improved a lot.)** **Then Danielle (did/did it) too.** **Did Danielle meet with a tutor? Y**
9	temporal	Jennifer cut the tall grass in the yard. (Then it grew back surprisingly quickly.) Two weeks later, Louis (did/did it) too. Did Louis cut the grass two weeks later? Y
10	**elaborative**	**Fred listened to music on the train.** **(His favorite bands were the Beatles and NSync.)** **Usually Wesley (did/did it) too.** **Did Wesley usually listen to music? Y**

11	cause-effect	Donald installed new windows in the apartment. (The cold wind was causing a draft.) That year, Harry (did/did it) too. Did the cold wind cause a draft? Y
12	**temporal**	**Christina shoveled the snow in the driveway.** **(That night it snowed another 3 inches.)** **The next day, Kathy (did/did it) too.** **Did it snow 3 inches that night? Y**
13	**temporal**	**Alex took the trash to the curb.** **(The next day it was picked up by the garbage men.)** **The next week, Bridget (did/did it) too.** **Did the garbage men pick up the trash? Y**
14	**elaborative**	**Pam invited friends over for dinner.** **(The menu was chicken and salad.)** **A few days later, Fred (did/did it) too.** **Was there chicken on the menu? Y**
15	cause-effect	Randy bought a new computer. (Her old one didn't work any more.) Then James (did/did it) too. Did Randy's old computer break? Y
16	elaborative	Nick drove to New York City last week. (He borrowed his roommate's car.) Then Leah (did/did it) too. Did Marilyn drive to New York City? N
17	**cause-effect**	**Marjorie went to a party last weekend.** **(She was bored of sitting at home.)** **The same night, Philip (did/did it) too.** **Did Marjorie go to a party this weekend? N**
18	**temporal**	**Derek stacked the boxes in a tall pile.** **(Then they all fell over.)** **Then Joe (did/did it) too.** **Did Derek stack the books in a tall pile? N**
19	**elaborative**	**Nicole ate her eggs with a spoon.** **(They were soft and mushy.)** **Afterwards, Jenna (did/did it) too.** **Did Nicole eat her eggs with a fork? N**
20	cause-effect	Ray returned his library books. (They were three weeks overdue.) Soon after, Anne (did/did it) too. Did Ray keep his library books? N
21	temporal	Virginia cut all her hair off last year. (Then it grew long again.) Last week, Kathleen (did/did it) too. Did Kathleen cut her hair today? N
22	**elaborative**	**Patrick hosted a dinner party yesterday.** **(Twelve of his friends came.)** **Then Martha (did/did it) too.**

		Did Mary host a dinner party? N
23	cause-effect	Carolyn went to an orchestra concert. (Classical music was her favorite.) That night, Harold (did/did it) too. Did Harold go to a rock concert? N
24	**temporal**	**Douglas bought tickets to a baseball game.** **(Then the game was cancelled.)** **The same day, Carl (did/did it) too.** **Did Carrie buy tickets to the baseball game? N**
25	elaborative	Christine took the kids to the park. (They played on the jungle gym.) The next day Marie (did/did it) too. Did Marie also take the dogs to the park? N
26	cause-effect	Irene broke her ankle last year. (It got caught in a hole in the sidewalk.) Last month, Jane (did/did it) too. Did Irene trip on a crack in the sidewalk? N
27	temporal	Aaron bought jelly beans at the store. (Then they all spilled on the floor.) Later, Andrea (did/did it) too. Did the jelly beans spill on the table? N
28	elaborative	Wanda got a haircut last week. (Her hair was long and brown.) Today, Howard (did/did it) too. Was Wanda's hair brown? N
29	**temporal**	**Lillian went ice skating on the pond.** **(The next week, the ice thawed.)** **That day, Todd (did/did it) too.** **Did the ice thaw that week? N**
30	**elaborative**	**Jesse bought a set of new curtains.** **(They were green with white stripes.)** **Soon after, Craig (did/did it) too.** **Did the curtains have orange stripes? N**
31	**temporal**	**Rebecca bought a new bike.** **(A week later, it was stolen.)** **Then Ruth (did/did it) too.** **Did Rebecca buy a bike? Y**
32	**elaborative**	**George took the trash out to the curb.** **(The trash was heavy and hard to carry.)** **Last week, Sally (did/did it).** **Did George take the trash to the curb? Y**
33	cause-effect	Paula chopped wood all day for a bonfire. (Her hands were sore and cracked.) To help out, Ken (did/did it) too. Did Paula chop wood? Y
34	temporal	Matt bought a new suit for an interview.

		(The next day, the pants ripped at the seam.) Later, John (did/did it) too. Did Matt buy a suit for an interview? Y
35	elaborative	Cathy made steak and potatoes for dinner. (It was very easy and quick to prepare.) The next day, Jennifer (did/did it) too. Did Cathy make steak and potatoes? Y
36	cause-effect	Amelia studied all night for the science test. (She really didn't want to fail.) That night, Karen (did/did it) too. Did Karen study for the science test? Y
37	temporal	Jeremy baked chocolate chip cookies. (Then they were all eaten immediately.) The next night, Julia (did/did it) too. Did Julia bake cookies the next night? Y
38	elaborative	Elena took the dog for a walk. (The trail went through the woods and over a river.) The day after, Lars (did/did it) too. Did Lars take the dog for a walk? Y
39	**cause-effect**	**Tom ate hamburgers every day for lunch.** **(Because of it, he was very fat.)** **Unfortunately, Pedro (did/did it) too.** **Did Pedro also eat hamburgers for lunch? Y**
40	temporal	Linda poured juice into a cup. (Then the juice spilled on the table.) After that, Mary (did/did it) too. Did Mary pour juice into a cup too? Y
41	elaborative	Marcus went to the opera last week. (It was La Boheme by Puccini.) Later that week, Vera (did/did it) too. Was the opera written by Puccini? Y
42	cause-effect	Lucy wrote the term paper by hand. (Her computer was broken.) Unfortunately, Andy (did/did it) too. Was Lucy's computer broken? Y
43	temporal	Samantha paid her bills on time. (Then they got lost in the mail.) Fortunately, Francine (did/did it) too. Did Samantha's bills get lost? Y
44	elaborative	Adam adopted a cat from a shelter. (It was orange and white with a pink nose.) Later that week, Regina (did/did it) too. Was Adam's cat orange and white? Y
45	cause-effect	Rita went to the doctor on Thursday. (She had a sore throat and a fever.) A few days later, Meredith (did/did it) too. Did Rita have a sore throat? Y

46	temporal	Arlene planted bulbs in a pot. (A month later, they were full grown.) Then Jesse (did/did it) too. Did Arlene plant bulbs outside? N
47	elaborative	Desmond dropped out of school. (He was only in tenth grade.) The next year, Charles (did/did it) too. Did Desmond stay in school? N
48	cause-effect	Maria swam across the pond. (Afterwards, she was very cold and tired.) Then Dorothy (did/did it) too. Did Maria swim across the river? N
49	temporal	Steve stole a laptop from the store. (The next day, the police found him.) The next day, Lisa (did/did it) too. Did Mack steal a laptop from the store? N
50	**elaborative**	**Betty went rollerblading by the river.** **(The weather was warm and sunny.)** **That day Brian (did/did it) too.** **Did Betty go rollerskating by the river? N**
51	**cause-effect**	**Arthur left the window open last week.** **(All the rain came in and soaked the floor.)** **That night, Janet (did/did it) too.** **Did Janet close the window that night? N**
52	temporal	Joyce spilled coffee on the table. (Then it dried into a hard brown crust.) Then Roger (did/did it) too. Did Roger also spill juice on the table? N
53	elaborative	Juan moved into a new apartment last year. (It was closer to his job than his old apartment.) This year, Jack (did/did it) too. Did Jack stay in his old apartment this year? N
54	cause-effect	Alice ate a burrito for lunch. (It made her stomach hurt a lot.) Yesterday, Theresa (did/did it) too. Did Theresa eat a burrito for dinner? N
55	**temporal**	**Albert bought a pet bird.** **(Later, it flew away.)** **Then Gloria (did/did it) too.** **Did Gloria also buy a pet cat? N**
56	cause-effect	Eugene went to the beach on Saturday. (The weather was too nice to stay indoors.) On Sunday, Carlos (did/did it) too. Was the weather bad? N
57	**temporal**	**Bonnie bought a house on the beach.** **(The next year it was destroyed by a hurricane.)**

		Then Ruby (did/did it) too. **Did the house stand up to the hurricane? N**
58	elaborative	Russell made dinner for his girlfriend. (It was romantic and delicious.) Later, Lois (did/did it) too. Was the dinner terrible? N
59	**cause-effect**	**Robin stayed home from the party.** **(Her head hurt too much to go out.)** **The same night, Peggy (did/did it) too.** **Did Robin's stomach hurt too much? N**
60	temporal	Alan put a scarecrow up in the yard. (Then it was knocked over by the wind.) After that, Crystal (did/did it) too. Was Alan knocked over by the wind? N
61	**cause-effect**	**Christopher washed his car in the driveway.** **(It was covered in dust and mud.)** **Yesterday, Amanda (did/did it) too.** **Did Christopher wash his car in the driveway? Y**
62	**temporal**	**Vera sent a memo to the secretary.** **(Then the memo was forwarded to the boss.)** **Later, Ned (did/did it) too.** **Did Vera send a memo? Y**
63	**elaborative**	**Larry read the newspaper over coffee.** **(Lots of terrible things were happening in the world.)** **That morning, Paul (did/did it) too.** **Did Larry read the newspaper over coffee? Y**
64	cause-effect	Mary entered the beauty pageant. (There was a huge prize for first place.) That day, Grace (did/did it) too. Did Mary enter the pageant? Y
65	temporal	Luke jumped off the cliff into the pond. (Later, he realized he was bruised.) Afterwards, Diana (did/did it) too. Did Luke jump off the cliff? Y
66	elaborative	Harry jumped off a bridge. (It was a very small bridge.) Then Kendra (did/did it) too. Did Kendra jump off a bridge? Y
67	cause-effect	Marlene slipped on the icy sidewalk. (The ice was hard to see and very slippery.) Later, Hank (did/did it) too. Did also Hank slip on the sidewalk? Y
68	temporal	Richard got Chinese food for dinner. (While he was out, his house was robbed.) That night, Greg (did/did it) too. Did Greg get Chinese food that night? Y

69	**elaborative**	**Vanessa asked the new boy out on a date.** **(He was very cute and smart.)** **Unfortunately, Laura (did/did it) too.** **Did Laura ask the new boy on a date too? Y**
70	cause-effect	Ike smoked a cigarette in bed. (His house caught on fire and burned down.) Later, Sandra (did/did it) too. Did Sandra smoke a cigarette in bed later? Y
71	temporal	Brendan bought milk at the store. (Then it spoiled in the hot car.) Later, Nick (did/did it) too. Did the milk spoil in the car? Y
72	elaborative	Miriam went to a hockey game on Tuesday. (The home team played poorly and lost.) On Wednesday, Suzanne (did/did it) too. Did the hockey team lose? Y
73	cause-effect	William stayed home from school yesterday. (His temperature was very high.) Last week, Holly (did/did it) too. Did William have a fever? Y
74	temporal	Theo threw a ball against the wall. (Then it bounced off and rolled into the woods.) Then Faith (did/did it) too. Did the ball roll into the woods? Y
75	elaborative	Adele swept the dust off the floor. (It was thick and disgusting.) The next week, Max (did/did it) too. Was the dust disgusting? Y
76	cause-effect	Ronald washed all of the sheets and towels. (They had not been washed in weeks.) The next week, Jason (did/did it) too. Did Ronald wash all the clothes? N
77	temporal	Donna put the bread in the oven. (An hour later, it had risen.) Then Carol (did/did it) too. Did Donna put the pizza in the oven? N
78	elaborative	Gary sent an email to his brother. (He wanted to write about his vacation.) Later, Sharon (did/did it) too. Did Gary send an email to his mother? N
79	cause-effect	Michelle drank a big glass of water. (She was unbelievably thirsty.) An hour later, Timothy (did/did it) too. Did Michelle drank a small glass of water? N
80	temporal	Jose sent a letter to the president. (Three weeks later, he got a reply.) Soon after, Deborah (did/did it) too.

		Did Jose send an email to the president? N
81	elaborative	Evelyn drank a bottle of soda. (It was delicious and refreshing.) Later, Justin (did/did it) too. Did Justin drink a bottle of juice later? N
82	**cause-effect**	**Gerald stayed up late reading last night.** **(He was not tired enough to sleep.)** **Unfortunately, Keith (did/did it) too.** **Did Ken also stay up late reading? N**
83	temporal	Jean took her umbrella to work. (Then the sun came out that afternoon.) That day, Cheryl (did/did it) too. Did Cheryl take her umbrella home too? N
84	elaborative	Ralph took a taxi to the airport. (The airport was far from his house.) That day, Ashley (did/did it) too. Did Ashley drive to the airport? N
85	cause-effect	Judith mopped the floor in the kitchen. (It was covered in grime.) Later that month, Lawrence (did/did it) too. Did Lawrence mop the bathroom floor? N
86	**cause-effect**	**Tina left the barn door open last week.** **(All the cows escaped.)** **Last night, Bobby (did/did it) too.** **Did all the goats escape from the barn? N**
87	temporal	Victor put some candy out on the table. (Then it was all devoured.) Today, Martin (did/did it) too. Did any candy remain on the table? N
88	elaborative	Dawn quit her job last month. (She had been working as an engineer.) Then Shawn (did/did it) too. Was Dawn working as a secretary? N
89	cause-effect	Clarence bought coffee at a cafe this morning. (There was no time to make it at home.) This afternoon, Earl (did/did it) too. Did Clarence have time to make coffee? N
90	temporal	Connie bought expensive tickets to a musical. (The next day, they went on sale.) Then Florence (did/did it) too. Did the tickets go on sale the next week? N
91	elaborative	Claire threw an egg at the house. (It was a rotten, stinky egg.) After that, Tom (did/did it) too. Did Claire throw an egg at the house? Y
92	cause-effect	Edward stretched before baseball practice.

		(He didn't want to get injured during practice.) At the same time, Josh (did/did it) too. Did Edward stretch before practice? Y
93	temporal	Melanie stole money from the cash register. (Then her boss found out.) The next day, Tammy (did/did it) too. Did Melanie steal money? Y
94	elaborative	Frank went ice-skating on the pond yesterday. (It was a beautiful winter day.) Later, Sarah (did/did it) too. Did Frank ice-skate on the pond yesterday? Y
95	**cause-effect**	**Jeanne watered the plants in the sunroom.** **(They were looking droopy and wilted.)** **Later on, Dan (did/did it) too.** **Did Jeanne water the plants? Y**
96	temporal	Trisha ate tainted meat at a restaurant. (Later, her stomach hurt.) The same night, Maurice (did/did it) too. Did Maurice eat meat at a restaurant? Y
97	elaborative	Reggie broke a glass at dinner. (It was an expensive wine glass.) Immediately, Joe (did/did it) too. Did Joe break a glass too? Y
98	cause-effect	Flora gave the dog a treat. (It was behaving very well.) Later, Carrie (did/did it) too. Did Carrie give the dog a treat later? Y
99	temporal	Mario went outside to water the flowers. (Soon after, it started raining.) The next day, Angela (did/did it) too. Did Angela water the flowers too? Y
100	elaborative	Wendy saw a new movie at the theater. (It was a comedy starring Jack Black.) That weekend, Patrick (did/did it) too. Did Patrick see a new movie at the theater? Y
101	temporal	Alicia bought new shoes last week. (Then they turned out to be too small.) Today, Danny (did/did it) too. Were Alicia's shoes too small? Y
102	elaborative	Jordan built a bookshelf in shop class. (It was made of beautiful oak wood.) Later, Steve (did/did it) too. Was the bookshelf made of wood? Y
103	**cause-effect**	**Justine passed her driving test.** **(She made only one mistake.)** **Fortunately, Rosa (did/did it) too.** **Did Justine make one mistake? Y**

104	cause-effect	Cal picked the books up from the floor. (They had fallen off the shelf.) Later, Sean (did/did it) too. Did the books fall off the shelf? Y
105	temporal	Allison bought a new scooter. (The next day it broke down.) The next week, Heather (did/did it) too. Did Allison's scooter break down? Y
106	elaborative	Scott turned on the television. (There were lots of good shows on.) Then Eric (did/did it) too. Did Scott turn off the television? N
107	cause-effect	Jessica went to the gas station yesterday. (Her car was out of gas.) Today, Shirley (did/did it) too. Did Irene go to the gas station? N
108	temporal	Andrew boarded the plane at noon. (Then the flight was delayed.) Minutes later, Cynthia (did/did it) too. Did Andrew board the train at noon? N
109	elaborative	Amy took a nap on the couch. (The couch was very uncomfortable.) Later, Jerry (did/did it) too. Did Amy take a nap on the bed? N
110	cause-effect	Dennis fed the cat only organic food. (He was worried it would get sick.) Oddly, Walter (did/did it) too. Did Dennis feed the dog organic food? N
111	temporal	Roy put up the storm windows in his house. (Later that month, it snowed.) That week, Bruce (did/did it) too. Did Bruce also put up screen windows? N
112	elaborative	Janice registered to vote today. (Registration was at the town hall.) At the same time, Kelly (did/did it) too. Did Mark also register to vote? N
113	cause-effect	Brandon bought the newspaper yesterday. (There had been a story worth reading about.) Today, Denise (did/did it) too. Did Denise buy a newspaper yesterday? N
114	temporal	Tammy went hiking on a volcano. (The next day, the volcano erupted.) Later, Wayne (did/did it) too. Did Wayne go jogging on a volcano? N
115	elaborative	Billy did his homework last night. (He had math and science homework.)

		Today, Louis (did/did it) too. Did Louis do homework last week? N
116	**elaborative**	**Phyllis borrowed a book from the library.** **(It was about the Spanish revolution.)** **On Friday, Norma (did/did it) too.** **Was the book about the American revolution? N**
117	cause-effect	Ernest took the train home for Thanksgiving. (He was scared of flying in an airplane.) Last year, Paula (did/did it) too. Was Ernest scared of taking the train? N
118	elaborative	Jimmy skipped class on Monday. (Nobody seemed to notice.) On Tuesday, Tiffany (did/did it) too. DId anyone notice that Jimmy skipped class? N
119	cause-effect	Carmen forgot her umbrella yesterday. (The rain soaked through her clothes.) Unfortunately, Antonio (did/did it) too. Did the rain stay off Carmen's clothes? N
120	temporal	Danny boarded the train at the last minute. (Then it left the station.) Fortunately, Cindy (did/did it) too. Did the bus leave the station? N

Appendix B: Fillers for distance experiments. Y/N after each item indicates the correct answer to the comprehension question. All distance fillers were used in all distance experiments.

1 Ron went to bed early last night.
Robert did the same thing.
Did Ron go to bed early last night? Y

2 Susan took the train to the beach last weekend.
Laura took the bus.
Did Susan take a train last weekend? Y

3 Terrence wrote a book about food and wine.
Julie wrote one about traveling.
Did Terrence write a book? Y

4 Claire bought a washing machine at the store.
Later, Nate bought one too.
Did Claire buy a washing machine? Y

5 Henry lifted the box up onto the shelf.
Later, Jerry took it down.
Did Henry lift the box onto the shelf? Y

6 Emily rode a bike to work today.
Kerry does on Mondays.
Did Emily ride a bike today? Y

7 Robert walked to the bus stop in the morning.
Jill walked there at night.
Did Robert walk to the bus in the morning? Y

8 Elena threw the lamp across the room.
The next day, Frank cleaned it up.
Did Elena throw a lamp? Y

9 Greg ate an entire pizza last night.
Wesley ate one too.
Did Greg eat a pizza last night? Y

10 Francine won first place at the spelling bee.
Kara won second place.
Did Francine win the spelling bee? Y

11 Jerry found ten dollars on the sidewalk.
Stella found twenty.
Did Jerry find money on the sidewalk? Y

12 Samantha took the dog for a walk.
Then Bob took it for a walk.
Did Samantha walk the dog? Y

13 Mark ate a turkey sandwich for dinner.
Later, Paul ate a chicken sandwich.
Did Mark eat a turkey sandwich? Y

14 Kara went for a run next to the river.
Mary didn't go.
Did Kara run near the river? Y

15 Richard baked a cake last night.
Then Patricia frosted it.
Did Richard bake a cake last night? Y

16 Elaine drinks tea every day.
Ken drinks tea too.
Did Elaine drink coffee today? N

17 Ben went to the gym on Tuesday.
On Wednesday, Terrence went too.
Did Ben go to the gym on Thursday? N

18 Rachel bought a new computer.
Afterwards, Irene bought one too.
Did Rachel sell a computer? N

19 Wesley drove to Maine for the holidays.
Amanda took the train.
Did Jeffrey drive to Maine? N

20 Mary called her mom on the phone.
The next day, Fred called his mom too.
Did Mary call her dad on the phone? N

21 Luke went to the dentist yesterday.
Henry went last month.
Did Luke go to the dentist today? N

22 Anita took a vacation to Hawaii last winter.
Nadia went on vacation too.
Did Anita take a vacation to Hawaii this winter?
N

23 Steve bought a car for his son.
Linda wanted to do the same.
Did Steve buy a truck for his son? N

24 Tracy read the newspaper at the table.
Bob read it on the train.
Did Tracy read the newspaper on the couch? N

25 Dave went to a baseball game last weekend.
This weekend, Eric is going too.
Did Dave go to a football game? N

26 Liz bought a present for a friend.
Alexis also bought a present.
Did Liz buy a present for her boyfriend? N

27 Jack quit his job last week.

Helen almost quit hers too.
Did Jack quit his job this week? N

28 Jane got new glasses yesterday.
So did Nick.
Did Paula get new glasses? N

29 Hector accidentally fell off a ladder.
John wasn't holding it properly.
Did Hector fall off a house? N

30 Ruby bought a new house last year.
This year, Eleanor bought one too.
Did Ruby buy a new car? N

31 Alex locked himself out of his house.
Then Talia did the same thing.
Did Talia lock herself out? Y

32 Alicia bought a book online.
Later, Greg bought the same one.
Did Greg buy a book online? Y

33 Larry went sledding in December.
Joseph went skiing in December.
Did Joseph go skiing in December? Y

34 Dorothy quit smoking last month.
Then Nancy quit too.
Did Nancy quit smoking? Y

35 Brian went for a walk in the woods.
Miriam went along too.
Did Miriam go for a walk in the woods? Y

36 Carol went on a plane for the first time.
Then Brett took a plane too.
Did Brett go on a plane? Y

37 Stanley ordered a beer at the bar.
Later, Nathan ordered wine.
Did Nathan order wine at the bar? Y

38 Leah stayed home from work yesterday.
Nina stayed home today.
Did Nina stay home from work? Y

39 Travis took a surfing lesson last summer.
Naomi took one too.
Did Naomi take a surfing lesson? Y

40 Jenny bought ice cream after lunch.
That night, Kyle had some too.
Did Kyle have some ice cream that night? Y

41 Al stayed at work late last night.
Brad left work early.
Did Brad leave work early? Y

42 Daisy bought a puppy today.
Lindsay wants to do the same thing.
Does Linda want to buy a puppy? Y

43 Barry watched television all day yesterday.
Cheryl read a book instead.
Did Cheryl read a book yesterday? Y

44 Natasha borrowed a bicycle from her sister.
Eddie bought one at the store.
Did Eddie buy a bicycle at the store? Y

45 Bernard built a bookshelf for his apartment.
Marcus did the same thing.
Did Marcus build a bookshelf? Y

46 Cora applied for graduate school last month.
Joanna also applied to school.
Did Joanna apply for a job? N

47 Jay started a rock band with his friends.
Sabrina joined the band.
Did Sabrina join a jazz band? N

48 Molly cleaned out the fridge today.
Then Tom also cleaned his fridge.
Did Tom clean the fridge yesterday? N

49 Derek planted a garden in the spring.
This year, Jerome will too.
Did Jerome plant a garden last spring? N

50 Amelia stayed home from the party.
Iris stayed home too.
Did Iris go to the party? N

51 Gordon got a speeding ticket last Friday.
Justin got one yesterday.
Did Justin get a speeding ticket today? N

52 Claudia got caught in the rain yesterday.
Tanya made it inside in time.
Did Tanya get caught in the rain? N

53 Martin celebrated his birthday in June.
Vera's was in May.
Did Vera have a birthday in March? N

54 Natalie joined the circus at age 20.
A year later, Andrew also joined.
Did Alex join the circus? N

55 Daniel went out to dinner last night.
Jack also went out.
Did Jack eat dinner at home last night? N

56 Vivian bought bananas at the supermarket.

Audrey bought apples.
Did Audrey buy apples at the orchard? N

57 Scott went to the ballet on Saturday.
On Sunday, Lynn went too.
Did Lynn go to the ballet on Saturday? N

58 Gail listened to music on the couch.
Patrick also listened to music.
Did Patrick read a book? N

59 Kevin donated money to the shelter.
Later, Harold donated too.
Did Harold donate money to the school? N

60 Amber auditioned for the school play.
Monica auditioned that day too.
Did Monica audition the next day? N

61 Theo ran a ten mile race last week.
He had been training for months.
Justin also ran that day.
Did Theo run a race last week? Y

62 Kathryn brought the books back to the library.
They were too boring to read.
Later, Ruth did the same thing.
Did Kathryn return the books to the library? Y

63 Oliver ate a sandwich for lunch yesterday.
It was ham and swiss on rye bread.
Nancy ate a meatball sub.
Did Oliver eat a sandwich yesterday? Y

64 Rachel learned Spanish last summer.
It was easier than she expected it to be.
Rick learned Spanish in the fall.
Did Rachel learn Spanish? Y

65 Colin sailed a boat to Maine.
It was a very exciting journey.
Dave decided not to go.
Did Colin sail a boat to Maine? Y

66 Allison took a ferry to the island.
It was a long and bumpy journey.
Leslie took the ferry too.
Did Allison take a ferry? Y

67 Gary went to a movie last night.
It was a scary movie about monsters.
Then Sheila went to the movie too.
Did Gary go to a movie last night? Y

68 Connie vacuumed the floor yesterday.
It was covered in dust and dog hair.
Meanwhile, Eric swept the hallway.
Did Connie vacuum the floor? Y

69 Carlos practiced the violin last night.
It had been a long time since he played.
Today, Wayne practiced too.
Did Carlos practice the violin last night? Y

70 Tiffany made chicken soup for dinner.
It turned out too salty.
Grace made split pea soup instead.
Did Tiffany make chicken soup? Y

71 Bill bought a new watch at the store.
It was made of stainless steel.
Sylvia also bought one.
Did Bill buy a new tie at the store? N

72 Edith took guitar lessons last year.
It was easy to learn the chords.
Bobby took lessons too.
Did Edith take guitar lessons this year? N

73 Todd did all of his chores on time.
He was a hard worker.
Shawn didn't do all of his chores.
Did Bob do his chores? N

74 Ellen threw out her old shoes.
They were smelly and falling apart.
Emma threw out hers too.
Did Ellen throw out her new shoes? N

75 Craig trained for the marathon all spring.
It was the hardest thing he'd ever done.
Lucy was also training.
Did Craig train for a triathlon? N

76 Eleanor mended the ripped jeans.
They were torn on the knees.
Jimmy brought his to the tailor.
Did Eleanor take her jeans to the tailor? N

77 Antonio won a hundred dollars in the lottery.
He wanted to put it in the bank.
Brandon didn't win anything.
Did Antonio win a thousand dollars? N

78 Tina took the cat to the vet yesterday.
It was time for the yearly checkup.
Jane is taking her cat on Thursday.
Did Tina go to the vet today? N

79 Roger went fishing last summer.
He caught three fish.
Katherine went fishing too.
Did Roger go fishing last spring? N

80 Martha bought a new CD at the store.
It was a very popular CD.

Richard bought the same one.
Did Martha buy a new DVD at the store? N

81 Ed ate his lunch under a big oak tree.
It was a beautiful sunny day.
Karl ate his at his desk.
Was it a beautiful day? Y

82 Janice had an interview at the company.
She was applying for a new job.
Later, Olivia had one too.
Was Janice applying for a new job? Y

83 Arthur bought tickets for the hockey game.
They were very expensive.
That night, Maggie bought some too.
Were the hockey tickets expensive? Y

84 Flora downloaded a new software program.
It was a helpful program for work.
Today, Gerald did the same thing.
Was the software program helpful for work? Y

85 Louis saved up money for vacation.
He wanted to go to Greece.
Russell didn't go on vacation.
Did Louis want to go to Greece? Y

86 Lola visited her friend in New York.
They went to lots of museums.
Natasha also went to New York.
Did Lola visit any museums? Y

87 Chad saw a zebra at the zoo.
It had black and white stripes.
That day, Nora saw one too.
Did the zebra have black and white stripes? Y

88 Bess went ice skating on the pond.
It was cold and windy.
Alan also went skating.
Was it a cold and windy day? Y

89 Howard ran through the sprinkler.
It was a very hot day.
Roy drank lemonade instead.
Was it a very hot day? Y

90 Wilma left her purse on the bus by accident.
She lost her wallet and keys.
Brittany also lost her purse.
Did Wilma lose her wallet? Y

91 Gerald cut his finger on the broken glass.
He had to get stitches.
Last year, Joy cut her finger too.
Did Marcia get stitches? N

92 Cassandra wrote a book about her life.
She tried to get it published.
Jeremy wrote a cookbook.
Did Cassandra publish a cookbook? N

93 Neil went skydiving with his friends.
It was a terrifying experience.
Wayne also went skydiving.
Was skydiving relaxing? N

94 Sabrina carried the boxes upstairs.
They were heavy and large.
Shelly carried some too.
Were the boxes small? N

95 Russell painted a picture yesterday.
It was of a beautiful landscape.
Today, Isabelle painted a picture too.
Was the picture of a person? N

96 Kayla studied English in college.
She wanted to be a writer.
Mike studied English too.
Did Kayla want to be a teacher? N

97 Norman planted flowers in the yard.
They were red tulips.
Later, Jake planted flowers too.
Were the flowers red roses? N

98 Susie baked cookies for her neighbors.
They had just moved in downstairs.
The next week, Julie also made cookies.
Did the neighbors live upstairs? N

99 Alfred walked along the beach last night.
It was beautiful and quiet.
This morning, Myra did the same thing.
Was the beach crowded? N

100 Rosalie called the police about the noise.
The music was shaking her house.
Later, John also called the police.
Was the noise from construction? N

101 Leo got stuck in a snowbank yesterday.
The snow was 5 feet high.
Herman did not get stuck.
Did Herman not get stuck? Y

102 Whitney had a picnic in the park today.
She brought sandwiches and lemonade.
Carla went to the picnic too.
Did Carla also go to the picnic? Y

103 Don bought wine at the liquor store.
It was for a dinner party.
Caroline bought the same wine.

Did Caroline bring wine too? Y

104 Nellie wrote an article for the newspaper.
It was about crime in the city.
The next week, Todd also wrote an article.
Did Todd write an article for the newspaper? Y

105 James made a pile of leaves outside.
He wanted to jump in them.
Josh played in the leaves too.
Did Josh play in the leaves outside? Y

106 Lucille ordered takeout for dinner.
She was too tired to cook.
That night, Renee ordered out too.
Did Renee order takeout for dinner? Y

107 Andrew learned to play the flute last year.
He had to practice a lot.
Stacy learned to play the oboe.
Did Stacy learn the oboe? Y

108 Veronica went to the aquarium on Thursday.
There were lots of dolphins and fish.
On Friday, Michael went too.
Did Michael go to the aquarium on Friday? Y

109 Dennis replaced the windows in the spring.
The old ones let in a draft.
Leo replaced the windows too.
Did Leo also replace the windows? Y

110 Emma hired an exterminator last week.
Her house was overrun with bugs.
Vivian needed an exterminator too.
Did Vivian need an exterminator? Y

111 Mario went to a jazz club last weekend.
The musicians were very talented.
Katrina went to the same club.
Did Katrina go to a different club? N

112 Harriet performed in a comedy show.
It was really hilarious.
Lewis was in the show too.
Did Lewis watch the show? N

113 Brent sent flowers to his girlfriend.
It was Valentine's Day.
Shane did the same thing.
Did Shane send chocolate to his girlfriend? N

114 Diana had a huge birthday party last year.
It was her thirtieth birthday.
Betsy had a small party.
Did Betsy have a large birthday party? N

115 Jason volunteered at the library last week.
He spent the day cataloguing books.
Later, Claudia volunteered too.
Did Claudia volunteer at the shelter? N

116 Agnes asked for a raise at work.
She had been working very hard.
Frank asked for one too.
Did Frank ask for a bonus? N

117 Willie went skiing in Maine this year.
It was an excellent vacation.
Charles went skiing too.
Did James also go skiing? N

118 Kay knit a sweater for her sister.
It was purple and fuzzy.
The next month, Heidi knit a sweater too.
Did Heidi knit a hat too? N

119 Bruce drank too much at the holiday party.
It was very embarrassing.
Maureen didn't drink anything.
Did Maureen drink too much? N

120 Gertrude broke a plate by accident.
She dropped it while washing it.
Then Floyd broke a plate too.
Did Floyd break a glass? N

Appendix C: Materials for particle shift experiments. All materials were used in the reading time experiment. Items in bold type were used in the fMRI experiment.

1 **Rebecca (dropped Jonathan off/dropped off Jonathan) at the pool.**
He was excited that she (drove/did/did it).
Rebecca (dropped Jonathan off/dropped off Jonathan) at the pool.

2 **Judy (turned Anthony down/turned down Anthony) in the meeting.**
He was flustered that she (refused/did/did it).
Judy (turned Anthony down/turned down Anthony) in the meeting.

3 Wendy (woke Vince up/woke up Vince) before school.
He was resentful that she (was loud/did/did it).
Wendy (woke Vince up/woke up Vince) before school.

4 Lauren (sent Ben out/sent out Ben) for groceries.
He was mad that she (was bossy/did/did it).
Lauren (sent Ben out/sent out Ben) for groceries.

5 Paula (let Fred down/let down Fred) at the competition.
He was disappointed that she (lost/did/did it).
Paula (let Fred down/let down Fred) at the competition.

6 Jenny (scared Lucas off/scared off Lucas) from the haunted house.
He was upset that she (was frightening/did/did it).
Jenny (scared Lucas off/scared off Lucas) from the haunted house.

7 Larissa (took Gary out/took out Gary) to the backyard.
He was happy that she (insisted/did/did it).
Larissa (took Gary out/took out Gary) to the backyard.

8 Edith (crossed Jake out/crossed out Jake) from the roster.
He was sad that she (was mean/did/did it).
Edith (crossed Jake out/crossed out Jake) from the roster.

9 Kathryn (knocked Peter out/knocked out Peter) by accident.
He was shocked that she (was strong/did/did it).
Kathryn (knocked Peter out/knocked out Peter) by accident.

10 Fran (asked Randall out/asked out Randall) to a movie.
He was flattered that she (was interested/did/did it).
Fran (asked Randall out/asked out Randall) to a movie.

11 Rosa (held Harry up/held up Harry) to see the show.
He was thankful that she (helped out/did/did it).
Rosa (held Harry up/held up Harry) to see the show.

12 Nicole (left Denny out/left out Denny) from the discussion.
He was furious that she (was rude/did/did it).
Nicole (left Denny out/left out Denny) from the discussion.

13 Anne (let the boy in/let in the boy) from the rain.
He was thankful that she (was kind/did/did it).
Anne (let the boy in/let in the boy) from the rain.

14 Justine (kicked the chairman out/kicked out the chairman) at the meeting.
He was embarrassed that she (was angry/did/did it).

Justine (kicked the chairman out/kicked out the chairman) at the meeting.

15 Nancy (laid the engineer off/laid off the engineer) after the budget cuts.
He was stressed that she (was making cuts/did/did it).
Nancy (laid the engineer off/laid off the engineer) after the budget cuts.

16 Bethany (dropped the inspector off/dropped off the inspector) at the site.
He was pleased that she (was helpful/did/did it).
Bethany (dropped the inspector off/dropped off the inspector) at the site.

17 Claire (turned the professor down/turned down the professor) at the conference.
He was displeased that she (refused/did/did it).
Claire (turned the professor down/turned down the professor) at the conference.

18 Meghan (woke the passenger up/woke up the passenger) on the train.
He was embarrassed that she (noticed/did/did it).
Meghan (woke the passenger up/woke up the passenger) on the train.

19 Julie (asked the banker out/asked out the banker) for dinner.
He was thrilled that she (was interested/did/did it).
Julie (asked the banker out/asked out the banker) for dinner.

20 Katie (held the boy up/held up the boy) to the window.
He was glad that she (was helpful/did/did it).
Katie (held the boy up/held up the boy) to the window.

21 Greta (left the new student out/left out the new student) from the party.
He was angry that she (was forgetful/did/did it).
Greta (left the new student out/left out the new student) from the party.

22 Olga (cleaned the child up/cleaned up the child) in the tub.
He was unhappy that she (was helping/did/did it).
Olga (cleaned the child up/cleaned up the child) in the tub.

23 Sally (brought the grandfather back/brought back the grandfather) to the house.
He was grateful that she (helped out/did/did it).
Sally (brought the grandfather back/brought back the grandfather) to the house.

24 Kara (lifted the patient up/lifted up the patient) from the chair.
He was glad that she (assisted/did/did it).
Kara (lifted the patient up/lifted up the patient) from the chair.

25 The stewardess (cheered Dan up/cheered up Dan) after the incident.
He was glad that she (was optimistic/did/did it).
The stewardess (cheered Dan up/cheered up Dan) after the incident.

26 The waitress (figured Thomas out/figured out Thomas) before the incident occurred.
He was mortified that she (knew/did/did it).
The waitress (figured Thomas out/figured out Thomas) before the incident occurred.

27 The matchmaker (set Matthew up/set up Matthew) on a date.
He was thrilled that she (was involved/did/did it).
The matchmaker (set Matthew up/set up Matthew) on a date.

28 The maid (let Jonathan in/let in Jonathan) from the porch.
He was relieved that she (was there/did/did it).
The maid (let Jonathan in/let in Jonathan) from the porch.

29 The babysitter (kicked Vincent out/kicked out Vincent) from the kitchen.
He was frustrated that she (was angry/did/did it).
The babysitter (kicked Vincent out/kicked out Vincent) from the kitchen.

30 The librarian (laid Danny off/laid off Danny) on Friday.
He was devastated that she (had to/did/did it).
The librarian (laid Danny off/laid off Danny) on Friday.

31 The policewoman (picked Bob up/picked up Bob) from jail.
He was delighted that she (offered/did/did it).
The policewoman (picked Bob up/picked up Bob) from jail.

32 The secretary (looked Adam up/looked up Adam) in the directory.
He was annoyed that she (was nosey/did/did it).
The secretary (looked Adam up/looked up Adam) in the directory.

33 The dancer (picked Ken out/picked out Ken) from the group.
He was thrilled that she (noticed/did/did it).
The dancer (picked Ken out/picked out Ken) from the group.

34 The doctor (cleaned Luke up/cleaned up Luke) on the table.
He was disgusted that she (was so thorough/did/did it).
The doctor (cleaned Luke up/cleaned up Luke) on the table.

35 The writer (brought Seth back/brought back Seth) in the van.
He was surprised that she (was helpful/did/did it).
The writer (brought Seth back/brought back Seth) in the van.

36 The nurse (lifted the baby boy up/lifted up the baby boy) to the window.
He was frightened that she (was careless/did/did it).
The nurse (lifted the baby boy up/lifted up the baby boy) to the window.

37 The hostess (sent the busboy out/sent out the busboy) to the grocery store.
He was not pleased that she (was bossy/did/did it).
The hostess (sent the busboy out/sent out the busboy) to the grocery store.

38 The lead actress (let the director down/let down the director) during the premier.
He was disappointed that she (performed poorly/did/did it).
The lead actress (let the director down/let down the director) during the premier.

39 The old hag (scared the little boy off/scared off the little boy) on Halloween.
He was unhappy that she (was ugly/did/did it).
The old hag (scared the little boy off/scared off the little boy) on Halloween.

40 The housewife (cheered the plumber up/cheered up the plumber) in the bathroom.
He was grateful that she (was friendly/did/did it).
The housewife (cheered the plumber up/cheered up the plumber) in the bathroom.

41 The milkmaid (figured the farmer out/figured out the farmer) in the field.
He was disconcerted that she (was smart/did/did it).
The milkmaid (figured the farmer out/figured out the farmer) in the field.

42 The seamstress (set the butcher up/set up the butcher) on a date.
He was nervous that she (was interfering/did/did it).
The seamstress (set the butcher up/set up the butcher) on a date.

43 The mother (picked the child up/picked up the child) at the bowling alley.
He was dismayed that she (came early/did/did it).

The mother (picked the child up/picked up the child) at the bowling alley.

44 The girl (looked the philosopher up/looked up the philosopher) in the encyclopedia.
He was unaware that she (was interested/did/did it).
The girl (looked the philosopher up/looked up the philosopher) in the encyclopedia.

45 The woman (picked the criminal out/picked out the criminal) from the lineup.
He was perturbed that she (remembered/did/did it).
The woman (picked the criminal out/picked out the criminal) from the lineup.

46 The soprano (took the conductor out/took out the conductor) to a bar.
He was confused that she (was flirtatious/did/did it).
The soprano (took the conductor out/took out the conductor) to a bar.

47 The party planner (crossed the uncle out/crossed out the uncle) from the guest list.
He was indignant that she (was in charge/did/did it).
The party planner (crossed the uncle out/crossed out the uncle) from the guest list.

48 The acrobat (knocked the ringmaster out/knocked out the ringmaster) at the circus.
He was concerned that she (messed up/did/did it).
The acrobat (knocked the ringmaster out/knocked out the ringmaster) at the circus.

49 Jennifer (put Jeff's jacket on/put on Jeff's jacket) to warm up.
He didn't mind that she (was needy/did/did it).
Jennifer (put Jeff's jacket on/put on Jeff's jacket) to warm up.

50 Francine (gave Peter's gift back/gave back Peter's gift) to the store.
He was offended that she (was displeased/did/did it).
Francine (gave Peter's gift back/gave back Peter's gift) to the store.

51 Alice (checked Tom's items off/checked off Tom's items) from the to-do list.
He was relieved that she (was organized/did/did it).
Alice (checked Tom's items off/checked off Tom's items) from the to-do list.

52 Lucy (backed Ron's van up/backed up Ron's van) into the garage.
He was glad that she (drove/did/did it).
Lucy (backed Ron's van up/backed up Ron's van) into the garage.

53 Joy (took Sean's hat off/took off Sean's hat) in the car.
He was disappointed that she (disliked it/did/did it).
Joy (took Sean's hat off/took off Sean's hat) in the car.

54 Roxanne (called Joe's vacation off/called off Joe's vacation) last week.
He was mad that she (was thoughtless/did/did it).
Roxanne (called Joe's vacation off/called off Joe's vacation) last week.

55 Donna (added Greg's earnings up/added up Greg's earnings) after work.
He was grateful that she (was good at math/did/did it).
Donna (added Greg's earnings up/added up Greg's earnings) after work.

56 Mary (passed Hank's photos out/passed out Hank's photos) at the dinner.
He was embarrassed that she (was sharing/did/did it).
Mary (passed Hank's photos out/passed out Hank's photos) at the dinner.

57 Pam (tried Ira's gloves on/tried on Ira's gloves) at the yard sale.
He was amused that she (was curious/did/did it).
Pam (tried Ira's gloves on/tried on Ira's gloves) at the yard sale.

58 Wendy (turned Paul's TV off/turned off Paul's TV) with the remote.
He was unaware that she (impatient/did/did it).
Wendy (turned Paul's TV off/turned off Paul's TV) with the remote.

59 Brenda (drew Ned's plan out/drew out Ned's plan) at the meeting.
He was appreciative that she (explained/did/did it).
Brenda (drew Ned's plan out/drew out Ned's plan) at the meeting.

60 Janet (turned Chris's oven on/turned on Chris's oven) before dinner.
He was aware that she (contributed/did/did it).
Janet (turned Chris's oven on/turned on Chris's oven) before dinner.

61 Jill (did the lawyer's report over/did over the lawyer's report) before the trial.
He was ungrateful that she (helped/did/did it).
Jill (did the lawyer's report over/did over the lawyer's report) before the trial.

62 Maude (mixed the doctor's instructions up/mixed up the doctor's instructions) after the appointment.
He was concerned that she (was confused/did/did it).
Maude (mixed the doctor's instructions up/mixed up the doctor's instructions) after the appointment.

63 Katie (put the chef's knives away/put away the chef's knives) in the drawer.
He was confused that she (was cleaning up/did/did it).
Katie (put the chef's knives away/put away the chef's knives) in the drawer.

64 Sheila (put the intern's sweater on/put on the intern's sweater) in the office.
He was disgusted that she (was cold/did/did it).
Sheila (put the intern's sweater on/put on the intern's sweater) in the office.

65 Nina (gave the teenager's iPod back/gave back the teenager's iPod) after the trip.
He was glad that she (was honest/did/did it).
Nina (gave the teenager's iPod back/gave back the teenager's iPod) after the trip.

66 Lily (checked the historian's dues off/checked off the historian's dues) in the ledger.
He was pleased that she (was responsible/did/did it).
Lily (checked the historian's dues off/checked off the historian's dues) in the ledger.

67 Kelly (turned the old man's light off/turned off the old man's light) in the bedroom.
He was irritated that she (was stingy/did/did it).
Kelly (turned the old man's light off/turned off the old man's light) in the bedroom.

68 Bessie (drew the actor's portrait out/drew out the actor's portrait) on some paper.
He was agitated that she (was slow/did/did it).
Bessie (drew the actor's portrait out/drew out the actor's portrait) on some paper.

69 Gail (turned the reporter's microphone on/turned on the reporter's microphone) before the interview.
He was pleased that she (assisted/did/did it).
Gail (turned the reporter's microphone on/turned on the reporter's microphone) before the interview.

70 Eileen (filled the boy's report card out/filled out the boy's report card) on Tuesday.
He was frightened that she (was harsh/did/did it).
Eileen (filled the boy's report card out/filled out the boy's report card) on Tuesday.

71 Stella (made the visitor's bed up/made up the visitor's bed) in the morning.
He was glad that she (was thoughtful/did/did it).
Stella (made the visitor's bed up/made up the visitor's bed) in the morning.

72 Jean (threw the doctor's notes away/threw away the doctor's notes) by accident.
He was livid that she (was careless/did/did it).

Jean (threw the doctor's notes away/threw away the doctor's notes) by accident.

73 The secretary (gave Ben's secret up/gave up Ben's secret) to the FBI.
He was outraged that she (talked/did/did it).
The secretary (gave Ben's secret up/gave up Ben's secret) to the FBI.

74 The reporter (wrote Nate's reply down/wrote down Nate's reply) on a notepad.
He was proud that she (listened/did/did it).
The reporter (wrote Nate's reply down/wrote down Nate's reply) on a notepad.

75 The salesgirl (wrapped Don's sandwich up/wrapped up Don's sandwich) in the cafe.
He was impressed that she (was fast/did/did it).
The salesgirl (wrapped Don's sandwich up/wrapped up Don's sandwich) in the cafe.

76 The editor (did Jerry's script over/did over Jerry's script) before the meeting.
He was resentful that she (was controlling/did/did it).
The editor (did Jerry's script over/did over Jerry's script) before the meeting.

77 The barista (mixed Peter's order up/mixed up Peter's order) yesterday morning.
He was frustrated that she (was wrong/did/did it).
The barista (mixed Peter's order up/mixed up Peter's order) yesterday morning.

78 The linguist (put Todd's book away/put away Todd's book) on the shelf.
He was irritated that she (was picky/did/did it).
The linguist (put Todd's book away/put away Todd's book) on the shelf.

79 The female terrorist (blew Bob's building up/blew up Bob's building) out of spite.
He was horrified that she (attacked/did/did it).
The female terrorist (blew Bob's building up/blew up Bob's building) out of spite.

80 The painter (got Lonnie's request across/got across Lonnie's request) to the manager.
He was pleased that she (understood/did/did it).
The painter (got Lonnie's request across/got across Lonnie's request) to the manager.

81 The bank teller (took Jack's check back/took back Jack's check) at the desk.
He was angry that she (was mistrusting/did/did it).
The bank teller (took Jack's check back/took back Jack's check) at the desk.

82 The accountant (filled Dan's tax form/filled form Dan's tax) out at the office.
He was happy that she (helped/did/did it).
The accountant (filled Dan's tax form/filled form Dan's tax) out at the office.

83 The florist (made Theo's bouquet up/made up Theo's bouquet) on the spot.
He was delighted that she (was artistic/did/did it).
The florist (made Theo's bouquet up/made up Theo's bouquet) on the spot.

84 The maid (threw Dean's leftovers away/threw away Dean's leftovers) on purpose.
He was shocked that she (was spiteful/did/did it).
The maid (threw Dean's leftovers away/threw away Dean's leftovers) on purpose.

85 The young woman (backed the lawyer's car up/backed up the lawyer's car) in the parking lot.
He was nervous that she (was distracted/did/did it).
The young woman (backed the lawyer's car up/backed up the lawyer's car) in the parking lot.

86 The model (took the designer's hat off/took off the designer's hat) after the shoot.
He was hurt that she (was unimpressed/did/did it).
The model (took the designer's hat off/took off the designer's hat) after the shoot.

87 **The groundskeeper (called the man's game off/called off the man's game) because of rain.**
He was frustrated that she (cancelled/did/did it).
The groundskeeper (called the man's game off/called off the man's game) because of rain.

88 The secretary (gave the CEO's plan up/gave up the CEO's plan) to the competition.
He was shocked that she (was disobedient/did/did it).
The secretary (gave the CEO's plan up/gave up the CEO's plan) to the competition.

89 The operator (wrote the man's complaint down/wrote down the man's complaint) in a booklet.
He was satisfied that she (listened/did/did it).
The operator (wrote the man's complaint down/wrote down the man's complaint) in a booklet.

90 The saleswoman (wrapped the boy's shirt up/wrapped up the boy's shirt) in a bag.
He was pleased that she (was careful/did/did it).
The saleswoman (wrapped the boy's shirt up/wrapped up the boy's shirt) in a bag.

91 The bride (blew the groom's balloons up/blew up the groom's balloons) at the party.
He was grateful that she (volunteered/did/did it).
The bride (blew the groom's balloons up/blew up the groom's balloons) at the party.

92 The secretary (got the consultant's message across/got across the consultant's message) to the board.
He was relieved that she (understood/did/did it).
The secretary (got the consultant's message across/got across the consultant's message) to the board.

93 The waitress (took the customer's food back/took back the customer's food) to the kitchen.
He was glad that she (was understanding/did/did it).
The waitress (took the customer's food back/took back the customer's food) to the kitchen.

94 The banker (added the man's money up/added up the man's money) at the counter.
He was impatient that she (was inefficient/did/did it).
The banker (added the man's money up/added up the man's money) at the counter.

95 The nun (passed the priest's letters out/passed out the priest's letters) at the service.
He was happy that she (volunteered/did/did it).
The nun (passed the priest's letters out/passed out the priest's letters) at the service.

96 The little girl (tried the man's coat on/tried on the man's coat) in the hallway.
He was amused that she (was short/did/did it).
The little girl (tried the man's coat on/tried on the man's coat) in the hallway.

97 **Bob (picked Amanda up/picked up Amanda) at the airport.**
She was grateful that he (was on time/did/did it).
Bob (picked Amanda up/picked up Amanda) at the airport.

98 **Alan (looked Cindy up/looked up Cindy) in the directory.**
She was annoyed that he (was pushy/did/did it).
Alan (looked Cindy up/looked up Cindy) in the directory.

99 Jacob (picked Joanne out/picked out Joanne) from the big crowd.
She was appreciative that he (was attentive/did/did it).
Jacob (picked Joanne out/picked out Joanne) from the big crowd.

100 Sam (let Amanda in/let in Amanda) from the storm.
She was ecstatic that he (offered/did/did it).
Sam (let Amanda in/let in Amanda) from the storm.

101 Henry (kicked Irina out/kicked out Irina) from the store.
She was enraged that he (was angry/did/did it).

Henry (kicked Irina out/kicked out Irina) from the store.

102 Nelson (laid Sadie off/laid off Sadie) last Friday.
She was distressed that he (was insensitive/did/did it).
Nelson (laid Sadie off/laid off Sadie) last Friday.

103 David (cheered Cathy up/cheered up Cathy) at the funeral.
She was thankful that he (was supportive/did/did it).
David (cheered Cathy up/cheered up Cathy) at the funeral.

104 Nathan (figured Tina out/figured out Tina) on the date.
She was ashamed that he (was astute/did/did it).
Nathan (figured Tina out/figured out Tina) on the date.

105 Jason (set Linda up/set up Linda) with a friend.
She was excited that he (offered/did/did it).
Jason (set Linda up/set up Linda) with a friend.

106 Ivan (cleaned Rita up/cleaned up Rita) after the food fight.
She was mortified that he (helped/did/did it).
Ivan (cleaned Rita up/cleaned up Rita) after the food fight.

107 Adam (brought Christina back/brought back Christina) after the party.
She was delighted that he (was a gentleman/did/did it).
Adam (brought Christina back/brought back Christina) after the party.

108 Tom (lifted Barbara up/lifted up Barbara) to the highest shelf.
She was pleased that he (was strong/did/did it).
Tom (lifted Barbara up/lifted up Barbara) to the highest shelf.

109 George (took his new girlfriend out/took out his new girlfriend) on a date.
She was happy that he (was interested/did/did it).
George (took his new girlfriend out/took out his new girlfriend) on a date.

110 James (crossed the nurse out/crossed out the nurse) in the photograph.
She was furious that he (was immature/did/did it).
James (crossed the nurse out/crossed out the nurse) in the photograph.

111 Cal (knocked the small girl out/knocked out the small girl) during the fight.
She was infuriated that he (attacked/did/did it).
Cal (knocked the small girl out/knocked out the small girl) during the fight.

112 Oliver (cheered the tour guide up/cheered up the tour guide) on the street.
She was glad that he (was cheerful/did/did it).
Oliver (cheered the tour guide up/cheered up the tour guide) on the street.

113 John (figured the psychologist out/figured out the psychologist) at the conference.
She was impressed that he (was intelligent/did/did it).
John (figured the psychologist out/figured out the psychologist) at the conference.

114 Scott (set the musician up/set up the musician) on a date.
She was thrilled that he (was matchmaking/did/did it).
Scott (set the musician up/set up the musician) on a date.

115 Steve (sent the intern out/sent out the intern) to the interview.
She was nervous that he (was lazy/did/did it).
Steve (sent the intern out/sent out the intern) to the interview.

116 Ed (let the manager down/let down the manager) at the presentation.
She was disappointed that he (underperformed/did/did it).
Ed (let the manager down/let down the manager) at the presentation.

117 Theodore (scared the young lady off/scared off the young lady) in the bar.
She was disgusted that he (was flirtatious/did/did it).
Theodore (scared the young lady off/scared off the young lady) in the bar.

118 Stan (picked the lifeguard up/picked up the lifeguard) at the beach.
She was overjoyed that he (was generous/did/did it).
Stan (picked the lifeguard up/picked up the lifeguard) at the beach.

119 Brad (looked the artist up/looked up the artist) in the phonebook.
She was surprised that he (was assertive/did/did it).
Brad (looked the artist up/looked up the artist) in the phonebook.

120 Gary (picked Karen out/picked out Karen) for the team.
She was proud that he (was interested/did/did it).
Gary (picked Karen out/picked out Karen) for the team.

121 The gentleman (asked Sally out/asked out Sally) for dinner.
She was flattered that he (offered/did/did it).
The gentleman (asked Sally out/asked out Sally) for dinner.

122 The gunman (held Nicole up/held up Nicole) in the dark alley.
She was frightened that he (was crazy/did/did it).
The gunman (held Nicole up/held up Nicole) in the dark alley.

123 The captain (left Maria out/left out Maria) during the roll call.
She was anxious that he (forgot/did/did it).
The captain (left Maria out/left out Maria) during the roll call.

124 The butcher (dropped Laura off/dropped off Laura) at the store.
She was pleased that he (offered/did/did it).
The butcher (dropped Laura off/dropped off Laura) at the store.

125 The trader (turned Edith down/turned down Edith) without an explanation.
She was irritated that he (refused/did/did it).
The trader (turned Edith down/turned down Edith) without an explanation.

126 The surgeon (woke Lynn up/woke up Lynn) after the procedure.
She was unaware that he (was done/did/did it).
The surgeon (woke Lynn up/woke up Lynn) after the procedure.

127 The carpenter (sent Sophie out/sent out Sophie) to the cafe.
She was angry that he (was bossy/did/did it).
The carpenter (sent Sophie out/sent out Sophie) to the cafe.

128 The composer (let Tanya down/let down Tanya) at the concert.
She was sad that he (was untalented/did/did it).
The composer (let Tanya down/let down Tanya) at the concert.

129 The firefighter (scared Gwen off/scared off Gwen) from the building.
She was amused that he (was silly/did/did it).
The firefighter (scared Gwen off/scared off Gwen) from the building.

130 The lumberjack (took Danielle out/took out Danielle) for the holiday.
She was shocked that he (was polite/did/did it).

The lumberjack (took Danielle out/took out Danielle) for the holiday.

131 The philanthropist (crossed Kelsey out/crossed out Kelsey) from the list.
She was confused that he (was angry/did/did it).
The philanthropist (crossed Kelsey out/crossed out Kelsey) from the list.

132 The milkman (knocked Miriam out/knocked out Miriam) with a bottle.
She was furious that he (was violent/did/did it).
The milkman (knocked Miriam out/knocked out Miriam) with a bottle.

133 The uncle (cleaned the child up/cleaned up the child) in the bathroom.
She was amused that he (was disgusted/did/did it).
The uncle (cleaned the child up/cleaned up the child) in the bathroom.

134 The truancy officer (brought the girl back/brought back the girl) in the car.
She was upset that he (was strict/did/did it).
The truancy officer (brought the girl back/brought back the girl) in the car.

135 The bodybuilder (lifted the student up/lifted up the student) to the top of the wall.
She was overjoyed that he (was strong/did/did it).
The bodybuilder (lifted the student up/lifted up the student) to the top of the wall.

136 The engineer (dropped the nanny off/dropped off the nanny) at the house.
She was glad that he (was friendly/did/did it).
The engineer (dropped the nanny off/dropped off the nanny) at the house.

137 The judge (turned the prostitute down/turned down the prostitute) in court.
She was agitated that he (refused/did/did it).
The judge (turned the prostitute down/turned down the prostitute) in court.

138 The manager (woke the beauty queen up/woke up the beauty queen) in the morning.
She was mad that he (was loud/did/did it).
The manager (woke the beauty queen up/woke up the beauty queen) in the morning.

139 The physician (let the girl in/let in the girl) to the examining room.
She was scared that he (was mean/did/did it).
The physician (let the girl in/let in the girl) to the examining room.

140 The guard (kicked the therapist out/kicked out the therapist) from the jail.
She was perplexed that he (was angry/did/did it).
The guard (kicked the therapist out/kicked out the therapist) from the jail.

141 The boss (laid the assistant off/laid off the assistant) unexpectedly.
She was unhappy that he (was insensitive/did/did it).
The boss (laid the assistant off/laid off the assistant) unexpectedly.

142 The politician (asked the socialite out/asked out the socialite) on Saturday.
She was repulsed that he (was interested/did/did it).
The politician (asked the socialite out/asked out the socialite) on Saturday.

143 The robber (held the clerk up/held up the clerk) in the store.
She was concerned that he (was dangerous/did/did it).
The robber (held the clerk up/held up the clerk) in the store.

144 The inventor (left the woman out/left out the woman) in the patent.
She was displeased that he (was dishonest/did/did it).
The inventor (left the woman out/left out the woman) in the patent.

145 **Ben (blew Jenny's balloons up/blew up Jenny's balloons) at the party.**
She was thankful that he (helped/did/did it).
Ben (blew Jenny's balloons up/blew up Jenny's balloons) at the party.

146 **Sam (got Allison's idea across/got across Allison's idea) to the committee.**
She was pleased that he (understood/did/did it).
Sam (got Allison's idea across/got across Allison's idea) to the committee.

147 Ken (took Laura's book back/took back Laura's book) to the library.
She was indifferent that he (was nice/did/did it).
Ken (took Laura's book back/took back Laura's book) to the library.

148 Louis (did Diana's work over/did over Diana's work) before class.
She was grateful that he (was smart/did/did it).
Louis (did Diana's work over/did over Diana's work) before class.

149 Mike (mixed Michelle's instructions up/mixed up Michelle's instructions) by accident.
She was peeved that he (was careless/did/did it).
Mike (mixed Michelle's instructions up/mixed up Michelle's instructions) by accident.

150 Harold (put Polly's clothes away/put away Polly's clothes) in the dresser.
She was relieved that he (was tidy/did/did it).
Harold (put Polly's clothes away/put away Polly's clothes) in the dresser.

151 Sidney (gave Nina's location up/gave up Nina's location) to the police.
She was furious that he (was disloyal/did/did it).
Sidney (gave Nina's location up/gave up Nina's location) to the police.

152 Andrew (wrote Patricia's number down/wrote down Patricia's number) on his hand.
She was excited that he (was interested/did/did it).
Andrew (wrote Patricia's number down/wrote down Patricia's number) on his hand.

153 Justin (wrapped Fran's leftovers up/wrapped up Fran's leftovers) in a container.
She was irritated that he (was picky/did/did it).
Justin (wrapped Fran's leftovers up/wrapped up Fran's leftovers) in a container.

154 Frank (filled Dana's application out/filled out Dana's application) at the store.
She was pleased that he (was thorough/did/did it).
Frank (filled Dana's application out/filled out Dana's application) at the store.

155 Wesley (made Brenda's speech up/made up Brenda's speech) before the ceremony.
She was happy that he (was creative/did/did it).
Wesley (made Brenda's speech up/made up Brenda's speech) before the ceremony.

156 Martin (threw Esther's book away/threw away Esther's book) by accident.
She was dismayed that he (was careless/did/did it).
Martin (threw Esther's book away/threw away Esther's book) by accident.

157 **John (added the spinster's money up/added up the spinster's money) at the table.**
She was relieved that he (was honest/did/did it).
John (added the spinster's money up/added up the spinster's money) at the table.

158 **Henry (passed the teacher's instructions out/passed out the teacher's instructions) in the classroom.**
She was delighted that he (was helpful/did/did it).
Henry (passed the teacher's instructions out/passed out the teacher's instructions) in the classroom.

159 Kyle (tried the girl's boots on/tried on the girl's boots) in the hallway.
She was irritated that he (bothered/did/did it).

Kyle (tried the girl's boots on/tried on the girl's boots) in the hallway.

160 Reggie (gave the princess's dog up/gave up the princess's dog) to the pound.
She was heartbroken that he (was thoughtless/did/did it).
Reggie (gave the princess's dog up/gave up the princess's dog) to the pound.

161 Ron (wrote the doctor's orders down/wrote down the doctor's orders) with a pen.
She was satisfied that he (was listening/did/did it).
Ron (wrote the doctor's orders down/wrote down the doctor's orders) with a pen.

162 Greg (wrapped the scientist's order up/wrapped up the scientist's order) with string.
She was impatient that he (was slow/did/did it).
Greg (wrapped the scientist's order up/wrapped up the scientist's order) with string.

163 Bill (backed the landscaper's truck up/backed up the landscaper's truck) onto the lawn.
She was irritated that he (was careless/did/did it).
Bill (backed the landscaper's truck up/backed up the landscaper's truck) onto the lawn.

164 Don (took the tennis player's jacket off/took off the tennis player's jacket) before the match.
She was happy that he (was helpful/did/did it).
Don (took the tennis player's jacket off/took off the tennis player's jacket) before the match.

165 Ned (called the photographer's exhibit off/called off the photographer's exhibit) at the last minute.
She was disappointed that he (cancelled/did/did it).
Ned (called the photographer's exhibit off/called off the photographer's exhibit) at the last minute.

166 Ryan (blew the woman's balloons up/blew up the woman's balloons) at the party.
She was appreciative that he (assisted/did/did it).
Ryan (blew the woman's balloons up/blew up the woman's balloons) at the party.

167 Max (got the economist's idea across/got across the economist's idea) on the radio.
She was excited that he (understood/did/did it).
Max (got the economist's idea across/got across the economist's idea) on the radio.

168 Alex (took the girl's gift back/took back the girl's gift) to the store.
She was glad that he (was honest/did/did it).
Alex (took the girl's gift back/took back the girl's gift) to the store.

169 The electrician (turned Susan's electricity off/turned off Susan's electricity) for the week.
She was angry that he (was unsympathetic/did/did it).
The electrician (turned Susan's electricity off/turned off Susan's electricity) for the week.

170 The architect (drew Marcia's plan out/drew out Marcia's plan) using the computer.
She was excited that he (was creative/did/did it).
The architect (drew Marcia's plan out/drew out Marcia's plan) using the computer.

171 The pool-boy (turned Jill's lawnmower on/turned on Jill's lawnmower) in the driveway.
She was surprised that he (was not lazy/did/did it).
The pool-boy (turned Jill's lawnmower on/turned on Jill's lawnmower) in the driveway.

172 The lawyer (put Hannah's record on/put on Hannah's record) in the living room.
She was embarrassed that he (liked Abba/did/did it).
The lawyer (put Hannah's record on/put on Hannah's record) in the living room.

173 The veterinarian (gave Nancy's cat back/gave back Nancy's cat) last Thursday.
She was overjoyed that he (was a good vet/did/did it).
The veterinarian (gave Nancy's cat back/gave back Nancy's cat) last Thursday.

174 The organizer (checked Maya's name off/checked off Maya's name) on the list.
She was impatient that he (was disorganized/did/did it).
The organizer (checked Maya's name off/checked off Maya's name) on the list.

175 The programmer (backed Liz's computer up/backed up Liz's computer) yesterday.
She was relieved that he (was in charge/did/did it).
The programmer (backed Liz's computer up/backed up Liz's computer) yesterday.

176 The designer (took Tessa's scarf off/took off Tessa's scarf) in the dressing room.
She was irritated that he (was unimpressed/did/did it).
The designer (took Tessa's scarf off/took off Tessa's scarf) in the dressing room.

177 The captain (called Flora's trip off/called off Flora's trip) because of the storm.
She was sad that he (backed out/did/did it).
The captain (called Flora's trip off/called off Flora's trip) because of the storm.

178 The auctioneer (added Linda's purchases up/added up Linda's purchases) at the register.
She was indifferent that he (was careful/did/did it).
The auctioneer (added Linda's purchases up/added up Linda's purchases) at the register.

179 The counselor (passed Gail's cookies out/passed out Gail's cookies) at the meeting.
She was proud that he (was socializing/did/did it).
The counselor (passed Gail's cookies out/passed out Gail's cookies) at the meeting.

180 The athlete (tried Bella's shoes on/tried on Bella's shoes) at the track.
She was unaware that he (attempted/did/did it).
The athlete (tried Bella's shoes on/tried on Bella's shoes) at the track.

181 The man (filled the saleswoman's form out/filled out the saleswoman's form) at the dealership.
She was satisfied that he (was thorough/did/did it).
The man (filled the saleswoman's form out/filled out the saleswoman's form) at the dealership.

182 The father (made the daughter's excuse up/made up the daughter's excuse) before class.
She was unappreciative that he (lied/did/did it).
The father (made the daughter's excuse up/made up the daughter's excuse) before class.

183 The grandfather (threw the girl's doll away/threw away the girl's doll) in the trash.
She was heartbroken that he (was careless/did/did it).
The grandfather (threw the girl's doll away/threw away the girl's doll) in the trash.

184 The technician (put the secretary's coat on/put on the secretary's coat) by accident.
She was amused that he (was confused/did/did it).
The technician (put the secretary's coat on/put on the secretary's coat) by accident.

185 The thief (gave the woman's money back/gave back the woman's money) at the trial.
She was glad that he (was atoning/did/did it).
The thief (gave the woman's money back/gave back the woman's money) at the trial.

186 The professor (checked the girl's exam off/checked off the girl's exam) in the office.
She was relieved that he (was impressed/did/did it).
The professor (checked the girl's exam off/checked off the girl's exam) in the office.

187 The policeman (did the clerk's reports over/did over the clerk's reports) on the computer.
She was frustrated that he (intervened/did/did it).
The policeman (did the clerk's reports over/did over the clerk's reports) on the computer.

188 The drummer (mixed the singer's music up/mixed up the singer's music) at the show.
She was exasperated that he (was forgetful/did/did it).

The drummer (mixed the singer's music up/mixed up the singer's music) at the show.

189 The tutor (put the girl's homework away/put away the girl's homework) after the session.
She was thankful that he (was patient/did/did it).
The tutor (put the girl's homework away/put away the girl's homework) after the session.

190 The landlord (turned the tenant's hot water off/turned off the tenant's hot water) after the fight.
She was outraged that he (was a jerk/did/did it).
The landlord (turned the tenant's hot water off/turned off the tenant's hot water) after the fight.

191 The artist (drew the musician's portrait out/drew out the musician's portrait) in a notebook.
She was flustered that he (was attentive/did/did it).
The artist (drew the musician's portrait out/drew out the musician's portrait) in a notebook.

192 The contractor (turned the hairdresser's power on/turned on the hairdresser's power) in the shop.
She was grateful that he (was helping/did/did it).
The contractor (turned the hairdresser's power on/turned on the hairdresser's power) in the shop.

Appendix D: Fillers for particle shift experiments. Y/N after each filler indicates the correct answer to the verification sentence. All particle shift fillers were used in all particle shift experiments.

1 Jack asked Lindsay out to the dance.
She was nervous that he was interested.
She was nervous that he was interested. Y

2 Paul added the girl's money up at the gate.
She was glad that he was smart.
She was glad that Paul was smart. Y

3 The policeman brought Kate back after the interrogation.
She was angry that he was mean.
She was angry that he was mean. Y

4 The woman backed the man's car up into the spot.
He was relieved that she was helpful.
The man was relieved that she was helpful. Y

5 Larissa blew Kevin's air mattress up on the floor.
He was happy that she was hospitable.
He was happy that she was hospitable. Y

6 Gary cheered up the teacher after class.
She was happy that he was nice.
She was happy that Gary was nice. Y

7 The principal called off Jenny's class unexpectedly.
She was upset that he was careless.
She was upset that he was careless. Y

8 The nanny cleaned up the boy in the tub.
He was mad that she was bossy.
He was mad that the nanny was bossy. Y

9 Susan crossed out George from the list.
He was upset that she was unfriendly.
He was upset that she was unfriendly. Y

10 Trudy checked off the traveler's documents at the counter.
He was glad that she was quick.
The traveler was glad that Trudy was quick. Y

11 The deliveryman dropped Rachel off at work.
She was pleased that he did.
She was pleased that he did. Y

12 The technician did the woman's repair over in the shop.
She was glad that he did.
She was happy that he did. Y

13 Nancy drew Richard's portrait out with a pencil.
He was flattered that she did.
He was flattered that she did. Y

14 Karen figured the liar out very quickly.

He was embarrassed that she did.
The liar was embarrassed. Y

15 The secretary filled Bob's form out during the meeting.
He was thankful that she did.
He was thankful that she did. Y

16 Frank held up Amanda during the earthquake.
She was relieved that he did.
Amanda was relieved. Y

17 Ben kicked out Diana from the bar.
She was angry that he did.
She was angry that he did. Y

18 Daniel got across the librarian's idea to the committee.
She was satisfied that he did.
The librarian was satisfied that he did. Y

19 The nurse knocked out Larry by accident.
He was furious that she did.
He was furious that she did. Y

20 The babysitter gave back the boy's toy after the nap.
He was overjoyed that she did.
He was really happy that she did. Y

21 Henry gave Sally's cat up to the shelter.
She was upset that he did it.
She was upset that he did it. Y

22 Gus laid the intern off last Friday.
She was devastated that he did it.
She was devastated that Gus did it. Y

23 The lawyer made Francine's defense up before the trial.
She was thankful that he did it.
She was thankful that he did it. Y

24 The mother left the son out at the game.
He was sad that she did it.
The son was sad that she did it. Y

25 Lucy let Gregory in through the door.
He was thankful that she did it.
He was thankful that she did it. Y

26 Ryan mixed up the manicurist's mail by accident.
She was irritated that he did it.
She was irritated that he did it. Y

27 The fireman let down Ruth with a rope.
She was grateful that he did it.
Ruth was grateful that the fireman did it. Y

28 The baker passed out the man's recipe to the crowd.
He was proud that she did it.
He was proud that she did it. Y

29 Julie put on James,Äò scarf before going outside.
He was annoyed that she did it.
He was annoyed that Julie did it. Y

30 Melissa lifted up the child above the railing.
He was excited that she did it.
The child was excited. Y

31 Michael put Lily's papers away in the drawer.
She was mad that he was cleaning.
Lily put Michael's papers away in the drawer. N

32 Hank looked the salesgirl up on facebook.
She was flattered that he was interested.
Hank looked up the salesgirl in the phonebook. N

33 The driver picked Lydia up at the train station.
She was happy that he was on time.
The driver dropped Lydia off at the train station. N

34 The woman took the boy's toy back to the store.
He was devastated that she was so mean.
The woman took back the girl's toy to the store. N

35 Sarah picked James out from the crowd.
He was relieved that she was observant.
James picked Sarah out from the crowd. N

36 Lawrence took off the director's coat in the foyer.
She was mad that he was an imposter.
Lawrence took the director's scarf off in the foyer. N

37 The mailman threw away Brenda's mail by accident.
She was livid that he was careless.
The mailman threw away Brenda's mail on purpose. N

38 The hairdresser scared off the client before the appointment.
He was frightened that she was inexperienced.
The client scared the hairdresser off before the appointment. N

39 Arielle tried on Jose's skates near the pond.
He was amused that she was trying.
Jose tried on Arielle's skate near the pond. N

40 Danielle sent out the intern to get lunch.
He was annoyed that she was bossy.
Danielle sent the intern out to get coffee. N

41 The artist set Jennifer up on a date.
She was excited that he did.
The artist set Bob up on a date. N

42 The plumber turned the homemaker's water off after the flood.
She was unhappy that he did.
The landlord turned off the homemaker's water. N

43 Tammy took Ed out to a fancy restaurant.

He was happy that she did.
Ed took Tammy out to a fancy restaurant. N

44 Rebecca turned the client's computer on at the table.
He was irritated that she did.
Rebecca turned on the client's computer at the desk. N

45 The saleswoman wrapped Jeff's purchase up in paper.
He was glad that she did.
The saleswoman wrapped Larry's purchase up in paper. N

46 The boater turned down the woman out of spite.
She was sad that he did.
The woman turned the boater down out of spite. N

47 Luke wrote down Sally's address in a notebook.
She was happy that he did.
Luke wrote down Sally's number in a notebook. N

48 Robert woke up the student in class.
She was embarrassed that he did.
Robert woke the student up in church. N

49 The florist asked out Rick to a party.
He was pleased that she did.
The dancer asked out Rick to a party. N

50 The waitress added up the customer's bill at the register.
He was impatient that she did.
The waitress added the customer's bill up at the table. N

51 Terrence brought Marie back to the hospital.
She was annoyed that he did it.
Terrence brought Marie back to the school. N

52 Andy backed the nurse's car up into a pole.
She was furious that he did it.
Andy backed up the nurse's bicycle into a pole. N

53 The assistant blew Cher's house up with dynamite.
She was livid that he did it.
The assistant blew Cher's house up with gasoline. N

54 The maid cheered up the father after dinner.
He was grateful that she did it.
The butler cheered the father up after dinner. N

55 Kim called Brian's game off yesterday.
He was displeased that she did it.
Kim called Larry's game off yesterday. N

56 Ronald cleaned up the librarian after the foodfight.
She was grateful that he did it.
Ronald cleaned the secretary up after the foodfight. N

57 The driver crossed out Donna from the schedule.
She was displeased that he did it.
The driver crossed out Donna from the schedule. N

58 The worker checked off the woman's orders on the list.
She was delighted that he did it.
The worker checked the man's orders off on the list. N

59 Cynthia dropped off Walter at school.
He was unhappy that she did it.
Cynthia picked up Walter from school. N

60 Martha did over the accountant's reports after the audit.
He was unaware that she did it.
Martha did the banker's reports over after the audit. N

61 Doug drew Amy's plan out on a sketchpad.
She was happy that he was artistic.
She was jealous that he was artistic. N

62 Jerry figured the flight attendant out during the announcement.
She was happy that he was listening.
She was annoyed that he was listening. N

63 The engineer filled Virginia's documents out at the site.
She was relieved that he was organized.
She was nervous that he was organized. N

64 The stenographer held the witness' evidence up in the trial.
He was embarrassed that she was loud.
He was embarrassed that she was misbehaving. N

65 Heather kicked Raymond out of the club.
He was angry that she was intolerant.
She was angry that he was intolerant. N

66 Charles got across the landscaper's idea to the client.
She was thankful that he understood.
She was thankful that Charles misunderstood. N

67 The drummer knocked out Carolyn with a drumstick.
She was upset that he was violent.
She was upset that he was friendly. N

68 The girl gave back the boy's toy after the fight.
He was glad that she repented.
He was mad that she was mean. N

69 Mildred gave up Eric's secret at the meeting.
He was shocked that she was disloyal.
He was shocked that she was loyal. N

70 Shirley laid off the statistician unexpectedly.
He was sad that she was unsympathetic.
He was glad that she was sympathetic. N

71 The mayor made Deborah's testimony up at the hearing.
She was surprised that he did.
She was mad that he did. N

72 The counselor left the girl out of the trip.

She was dismayed that he did.
She was pleased that he did. N

73 Janet let Tim in from the blizzard.
He was shocked that she did.
She was shocked that he did. N

74 Andrea mixed the critic's column up in the paper.
He was angry that she did.
He was glad that she did. N

75 The therapist let Patrick down in the session.
He was disappointed that she did.
He was disappointed that the teacher did. N

76 The motorcyclist passed out the activist's paper to the group.
She was unaware that he did.
She was aware that he did. N

77 Roy put on Rose's dress for fun.
She was disgusted that he did.
She was disgusted that he put on her shoes. N

78 Keith lifted up the dancer in the show.
She was proud that he did.
She was angry that he did. N

79 The student put away Al's data in the binder.
He was confused that she did.
She was confused that he did. N

80 The writer looked up the man on Google.
He was surprised that she did.
He was upset that she did. N

81 Nick picked Irene up at the airport.
She was shocked that he did it.
He was shocked that she did. N

82 Aaron took the babysitter's movie back to the store.
She was pleased that he did it.
She was unhappy that he did it. N

83 The vendor picked Janice's hot dog out from the fryer.
She was ambivalent that he did it.
Janice was ambivalent that the waiter did it. N

84 The stylist took off the actor's jacket in the photo shoot.
He was mad that she did it.
The actress was mad that she did it. N

85 Jane threw Don's shoes away in the trash.
He was upset that she did it.
He was happy that she did it. N

86 Edwin scared off the Jehovah's witness at the door.
She was indignant that he did it.
She was proud that he did it. N

87 The male model tried on the designer's suit backstage.
She was impatient that he did it.
She was impatient that she did it. N

88 The cartoonist sent out the man's draft to the editor.
He was nervous that she did it.
He was nervous that he did it. N

89 Ashley set up Leon on a date.
He was unaware that she did it.
He was aware that Ashley did it. N

90 Bonnie turned off Spencer's heat in the summer.
He was frustrated that she did it.
He was frustrated that Linda did it. N