Toward an account of accented pronoun interpretation in discourse context: Evidence from eye-tracking

Jennifer J. Venditti[⋆], Matthew Stone[†], Preetham Nanda[†], Paul Tepper[†]

[⋆]University of Pennsylvania Institute for Research in Cognitive Science
[†]Rutgers Center for Cognitive Science and Dept. of Computer Science

December 21, 2001

Abstract

Pronouns uttered with intonational prominence (i.e. 'pitch accent') are interpreted differently from those uttered without such prominence. Studies have shown that an accented pronoun will shift attention from the most salient entity in the discourse context to some other less salient entity. For example, in the spoken utterance *John hit Bill and then HE hit George*, listeners agree that the accented *HE* refers to the less salient *Bill*. While this judgment has been discussed numerous times in the literature, the majority of previous studies have relied on introspective or off-line judgments, and have focused on interpretation in strictly parallel clausal sequences. This paper reports the results from an eye-tracking study of on-line interpretation of nuclear-accented (subject) pronouns in differing discourse contexts. We present data suggesting that (i) the type of inferred discourse coherence relation, and (ii) the ability to locally resolve the presupposition of contrast evoked by the accent influence the interpretation of accented pronouns. In addition, our data tell us something about the time-course of incremental interpretation of utterances with accented subject pronouns. We find that both potential antecedents are evoked immediately upon hearing the accented pronoun. A preference for one referent over the other only emerges once subsequent propositional information is encountered which lends support for the inferred discourse relation.

Contents

1 Introduction

A number of studies in both theoretical and computational linguistics have observed that pronouns uttered with intonational prominence (i.e. 'pitch accent') are interpreted differently from those uttered without such prominence. For example, consider the now famous example about John and Bill's aggressions, given in (1).

(1) a. John hit Bill and then he hit George. (*he* = John)
b. John hit Bill and then HE hit George. (*HE* = Bill) [1] [2]

In these examples, the gender information conveyed by the pronoun is consistent with either of the two male referents in the previous clause. However, despite this potential ambiguity, native intuitions are unambiguous: the pronoun uttered without an accent (1a) is interpreted as referring to *John*, while the accented pronoun (1b) is taken to refer to *Bill*.

The current study takes an experimental approach to documenting this intuition. Specifically, we present data from a psycholinguistic experiment using eye-tracking which examines the on-line and off-line interpretation of accented and unaccented pronouns in discourse context. The main research questions addressed by this study are: Is the information provided by the accent interpreted on-line, as listeners parse a spoken utterance? Does the discourse context affect interpretation preferences? Is eye-tracking a valid methodology for investigating these questions?

Our data suggest that accent alone is not sufficient to switch reference to a less salient entity. Rather, we find that (i) the type of inferred discourse coherence relation, and (ii) the ability to locally resolve the presupposition of contrast evoked by the accent influence the interpretation of accented pronouns. In addition, our data tell us something about the time-course of incremental interpretation of utterances with accented subject pronouns. We find that both potential antecedents are evoked immediately upon hearing the accented pronoun. A preference for one referent over the other only emerges once subsequent propositional information is encountered which lends support for the inferred discourse relation.

1.1 Constraints on accented pronoun interpretation

Anaphoric expressions such as pronouns must be interpreted with reference to the previous discourse context. The antecedent of a anaphor is generally some salient entity mentioned recently in the discourse. A great number of linguistic theories have proposed mechanisms by which such reference resolution can effortlessly occur in language comprehension. However, a majority of these studies have investigated off-line judgments, and have focused on pronoun interpretation in written language, or on unaccented pronouns in spoken language. In this study, we focus our attention on the on-line interpretation of intonationally prominent, or 'accented', pronouns in spoken discourse.

Early research on accented pronouns relied on introspective judgments about pronoun interpretation in sequences of conjoined clauses, such as those shown in (1a) and (1b) above (e.g. [Gleit61, AJ70, Lak71], etc.). The general conclusion from these studies is summarized by Akmajian and Jackendoff's observation that "contrastive stress on either a pronoun or noun will prohibit coreference" [AJ70, p. 124]. In other words, given some heuristic which determines which entity in a discourse model is likely to be the antecedent of (i.e. 'coreferent with') a pronominal, putting an accent on that pronominal will cue that the antecedent is in fact some 'other' entity. This observation as stated reflects our intuition that an intonationally prominent pronoun refers to an entity which is not the one that we would 'expect', though it does not specify details of how to go about determining which referent in the discourse this 'other guy' is.

1.1.1 The parallel function strategy

Following up on this early hypothesis, Solan and others suggested that unaccented pronoun interpretation is driven by a heuristic called the *parallel function strategy*, and that "contrastively stressing the pronoun in a

[1] Following the usual convention in theoretical linguistics, capitalization will be used in examples here as a shorthand for indicating that the word bears a prominent accent. What type of accent this represents is an open question which we are currently investigating.

[2] This example of accent on a subject pronoun is taken from Oehrle's [Oeh81] extension of Akmajian and Jackendoff's [AJ70] original examples in which the accent occurs on an object pronoun.

sentence has the effect of undermining the parallel function strategy ... [that is, it has] the effect of shifting preferred antecedents" [Sol83, p. 163]. The *parallel function strategy* is a general heuristic first proposed by Sheldon [Shel74], by which listeners/readers interpret an unaccented pronoun to be coreferent with the entity which was mentioned in the same grammatical position (e.g. subject, object, etc.) in the previous clause. Given this strategy, Solan's claim is that placing intonational prominence on the pronoun results in an interpretation in which the pronoun now refers to an entity which is in another grammatical position (see also [Smyth92, Smyth94]). As Solan notes, "stressing a pronoun informs the hearer that the speaker intends its antecedent to be something unexpected" [Sol84, p. 176]. Here, what is 'expected' is defined by the parallel function strategy.

1.1.2 Determining which referent is 'expected' or 'salient'

According to the parallel function strategy, an 'expected' referent is that which is in the same grammatical position as the pronoun in question. Recent studies in theoretical and computational linguistics have proposed an alternative means to define which referents are 'expected' in a discourse, or more importantly here, which referents are 'unexpected' antecedents of pronominal forms. Attention-driven studies of discourse coherence and pronoun resolution have proposed that entities in the discourse model are ranked according to their attentional 'salience', and that this ranking determines the 'expectedness' for coreference with a subsequent pronominal (see, for example, the proposals associated with Centering Theory, e.g. [GJW95, WJP98]). In this approach, the set of salient entities which are potential referents of a pronominal form in clause U_i is defined as those entities which are realized in the immediately preceding clause U_{i-1}.[3] We will call this set of referents the 'salient subset' (aka. the 'forward-looking center list' in Centering Theory). The salience ranking of the entities in this salient subset is often determined by their grammatical position in the clause, with subjects being more salient than objects, etc. [Chafe76, WJP98]. Given this type of ranking (and an underlying preference for discourses to be coherent), a subject in U_{i-1} will be a more 'expected' antecedent of a pronominal in U_i than would be an object.[4] Note that in the case of pronominals in subject position in U_i, as in (1) above, the predictions of the parallel function strategy and the salience ranking/coherence hypothesis are identical: the preferred antecedent of an unaccented subject pronoun will be the grammatical subject realized in the preceding clause.[5]

This simple salience ranking (accompanied by the preference for discourse coherence via center continuation) can explain the 'default' interpretation of a great majority of written pronouns and unaccented pronouns in spoken language. However, these default preferences do not hold when the pronoun receives intonational prominence, as described above. To account for these cases, there have been a number of proposals within the attention-driven framework. Among the proposals, Cahn states that "when a pitch accent is applied to a pronominal, its main effect is attentional, on the order of items in [the salient subset]". [Cahn95, p. 291] (see also [Cahn]). Nakatani suggests that accented pronouns mark a "shift in attention away from the current discourse center to a new discourse entity that was indeed salient in the immediate discourse context" [Naka93, p. 166] (see also [Naka97a, Naka97b]). Terken remarks that "accented pronouns ... signal that the intended entity is not the most accessible entity at that moment. That is, we expect the occurrence of an accented pronoun in situations where the pronoun violates the prominence ranking." [Terk93]. And finally, Kameyama proposes that "a focused pronoun takes the complementary preference of the unstressed counterpart" [Kame99, p. 315]. What all of these proposals have in common is that a candidate set of currently salient entities is defined, and is ranked according to relative salience or acces-

[3] We use the shorthand 'clause' here to denote the unit of structure over which the set of salient entities is defined and updated. However, there has been some debate about what the exact nature of this unit should be: a sentence? a tensed clause? etc. (see [Kame98, Milt] and also Section 3.4.5 below).

[4] This parenthetical about coherence is included here because in the most prominent attention-driven analysis of discourse interpretation, Centering Theory, pronoun interpretation is based on two interacting constraints: (i) the ranking of the salient subset, and (ii) the preference for a discourse to be coherent. That is, readers/listeners prefer to continue centering the same discourse entity across utterances pairs. Therefore, resolution depends on more than just the salience ranking.

[5] This study investigates the interpretation of pronominals in subject position only. In the case of object pronominals, such as *John hit Bill and then George hit him*, the predictions of the parallel function strategy and the salience ranking/coherence hypothesis differ: the 'expected' antecedent of the object pronoun is the previous object according to the parallel function strategy, but is the previous subject according to the attention-driven hypothesis. We are currently conducting experiments which examine the on-line interpretation of accented and unaccented object pronominals.

sibility (in practice, grammatical role). The preferred antecedent of an accented pronoun is some salient entity contained within this set, but crucially is not the most salient entity (according to the default ranking). This is essentially the same claim made by Akmajian and Jackendoff [AJ70], Solan [Sol83, Sol84], Smyth [Smyth92, Smyth94] and others, though couched in a different framework. In the attention-driven approaches, the definition of the salient subset and the ranking of entities within it is made explicit.

1.1.3 Kameyama's account of accented pronoun interpretation

To clarify the predictions made by the attention-based theories, we will briefly outline Kameyama's [Kame99] account of accented pronoun interpretation. This is perhaps the most explicit and well-documented of the current proposals (and is generally representative of the assumptions about accented pronoun interpretation made by [Cahn, Cahn95, Terk93, Naka93, Naka97a, Naka97b], though not necessarily in the exact details).

Kameyama claims that accented pronoun interpretation results from an interaction of the semantic focus associated with the pitch accent, and the definition and ranking of the salient subset used for interpreting unaccented pronouns. She states that "a stressed *HE* presupposes a constraint ∼F that there is a contextually determined set of entities ($[[HE]]^f = \{x \mid x \in F \subseteq E\}$ where E is the domain of individuals) with at least two members — the denotation of *HE* ($[[HE]]^o$) and at least one more contrasting individual" [Kame99, p. 308]. This constraint is due to the semantic focus interpretation of the (narrow focus) pitch accent itself, as described by Rooth [Rooth92]. Kameyama implicitly assumes that the type of intonational prominence on a 'stressed pronoun' is the same as a narrow focus, or 'contrastive', pitch accent.[6] This hypothesis is consistent with the characterization of accented pronouns in the previous literature as 'contrastively stressed' (e.g. [AJ70, Sol83, Smyth94]).[7]

The "contextually determined set of entities" which Kameyama refers to here is functionally defined as the set of salient entities in the local attentional state (i.e. the entities realized in the immediately preceding clause). This salient subset is the same for both accented and unaccented pronouns — the only difference being the relative salience ranking within the set. She proposes that accented pronoun interpretation is driven by a preference ranking which is 'complementary' to the ranking for unaccented pronouns. This means that the salience ranking of possible antecedents for accented pronouns in equivalent to the *reverse* of the ranking used for interpretation of unaccented pronouns. Kameyama suggests that accented pronoun interpretation proceeds via the sequence of computations described below (paraphrased here, see [Kame99, p. 315] for full details). Consider again the example in (2).

(2) John hit Bill. Then HE ...

- Determine the salient subset based on the local attentional state of U_{i-1}:
  $\{$*John*, *Bill*$\}$
- Determine the salience ranking for the unaccented pronoun ('default') case:
  $\{$*John* $>$ *Bill*$\}$
- Compute complementary preference (i.e. re-rank):
  $\{$*Bill* $>$ *John*$\}$
- Discharge the presupposed constraint of contrast ∼C for the utterance U_i.

By this account, interpretation involves re-ranking of entities in the salient subset, and choosing the most salient entity of that newly ranked set as the preferred antecedent of the accented pronoun.[8] Note that if the initial salience ranking is based only upon grammatical role, such an account predicts the incorrect interpretation of an accented (or unaccented) pronoun in object position in parallel structures such as *John*

[6] The function and distribution of narrow focus pitch accents has been discussed extensively in the intonation literature: see Rooth [Rooth92], Bolinger (e.g. [Bol61]), Ladd (e.g. [Ladd80]), among many *many* others.

[7] Note however that in other recent studies, there is some debate about what the exact nature of this intonational prominence is. Cahn suggests that the pitch accent must be L+H* [Cahn, Cahn95], while Nakatani claims that the shift in preferred interpretation occurs with H* accents as well (e.g. [Naka97b]). We are currently conducting experiments which examine this question in detail.

[8] Note that this proposal predicts that the lowest-ranked entities in the default order of salience will become the highest-ranked entities in the re-ranked set. This suggests that the preferred antecedent of the accented pronoun in a sequence like *John hit Bill using the bat owned by Sam. Then HE ...* will be *Sam*. This prediction has yet to be empirically tested.

hit Bill then George hit him/HIM. That is, the previous subject should be the preferred antecedent in the unaccented case, while the previous object should be preferred in the accented case. Both of these predictions result in an incorrect interpretation. Pronoun interpretation in strictly parallel sequences is a general problem encountered by approaches which consider only grammatical role in determining salience ranking (e.g. most of the implementations of Centering Theory). To account for this, Kameyama has proposed an additional *property sharing constraint*, similar to the parallel function strategy, which she claims comes into play in the (default) salience ranking step (see [Kame86, Kame99]).[9] We will return to discussions of pronoun interpretation in strictly parallel vs. non-parallel sequences in Sections 3.4.3, 3.4.4 and 4 below.

It is important to ask at this point whether Kameyama's proposed computations for accented pronoun interpretation can be used in a psycholinguistic model of on-line interpretation. Kameyama is careful to note that that "no sequential order is assumed" among the computations [Kame99, p. 308], but we can assume that some steps do precede others — for example, the salience ranking probably does precede the discharging of the presupposition ∼C for U_i. How might this process work on-line? One can imagine that upon parsing the clause U_{n-1}, listeners will have in memory the salient entities which were just encountered, and they may even be able to form a hypothesis about the salience ranking of these entities at this point, based on (at least) knowledge about grammatical roles. Then, upon hearing the discourse connective *then* and the pronoun *HE* in the following clause U_i, listeners may be cued to initiate the complementary preference computation. This may occur rapidly as the accent is perceived. The final step in interpretation is to discharge the presupposed constraint of contrast for the utterance U_i. If we follow Kameyama's description, this step necessarily cannot be computed immediately after the accented pronoun is perceived. This is because the presupposed constraint which is discharged indicates that there is "a contextually determined set of *propositions* [our emphasis] obtained by instantiating a set abstraction with the alternative values of the focused element" [Kame99, p. 308]. While the 'alternative values' may be immediately available (after parsing the previous clause and the accented *HE*), the proposition carried by the target clause U_i may not be known in full until the whole clause is parsed. That is, since the contrast set is obtained by instantiating $\{$*hit George*(*x*) $\mid$ $x \in F\}$ with the alternative values of $[[HE]]^f \in F$ (see [Kame99, p. 308] for full details)[10], one must first know what the proposition carried by U_i is. Therefore, this step, as worded by Kameyama, must take place after U_i has been parsed. Of course, whether or not a listener can build on-line an incremental hypothesis of the propositional content of an utterance is one of the ultimate questions in sentence processing (see e.g. [TT95] for a review). If listeners can glean information (albeit incomplete) about propositional content on-line, then it is highly likely that they may also be able to proceed on-line with the discharging of the presupposed constraint of contrast which Kameyama describes. In such a case, the contrast set would be computed based on incomplete information and may be ultimately wrong and subject to subsequent revision. We will return to this issue of the incremental nature of accented pronoun interpretation in the discussion in Section 4.

1.1.4 When accented pronouns do not shift attention

Although a majority of previous accounts describe accented pronouns as shifting the center of attention to a less salient entity in the discourse context, there are many cases in which the attention is not shifted. Instead, the accent serves to cue a contrast between the salient entity and some other unspecified set of entities (e.g. see discussions in [Prev95, Prev96]). Consider the example given in (3).

(3) Jack is a physicist. HE ...
[Kame99, p. 317]

Here, native speakers unambiguously interpret the accented pronoun as referring to *Jack* and not some 'other guy', even when *Jack* is the most salient entity in the current discourse context. These sorts of examples are clear counterexamples to the many claims that accented pronouns shift attention away from the most salient entity. How then can we resolve this apparent contradiction?

[9] But see the discussion in Section 4 below of Kehler's [Kehl01] unified account of discourse coherence and pronoun interpretation in both parallel and non-parallel sequences.

[10] For this example, this computation would result in the set of propositions $\{$*John hit George, Bill hit George*$\}$.

Kameyama suggests that these cases can also be accounted for by the same mechanisms used in interpreting accented pronouns which shift attention. The key is the size of the salient subset of discourse entities. Kameyama proposes that these cases (of no shift) occur when there is only one salient entity in the local attentional state. Given this, their interpretation falls out from the proposed steps of interpretation outlined above. That is, if the salient subset contains only one entity, then re-ranking the set results in that single entity remaining the most salient, and thus the preferred antecedent of the accented pronoun. In this way, Kameyama predicts that the single salient entity will be the preferred antecedent of both an unaccented pronoun as well as an accented pronoun. She notes that in the accented pronoun interpretation, there is an additional presupposition of contrast: in cases where the salient subset is a singleton, "at least one contrasting individual is accommodated" when this constraint is discharged [Kame99, p. 315]. However, the exact nature of this accommodation remains unspecified.

Kameyama's account unifies the apparent discrepancy in interpretation of the two contrasting classes of accented pronouns: those in which the attention is shifted to a less-salient entity, and those in which the attention is not shifted. In this paper, we will restrict our focus to cases in which the attention is shifted. That is, we will examine cases in which there is more than one salient entity in the immediate context.

1.2 Building a discourse context

Most of the early descriptions of accented pronouns describe intuitions about coreference only in very parallel clause sequences like *John hit Bill and then HE hit George*, presented in isolation (e.g. [Gleit61, AJ70, Lak71, Oeh81, Sol83, Sol84], and also more recently by [Smyth94, BST98]). In more recent work on pronoun resolution within the context of computational linguistics and artificial intelligence, researchers have generalized the use and interpretation of accented pronouns beyond strictly parallel structures (e.g. [Cahn, Naka93, Terk93, Cahn95, Prev95, Prev96, Naka97a, Naka97b, Kame99]). In Section 1.1.3 above, we outlined the details of one such general account, in which pronoun interpretation is determined by the relative salience ranking of entities in the current discourse context (in conjunction with the preference for coherence across utterance pairs). In the sections that follow, we test this hypothesis empirically by examining the on-line and off-line interpretation of unaccented and accented pronouns in structures which are not strictly syntactically parallel.

In constructing a discourse context for our experimental stimuli, we were faced early-on with the crucial question: In what discourse contexts is the use of an accented pronoun felicitous? Based on the discussions provided by [Rooth92, Prev95, Prev96, Kame99] and others, our working assumption is that accented pronouns are felicitous in contexts in which there is a basis for contrast among members of a set of salient entities.[11] In order to create such a context, we constructed narratives in which the discourse participants are collaborating on a joint action. The discourse provides information about what each participant contributes to the joint goal. The progression of the discourse is driven by the open questions under discussion, also known as *QUDs* (see e.g. [Rob96]), and nature of the QUDs is what sets up the basis for contrast. Example (4) shows the QUD structure of one of the discourse stimuli used in this experiment. This structure is representative of all the test stimuli (see Set 1 in the Appendix for all discourses).

(4) ⇒ *QUD: What happened?*
The zebra and the pig wanted to wash the car together.

⇒ *QUDs: What did the zebra contribute? What did the pig contribute?*
The zebra put a bucket of soapy water next to the pig near the front of the car.
(answers *What did the zebra contribute?*)

⇒ *QUDs: What else did the zebra contribute? What did the pig contribute?*
Then he/HE got out some sponges.
('*he*' answers *What else did the zebra contribute?*)
('*HE*' answers *What did the pig contribute?*)
...

[11] We are currently investigating this question further through analyses of large-scale corpora.

At the beginning of the discourse, the open question is the general *What happened?*. The introduction of the two discourse participants (*the zebra* and *the pig*) and the joint goal (*washing the car together*) in sentence 1 motivates the subsequent QUDs: *What did the zebra contribute?* and *What did the pig contribute?*. Sentence 2 proceeds to answer *What did the zebra contribute?*, thereby removing it from the QUD list. At this point in the discourse, we are left with the open QUD *What did the pig contribute?*, but have also added another possibility: *What else did the zebra contribute?*. Sentence 3 could answer either of these questions. If the subject pronoun in sentence 3 is unaccented, it is taken to answer *What else did the zebra contribute?*, while if it is accented, it answers *What did the pig contribute?*. That is, the salience ranking/coherence constraints drive the interpretation of the unaccented *he* — the listener assumes that the QUD about zebra's contribution is still being answered. In contrast, the accented *HE* cues the listener that next QUD (about the pig's contribution) is to be addressed. The experimental stimuli were constructed based on this type of QUD structure. The discourse in (5) shows the example from (4), with QUDs removed. All stimuli are listed in the Appendix.[12]

(5) 1. The zebra and the pig wanted to wash the car together.
2. The zebra put a bucket of soapy water next to the pig near the front of the car.
3a. Then he got out some sponges.
3b. Then HE got out some sponges.
4. And together they started washing the hood and the fenders.

In the current study, we examine listener preferences about which QUD will be answered by the target sentence 3. We also document the incremental on-line interpretation of both the accented and unaccented pronouns, as well as off-line judgments. We now turn to a detailed discussion of those experiments.

2 Experiment 1: Determining potential biases

2.1 Motivation

The purpose of Experiment 1 was two-fold. First, we wanted to experimentally determine whether there are any biases in our visual and audio stimuli which would cause listeners to prefer one character over the other in their interpretation of the pronoun in target sentence 3. The (visual and auditory) stimuli were designed with the intention that either character would be equally plausible as do-er of the target action, and this norming experiment assessed our success in doing so.[13]

Another main motivation for Experiment 1 was to investigate whether the intentional structure we set up for the discourses, defined here by the open questions under discussion (QUDs), affects listeners' preferences for who will be the do-er of the target action. That is, in the discourse given in (4) above, if listeners entertain the QUD *What else did the zebra contribute?* after hearing sentence 2, they will prefer sentence 3 to describe the action of the *zebra*. If, on the other hand, listeners entertain the QUD *What did the pig contribute?*, preferences will be for the *pig* to be doing the action. If either QUD is equally plausible, preferences should be mixed. In this experiment, we examined listener preferences for agent of the target action described by sentence 3. We included discourses like that in (4), and also included ones in which the QUD structure was slightly different. Section 2.2.1 describes the discourses in detail.

2.2 Materials

2.2.1 Discourses

All discourses describe a joint collaborative action between two cartoon animals. Two variants of a 3-sentence discourse were constructed, as shown in (6).

[12] The test stimuli are listed in Set 1, and the fillers are listed in the other sets.

[13] Thanks to Mark Steedman for suggesting this experiment.

(6) 1. The zebra and the pig wanted to wash the car together.
2a. The zebra put a bucket of soapy water next to the pig near the front of the car.
2b. The zebra told the pig to put a bucket of soapy water near the front of the car.
3. Then someone got out some sponges.

N1 (*the zebra*) and N2 (*the pig*) are introduced as a conjoined NP in the first sentence. N1 remains the subject of the second sentence, but the do-er of the action is varied: in (6.2a) N1 does the action, while in (6.2b) N2 does the action. In the third sentence, the identity of the actor is unspecified. We chose to use *someone* instead of the pronominal *he* to refer to the agent in order to avoid preferences based on discourse coherence strategies which would result from using a pronoun. Since the open questions under discussion after hearing sentence 2 are necessarily different depending on the variant, we suspect that this may influence listener judgments. The different QUDs are shown in (7).[14]

(7) 2a. The zebra put a bucket of soapy water next to the pig near the front of the car.
⇒ *QUDs: What else did the zebra contribute? What did the pig contribute?*

2b. The zebra told the pig to put a bucket of soapy water near the front of the car.
⇒ *QUDs: What else did the pig contribute? What did the zebra contribute?*

2.2.2 Auditory stimuli

All utterances in the experiments reported here were recorded by the first author using a Shure SM10A unidirectional head-mounted microphone and a TASCAM TEAC PA-1 portable DAT recorder. The utterances were recorded at 48KHz then transferred to UNIX/Linux workstations and downsampled to 16KHz for analysis and playback. Acoustic analysis was performed using Entropic Research Labs ESPS/Waves+ software, version 5.3.1.

Care was taken to utter the discourses in a uniform yet natural manner in order to minimize acoustic and prosodic variability. Because of the range of text material used in the discourses (see Appendix), the prosodic structures necessarily were not identical. However, certain relevant prosodic features were intentionally kept constant: the noun phrases referring to N1 and N2 received pitch accents in sentence 1 as well as in sentence 2. Sentence 3 was uttered with a 'hat-pattern' intonation, with a pre-nuclear H* pitch accent on *someone*.[15]

2.2.3 Visual stimuli

All visual stimuli used in the experiments reported here were generated in Abode Photoshop 6.0 using animal characters hand-drawn by Paul Tepper and clip art available on the internet.

Each stimulus shows a scene containing animals and objects which are involved in the actions described by the discourse. Figure 1 gives an example of the visual stimulus paired with the discourse in (6) above. In each scene, the two characters involved in the joint activity are located diagonal to one another and equidistant from the object mentioned at the end of sentence 2 (here, the *car*), which is immediately previous to the mention of *someone* in the target sentence 3. The relative positioning (i.e. left-right, top-bottom) of these characters was balanced across items. In addition, the object described by the action in the target sentence (here, the *sponges*) is located beside each of the characters. Such placement of the characters and target object was intended to prevent any bias due to one character being closer to the target object, or bias due to one character consistently being in a particular region of the scene. However, the placement of the object of the action in sentence 2 (here, the *bucket*) may in fact produce a bias toward N2 in the discourses. In sentence 2, N1 places (or tells N2 to place) the object in the vicinity of N2. The auditory

[14] See Set 1 in the Appendix for a list of all discourses and visual stimuli used in Experiment 1. The visual stimuli were exactly as shown. The discourse stimuli had the following structure: sentence 1 introduced N1 and N2 as a conjoined NP subject, followed by the introduction of the collaborative action in the predicate. Sentence 2 occurred in both (a) 'doing' and (b) 'telling' variants. In sentence 3, *someone* was used instead of *he*. Sentence 4 was omitted in Experiment 1. This structure is the same as the example given in (6).

[15] See [BE94] for a full description of the ToBI-style intonation notation used in this paper.

Figure 1: Example of the visual stimulus paired with the discourse in (6).

stimulus describes this action, and the visual stimulus shows the object in its resulting location near N2. It is possible that this positioning may cause listeners to prefer N2 to be the do-er of the target action in sentence 3, since N2 is closest to the location that the last action occurred. That is, while we intended for the auditory and visual stimuli to be unbiased toward either of the two characters, this positioning of the object in sentence 2 may produce a slight bias toward N2 in the target sentence. However, this prediction would hold for both type (a) 'doing' and (b) 'telling' variants.[16]

2.3 Methods

2.3.1 Subjects

Forty students from Rutgers University participated in exchange for course credit. All subjects were either English mono-linguals or English-dominant bi-lingual speakers. All reported normal or corrected vision and normal hearing.

2.3.2 Experiment design and procedure

Sixteen discourse-scene pairs were used in Experiment 1 (see Appendix Set 1 for discourse and scene content, and example (6) above for discourse structure). In addition, 16 filler pairs were also included (see Sets 2 & 3 in the Appendix for content and (6) for discourse structure), resulting in 32 stimuli in total. Test and filler stimuli were presented in a fixed random order. The content of sentence 2 (N1 action (2a) vs. N2 action (2b)) was counterbalanced across two lists, and both lists were presented in both ascending and descending order.

Visual scenes and auditory discourse stimuli were presented using the psycholinguistic experimentation software DMDX, running on a PC desktop.[17] The manner of presentation was as follows: first, the scene

[16] Further details of the placement of the characters and objects are more relevant to the on-line eye-tracking study, and will be described below in Section 3.2.3.

[17] DMDX is the Windows version of DMASTR, authored by Kenneth Forster and Jonathan Forster at University of Arizona Psychology.

was presented on the display (800x600 pixel, 16 bit resolution). Approximately 3 seconds after onset of this visual display, the 3-sentence discourse was presented over external speakers.[18] Sequential sentences in the discourse were separated by approximately 750ms of silence. Subjects were instructed to follow along in the picture while listening. After hearing the third sentence containing the subject *someone* and the associated action, subjects indicated their preference for the do-er of this action on a separate answer sheet, and then pressed a key to continue to the next item.

The answer sheet contained three choices for each item: (1) the name of the character positioned on the left side of the scene (e.g. *the pig*), (2) the name of the character positioned on the right side of the scene (e.g. *the zebra*), and (3) the word 'either'. Subjects were instructed to circle the name of the character whom they preferred to be the do-er of the action described in sentence 3, or to circle 'either' if either character could have plausibly done the action. They were instructed to base their answer on both the picture as well as on the story. In addition, they were told to rely on their first instinct, even if they were unsure. One practice trial was given, and subjects had a chance to ask questions after completing the practice. A detailed debriefing was given upon completion of the experiment.

2.3.3 Data coding and analysis

Subject responses were logged by hand using the following scheme: a score of +1 was given for a N1-as-doer preference, a score of -1 for a N2-as-doer preference, and a score of 0 was given for an 'either' response. The item condition (either N1 (2a) or N2 (2b) action in sentence 2) was also logged.

2.4 Results

If all discourses and visual stimuli were indeed completely unbiased, we would predict that responses would be either all 'either' (score=0), or an even mix of N1-as-doer and N2-as-doer preferences (also resulting in score=0). If the placement of the object in sentence 2 (the *bucket*) in the vicinity of N2 produced a bias toward N2 as the agent of the subsequent target sentence 3, we predict that responses should favor N2 (i.e. a negative score) in both the 'doing' (2a) and 'telling' (2b) conditions. Figure 2 shows scores for the 16 test stimuli, arranged by item.

There is a clear main effect of intentional structure (i.e. open question under discussion) on subject responses.[19] In discourses in which sentence 2 describes what N1 contributed to the joint action, subjects prefer the subsequent target sentence 3 to describe N2's contribution. In contrast, if sentence 2 describes what N2 contributed to the joint action, then subjects prefer the following sentence to describe N1's contribution. That is, in the 'doing' condition (N1 act, filled circles), subjects take sentence 2 to be the answer to the open question *What did N1 contribute to the joint action?* and the target sentence 3 to be the answer to the outstanding question *What did N2 contribute to the joint action?*. The exact reverse is true for the 'telling' condition (N2 act, hollow circles). A number of subjects reported during the debriefing session that, since N1 and N2 were 'cooperating' to achieve a specific goal (e.g. here, the goal of *washing the car together*), they expected the contribution of both to be described, thus resulting in the preference for 'the other guy' to be the do-er of the target action.

These results are inconsistent with both predictions outlined at the beginning of this section. Instead of the stimuli being completely unbiased (score=0), or biased toward N2 (negative score), results show that the question under discussion (QUD) structure greatly influences listeners' preferences in determining the do-er of the target action. In Experiment 2 we will examine whether these preferences also hold when the pronoun *he* is used in place of *someone* in the target sentence, and whether listeners' eye movements reflect their preferences on-line.[20]

[18] In all experiments reported here the measure of silent inter-sentence interval durations is approximate. This is because the DMDX script used for presentation measured silent intervals in terms of machine 'ticks'. The refresh-rate varied from 14 to 16ms depending on the session (due to accidental resetting after machine reboots), so a rate of 15ms is used for calculation of absolute millisecond durations for purposes of descriptions presented here.

[19] We have not yet done a chi-square analysis on the data, so we will just rely on clear patterns in the data in our discussion here.

[20] It is possible that instead of a bias based solely on the salience of one QUD over another (as suggested here), listener preferences may be due to the QUDs interacting with the indefiniteness of *someone*. That is, instead of a general preference for *What did N2 contribute?* to be more salient than *What else did N1 contribute?* after hearing (6.2a), it is possible that the indefiniteness of *someone*

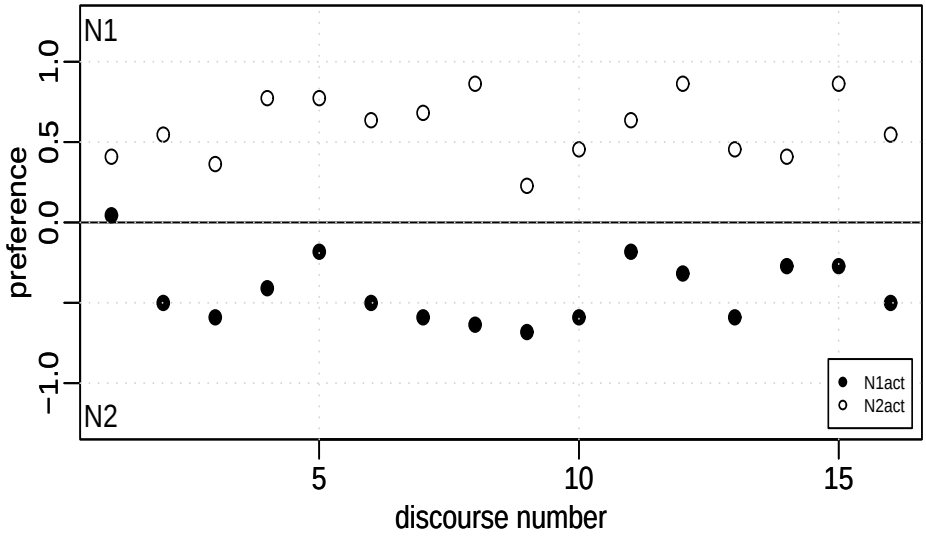

Figure 2: Preferences for agent of the target sentence 3: *Then someone did ....* Two conditions are plotted: the 'doing' condition (see (6.2a)) in which the pre-target sentence describes an N1 action (filled circles), and the 'telling' condition (see (6.2b)) in which the pre-target sentence describes an N2 action (hollow circles).

3 Experiment 2: Tracking eye fixations on-line in discourse

3.1 Motivation

In one of the earliest studies of spoken language comprehension using eye-tracking, Cooper observed that subjects' eye movements to images are time-locked to information in an auditorily presented story [Coop74]. He found that subjects would fixate an object in a visual display shortly after the object was mentioned in a spoken narrative. The probability of fixating objects which were semantically related to the narrative was higher than in a control condition, in which the objects were unrelated to the story line. Arnold and colleagues also examined eye fixations during the on-line comprehension of discourse [AEBST00]. Specifically, they examined the use of gender information and accessibility in the interpretation of (unaccented) gender-ambiguous and unambiguous pronouns. They also found that subjects' fixations on characters in a visual display are closely related in time to the mention of a pronominal referring expression in the discourse. These studies suggest that eye-tracking is indeed a useful methodology for tracking listeners' on-line comprehension in spoken discourse. Given these previous findings, the purpose of Experiment 2 was to document the time-course of the interpretation of both accented and unaccented pronouns in our discourse stimuli.

3.2 Materials

3.2.1 Discourses

To simplify the experimental design, the discourses used as test stimuli in this experiment were all of the 'doing' type, in which sentence 2 described N1's contribution to the joint goal. The example in (8)

requires that the QUD being answered refers to the contribution of the non-'centered' entity (here N2). Therefore, this effect of *someone* may have caused the interpretation preferences shown in Figure 2. Possibly a better way to test general biases based on QUD structure would be to ask subjects to complete the target sentence (*Then ...*), or fill in a missing subject (*Then __ got out some sponges*), without resorting to the use of indefinite *someone* or definite *he*. Thanks to members of the Fall 2001 CUNY Psycholinguistics Supper Club for discussion of this point.

shows the structure of the discourses used as test stimuli in Experiment 2.[21] In addition to the test stimuli, this experiment contained filler discourse-scene pairs which were a mixture of the 'telling' type, parallel structures, and discourse 'digressions'. Data from these will be presented separately in Sections 3.4.2–3.4.5 below.[22]

(8) 1. The zebra and the pig wanted to wash the car together.
2. The zebra put a bucket of soapy water next to the pig near the front of the car.
3a. Then he got out some sponges. ('unaccented pronoun' = un)
3b. Then HE got out some sponges. ('nuclear-accented pronoun' = nuc)
3c. Then the zebra got out some sponges. ('N1 full NP' = N1)
3d. Then the pig got out some sponges. ('N2 full NP' = N2)
4. And together they started washing the hood and the fenders.

In the test discourses, sentence 1 introduces both characters as a conjoined NP and describes a joint action which is their goal in the discourse. Sentence 2 describes N1's contribution to this joint action, using a full NP subject to refer to N1 (here, *the zebra*). Then, the target sentence 3 describes a subsequent action, which can potentially be interpreted as a continued description of N1's contribution, or a shift to mention N2's contribution to the joint action. Results from Experiment 1 suggest that listeners prefer N2 to be the do-er of the action in sentence 3 when the indefinite *someone* is used as subject. However, Experiment 2 uses a pronoun to refer to the do-er. Such a referring expression carries with it different presuppositions than the indefinite *someone*: it signals coreference to an entity mentioned in the previous discourse. Attention-driven theories of pronoun resolution and discourse coherence described in Section 1.1 above predict that an unaccented pronoun will refer to the most salient entity in the previous utterance, namely N1 (the *zebra*). This results in contrasting predictions about who listeners will prefer as the do-er of the action in sentence 3: if the agent is described by the indefinite *someone*, N2-as-doer is preferred (see Experiment 1). On the other hand, if the agent is described by an unaccented pronoun, then N1-as-doer is preferred.

In addition to the condition in which the pronominal subject is unaccented (8.3a), this experiment also includes a condition in which the pronominal subject is uttered with a nuclear accent (8.3b).[23] As described in Section 1.1, previous accounts of pronoun interpretation in discourse predict that an accented pronoun will refer to a salient entity in the previous utterance, but not to the most salient entity. In our discourse stimuli, both N1 and N2 are mentioned in sentence 2: N1 as subject (most salient) and N2 as object of a preposition (less salient). Therefore, the antecedent of the accented pronoun in the target sentence 3 is predicted to be N2 (see Kameyama's account described in detail in Section 1.1.3). In the accented pronoun condition, the predictions based on biases due to the QUD structure (see Experiment 1) and those based on the salience ranking/coherence accounts are identical: both predict that N2 will be interpreted as the antecedent of the accented pronoun.

Two other conditions are included in this experiment as controls: a full NP referring either to N1 (8.3c) or N2 (8.3d) are included in order to compare the fixation patterns on full N1 vs. *he*, and full N2 vs. *HE*.

3.2.2 Auditory stimuli

All utterances for Experiment 2 were recorded using the same methods described in Section 2.2.2 above. Certain prosodic features were also controlled in this experiment: the noun phrases referring to N1 and N2 received pitch accents in sentence 1 as well as in sentence 2, as in Experiment 1.[24] The target sentence 3 was uttered using the intonation tunes shown below.[25]

[21] See the Appendix for a list of the exact content and structure of the discourses and visual stimuli used in Experiment 2: Set 1 lists the test stimuli and Sets 2–5 list the various types of fillers used.

[22] Data from the discourse 'digressions' subset (Appendix Set 4) will not be described in this report. We will return to this type in future experiments and manuscripts.

[23] As mentioned above, we take descriptions of 'contrastively stressed' pronouns as indicating that they bear a narrow focus, or 'nuclear', pitch accent. Whether or not this is an appropriate characterization is of course still a matter of some debate, which we are currently investigating experimentally.

[24] In fact, the recordings of sentences 1 & 2 used in Experiment 2 were the very same ones used in Experiment 1.

[25] Note that the entire tune of each condition is described here using ToBI notation [BE94]. This contrasts with the all too common practice of only marking selected prominent accents with capitalization (e.g. the *HE* in 8.3b). However, full transcriptions are preferred in order to avoid ambiguity about how the remainder of the utterance is intoned.

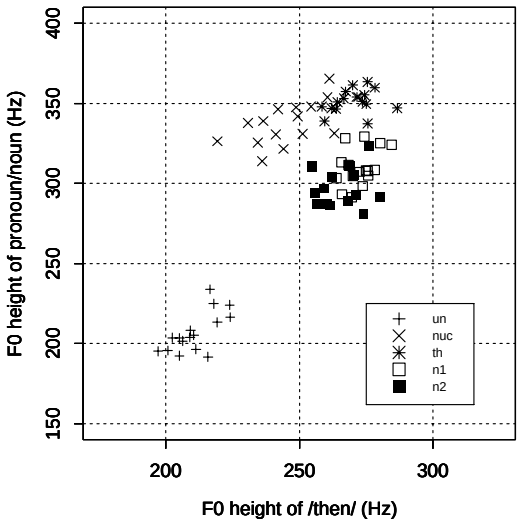

Figure 3: F0 relation of pronoun (or full NP) and *then* in the auditory stimuli used in Experiment 2. Five experimental conditions are plotted: unaccented pronoun (un), L+H*L- nuclear-accented pronoun (nuc), H*L-H% 'thematic' nuclear-accented pronoun (th, see footnote for description), H* pre-nuclear-accented full N1 (n1), and H* pre-nuclear-accented full N2 (n2).

```
3a. Then he          got out some sponges.
                     H*              H*    L- L%
3b. Then HE          got out some sponges.
       L+H* L-       H*              H*    L- L%
3c. Then the zebra   got out some sponges.
             H*    (H*)              H*    L- L%
3d. Then the pig     got out some sponges.
             H*    (H*)              H*    L- L%
```

In (3a), the pronoun *he* is unaccented, while in (3b) it bears a nuclear L+H* pitch accent (i.e. the last accent in its intermediate phrase, followed by a (L-) phrase tone). In versions (3c) and (3d), the utterances were produced using a 'hat-pattern' intonation with a pre-nuclear H* pitch accent on the subject NP.[26]

Figure 3 shows raw fundamental frequency (F0) heights of the pronoun/full NP plotted against the connective *then*, for each experimental condition. The height of the pronoun was measured at the vowel midpoint in the unaccented condition, and at the F0 peak in the nuclear-accented condition. The measurement was taken at the end of the rise in the full NP conditions. The F0 measurement of *then* was taken at the rhyme midpoint. In the unaccented condition (un), the height of both *then* and *he* are low in the speaker's range. The other conditions have relatively higher F0 during these words. The L+H* nuclear-accented condition (nuc) has a slightly higher F0 peak on *HE* and a slightly lower F0 on *then*, compared to the full NP H* conditions (n1 and n2).[27]

In addition to the discourse sentences, three 'look at' instruction sentences for each discourse were also recorded (henceforth, the 'instructions'). The instructions in (9) are an example of those matched up with the (car washing) discourse in (8). The first instruction directed the subject to fixate on the object mentioned at the end of sentence 2 in the discourse (here, the *car*). The second instruction directed the subject to then fixate on the character who did the action described by the target sentence 3. The third instruction was used as a filler.

[26]The nature of a 'hat-pattern' intonation is such that it is difficult to determine whether there are any pitch accents in the plateau which spans between the two leftmost/rightmost accents, hence the use of parentheses in the transcription of the verb *got out* in (3c) and (3d).

[27]The th condition is a H*L-H% 'thematic' nuclear accent which was used in a separate pilot experiment not discussed here.

(9) <Now look at the left headlight of the car.>
<Now look at the guy who got the sponges.>
<Now look at the guy wearing the red hat.>

The second instruction served as the off-line check of pronoun interpretation. That is, if subjects interpret N1 to be the referent of the pronoun (i.e. the do-er of the target action described in sentence 3), we expect more fixations on N1 after this instruction. If, on the other hand, subjects prefer N2 as the referent of the pronoun, we expect more fixations on N2 here. In the control conditions in which N1 or N2 are explicitly mentioned using a full NP, we expect fixations on N1 and N2, respectively.

3.2.3 Visual stimuli

The visual stimuli used in Experiment 2 were identical to those used in Experiment 1, and can be found in the Appendix. There are a few properties of the scenes relevant to eye-tracking that are important to describe here.

In constructing the visual stimuli, the scene was divided into four quadrants. The two characters involved in the joint activity were positioned in quadrants diagonal to one another, and were both equidistant from the object mentioned at the end of sentence 2 (e.g. the *car*), which was positioned in a quadrant adjacent to the characters. The purpose of such placement was to ensure that subjects were fixating this object (henceforth, the 'location') immediately prior to the mention of the target character at the beginning of sentence 3. In addition, the object which was acted upon sentence 2 (henceforth, the 'object') was consistently placed midway between the 'location' and the character referred to by N2. This placement is described by the action in sentence 2. As mentioned in Section 2.2.3 above, such placement has potential to cause a bias toward N2 as the do-er of the action in sentence 3. However, results of Experiment 1 showed that it was the QUD structure of the discourse, not this visual bias, which had the most influence on subjects' off-line preferences for the agent of the action in sentence 3.

One last constraint on the placement of objects in the visual stimuli was in the positioning of the object acted upon in the target sentence 3. Identical copies of the object were placed immediately adjacent to both N1 (henceforth, 'o1') and N2 (henceforth, 'o2'). This was to ensure that each character would have equal access to the object, and there would be no resulting bias due to one character being closer to the object than the other.[28] Also, based on a previous pilot experiment not reported here, we were concerned that subjects would not fixate the character referred to by the subject in sentence 3 (since that information is already 'old' and highly salient in the discourse), but rather would fixate only on the object itself, which is the 'new' information. Therefore, the placement of this new object next to each character provides a way to observe on-line interpretation of the pronoun, even in the absence of looks to the actual character himself. That is, we expect subjects to fixate the object next to the character whom they take to be the do-er of the target action.[29]

3.3 Methods

3.3.1 Subjects

Eight undergraduate students from Rutgers University participated in exchange for course credit. All subjects were English mono-linguals whose parents spoke only English to them at home while growing up. All reported normal or corrected vision and normal hearing.

3.3.2 Experiment design and procedure

Sixteen discourse-scene pairs were used as test stimuli in Experiment 2 (see Appendix Set 1 for the structure and content of the discourses and scenes). In addition, 20 filler pairs were also included (see Sets 2, 3, 4 & 5 in the Appendix), resulting in 36 stimuli in total. The fillers were comparable to the test stimuli in some

[28] See Section 3.4.5 below for discussion of some of the filler discourse-scene pairs in which there was only one object which was equidistant from the two characters.

[29] Thanks to Bonnie Webber for suggesting this crucial strategy.

respects, and discussion of the data collected from them will be presented in Sections 3.4.2–3.4.5 below. Test and filler stimuli were presented in a fixed random order. The intonation and form of the subject NP in sentence 3 (unaccented pronoun, nuclear-accented pronoun, pre-nuclear accented full N1, pre-nuclear-accented full N2) were counterbalanced across four lists, and all lists were presented in both ascending and descending order.

The presentation of the visual scenes and auditory discourse stimuli was the same as in Experiment 1. The scenes subtended approximately 20°x20° of visual angle.[30] For each trial, the scene was displayed for approximately 4.5 seconds, during which subjects had a chance to view the objects in the scene. Then there was an auditory prompt to "look at the cross" which was placed in the center of the scene. After a 1.5 second silent interval, the discourse was auditorily presented. Subjects were instructed to 'follow along' while listening to the discourse (though a definition of what it means to 'follow along' was not provided). Sequential sentences in the discourse were separated by approximately 750ms of silence. After hearing the final sentence of the discourse (sentence 4), there was a 2.25 second silence before the first 'look at' instruction sentence. Each of the instructions were separated by 2.25 seconds of silence.[31] When the trial was completed, subjects pressed a key to continue to the next trial. Two practice trials were given, and subjects had a chance to ask questions after completing the practice. A detailed debriefing was given upon completion of the experiment.

Eye movements were monitored using an ISCAN, Inc. head-mounted eye-tracking system. The point-of-regard (i.e. fixation location) was logged and overlaid onto the scene image, then this composite was recorded along with the simultaneous audio information onto a SONY DV-CAM digital video tape. The sampling rate of the video was 30 fps (frames per second).

3.3.3 Data coding and analysis

The digital video recordings were downloaded directly to a PC hard drive using a firewire connection. Fixation locations for each frame of the test and filler trials were hand coded by the first and third authors using Adobe Premiere 6.0 software. Objects in the scenes were assigned to categories (e.g. N1, N2, object, location, o1, o2, etc.), and coding was conducted using this categorization. The onset of a fixation of a given object was operationally defined in this study as the frame at which the saccade to that object was launched (following Cooper [Coop74] and other studies including [TSKES95, AMT98, AEBST00, TMDC00], etc.). The duration of the fixation included the duration of this saccade, as well as the duration when the point-of-regard was steady on the object. The fixation offset was defined as the frame immediately preceding the launch of a subsequent saccade outside of the current fixation region. Fixations of 2 frames (66.66ms) or greater were included in the log. That is, objects were considered 'fixated' if the point-of-regard remained on the object (or on its way to the object) for 2 frames or more. This measure is more liberal than the 3 frames (99.99ms) used in the eye-tracking studies conducted by Tanenhaus and colleagues. However, we opted for a lower threshold based on studies which report that fixation duration can be quite short if more than one saccade is programmed concurrently (e.g. [Beck91, TKH$^+$99]). We don't expect this subtle difference in coding to affect our results in any significant way. Blinks were ignored if the object fixated immediately prior to and after the blink was the same object. In cases were blinks intervened between fixation of one object and another, the blink was considered part of the following saccade, hence the onset of fixation of the following object began at the frame in which the blink began.

In addition to coding fixation locations, the acoustic onsets of each sentence in the discourse and 'look at' instructions were also hand-coded by a trained phonetician (first author) from the waveform representation of the acoustic signal displayed using Adobe Premiere 6.0. The onsets of each word within the sentences were automatically determined using Entropic Research Labs Aligner version 1.2 software, and hand-corrected by the first author. These locations (measured in milliseconds from start of speech file) were then synchronized with the sentence onset locations (measured in frames from start of video file), and were used to align the fixation data with the acoustic data.

[30] The positioning of the subjects from the display was approximately 70cm, but could not be controlled exactly.

[31] Again, all silent interval durations were approximate, due to varying refresh rate of the computer.

3.4 Results

In the following sections, we discuss a number of results from our investigation of eye fixations in on-line discourse comprehension. Section 3.4.1 discusses a linking hypothesis between eye fixations and spoken language comprehension, and asks whether listeners can 'follow along' with their eyes in spoken discourse. Section 3.4.2 examines potential effects of surface form or discourse salience on fixation probabilities. Discussions in these two sections set up the experimental context in which our results on the effect of accent can be interpreted. Section 3.4.3 then discusses the effect of accent on pronoun interpretation in the test stimuli used in our study. We provide data about both on-line and off-line interpretation preferences. Finally, Sections 3.4.4 and 3.4.5 present a brief analysis of some of the filler stimuli also included in the study: we discuss some preliminary results about the effects of syntactic parallelism, and also data from cases in which the accent appears not to switch reference.

3.4.1 Can our eyes 'follow along' with a spoken discourse?

In order for eye movement/fixation behavior to shed light on the interpretation of referring expressions in discourse, we must first formulate some hypothesis about the relation between eye movements and language comprehension in general. Tanenhaus and colleagues have proposed such a 'linking hypothesis', which relates movements to spoken language understanding: "Informally, we have automated behavioral routines link a name to its referent; when the referent is visually present and task relevant, then recognizing its name accesses these routines, triggering a saccadic eye movement to fixate the relevant information" [TMDC00, p. 565]. We will adopt this hypothesis in describing fixation behavior in the present study. One part of this linking hypothesis which is subject to question is how to interpret what is meant by 'task relevant'. In many previous studies of eye movements in language comprehension, the task was to pick up, move, or point at an object which was mentioned (e.g. [AMT98, TSKES95] and others). In such studies, the referent is necessarily task relevant. In our study, the off-line instructions directed subjects to 'look at' a visual object, which also makes the referent named by this instruction task relevant. However, the on-line portion of our study instructed subjects to simply 'follow along' while listening to the story. Can we assume that the characters and objects mentioned in the story are relevant to the task of following along?

There are two studies that we know of which have used eye-tracking to track listener comprehension in connected discourse. One is any early study by Cooper [Coop74], which examined subjects' eye movements while they listened to a short narrative passage. Cooper misinformed subjects by telling them that their pupils would be monitored as they listened to discourses, and instructed them to look anywhere they wanted within the visual display. That is, no explicit instruction was given to 'follow along' while listening. Cooper was able to identify three types of visual behavior displayed by his experimental subjects: "(1) a visual-aural interaction mode, in which fixation of targets was correlated with the meaning of concurrently heard language, (2) a free-scanning mode, in which [the subject] continually altered his direction of gaze in a manner independent of the meaning of concurrently heard language, and (3) a point-fixation mode, in which [the subject] continued to fixate the same location independent of the meaning of concurrently heard language" [Coop74, p. 102].

Arnold et al. [AEBST00] investigated the time course of interpretation of gender-ambiguous and unambiguous pronouns in connected discourse. Their experimental task was to indicate whether or not the discourse which listeners heard was consistent with the visual scene. Again, no explicit instruction was given to 'follow along', yet scanning of the scene was implicitly encouraged by the fact that subjects often had to detect minor differences between the discourse and the scene. For example, the only discrepancy between the discourse and the scene might have been the time depicted on a clock hanging on the wall in the background (J. Eisenband, personal communication). Therefore, in such a task, each referent mentioned in the discourse becomes task relevant.

In the present study, attempts were made to elicit the visual-aural interaction mode of fixation behavior observed by Cooper and Arnold et al. In order to minimize the use of the free-scanning mode during the discourse presentation, subjects were encouraged to look around and examine objects in the scene during the 4.5 second silent interval between the onset of the visual stimulus and the onset of the auditory stimulus (i.e. the narrative). In addition, subjects were explicitly told to 'follow along' while listening to the discourse.

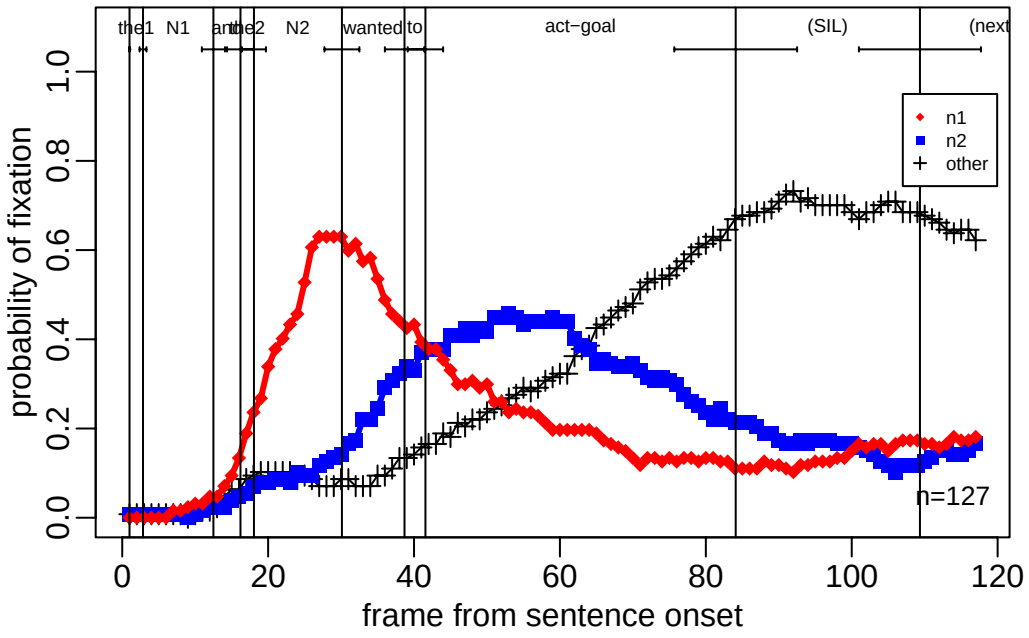

Figure 4: Probability of fixation on objects in visual display for each successive frame in sentence 1 (e.g. [*The*] $[zebra]_{N1}$ [*and*] [*the*] $[pig]_{N2}$ [*wanted*] [*to*] $[wash\ the\ car\ together]_{act-goal}$). Vertical lines mark mean (with standard deviations) onsets of words in the utterance. Red (diamond) lines plot the probability of fixation on the first-mentioned character (e.g. *zebra*) and blue (square) lines refer to the second-mentioned character (e.g. *pig*).

Figure 4 shows fixation probabilities during the auditory presentation of sentence 1: *The zebra and the pig wanted to wash the car together*. At each successive frame in sentence 1, the probability that a given object is fixated was calculated across subjects and discourse items. All plots are aligned at the sentence onset. The objects plotted are: the first-mentioned character (N1: *zebra*), the second-mentioned character (N2: *pig*), and all other objects in the scene (fixations on these objects have been collapsed for ease of presentation). At the onset of sentence 1, subjects are fixating on the cross (not plotted here). As the first utterance is presented, the probability that N1 and N2 will be fixated increases markedly after these referents are named in the discourse.[32] Since the remainder of sentence 1 was not controlled for content, no predictions are made here about which objects will be fixated.

Figure 5 shows fixation probabilities during the presentation of sentence 2: *The zebra put the bucket of soapy water next to the pig near the front of the car.* The probability of fixation on N1 (red diamonds), the object (green asterisks), and N2 (blue squares) increases substantially after each of these referents is named in the discourse. This fixation behavior observed in Figures 4 and 5 is consistent with the linking hypothesis discussed above, which states that recognizing the name of a referent results in a saccade to fixate the visual representation of that referent. It is also consistent with the behavior displayed by subjects in Arnold et al.'s study, and with the 'visual-aural interaction mode' observed by Cooper. Therefore, we conclude that, in the task of 'following along', subjects are able to make eye movements which have a meaningful and closely time-locked relation to the mention of referents in a spoken discourse.

One thing to notice in Figures 4 and 5 is that the probability of fixating N1 is greater at its peak than the probability of fixating N2, even though both referents are named in each sentence. Does this mean that

[32] A number of studies have observed an approximately 200ms delay between the speech cue and the launch of a saccade to the visual object to which the speech refers (e.g. [AMT98, TSKES95], among many others). This is attributed to the time it takes for the oculomotor system to program a saccade. Therefore, this delay should be taken into consideration when interpreting all the fixation graphs presented here.

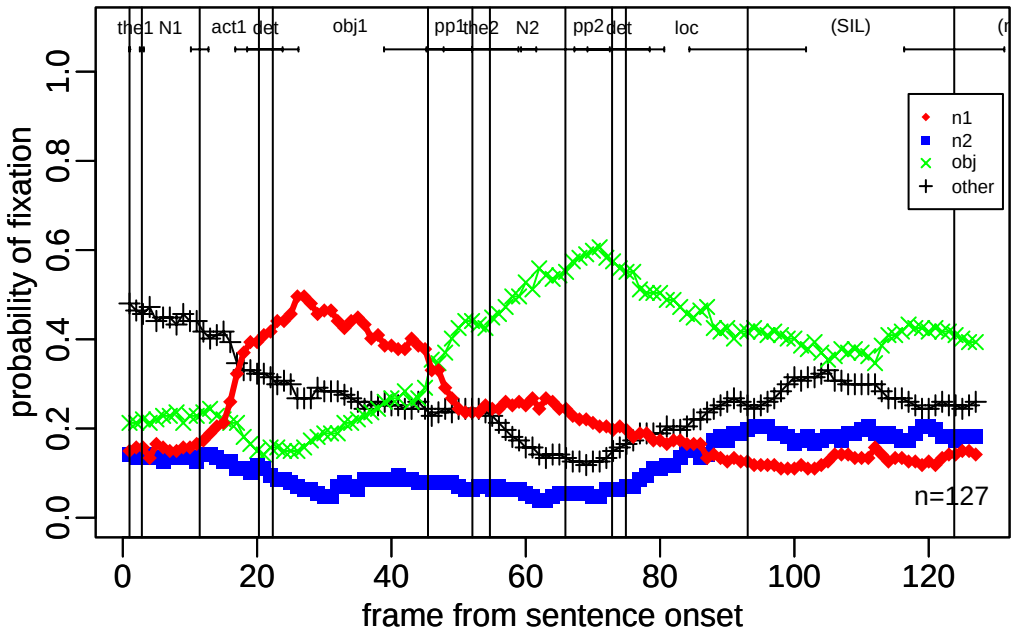

Figure 5: Probability of fixation on objects in visual display for each successive frame in sentence 2 (e.g. [*The*] $[zebra]_{N1}$ $[put]_{act1}$ $[the]_{det}$ $[bucket\ of\ soapy\ water]_{obj1}$ $[next\ to]_{pp1}$ [*the*] $[pig]_{N2}$ $[near]_{pp2}$ $[the]_{det}$ $[front\ of\ the\ car]_{loc}$). Vertical lines mark mean (with standard deviations) onsets of words in the utterance.

listeners devote less (visual) attention to a second-mentioned referent? This is an interesting open research question that unfortunately cannot be answered by this graphic representation of the data. One reason why the peak of N2 is lower than that of N1 may be due to competition from other objects in the visual scene: anticipatory (or perseveratory) looks to other objects may be more likely as the sentence progresses, thus making a first-mentioned referent have less competition than a second-mentioned referent. This may or may not reflect the distribution of actual linguistic attention in processing the utterance. Another potentially confounding factor is the cumulative misalignment of word onsets as the sentence progresses. In these figures (and in all figures shown in this paper), the utterance is aligned at the sentence onset. The probability of fixation is calculated over all discourses in a given condition, for each successive frame, with the data aligned at the onset of the utterance. Since there is a fair amount of variability in the phonetic length of constituents in each item (for example, *a bucket of soapy water* in one item vs. *a stake* in another), points further on in the sentence necessarily become misaligned when pooling data across items. The result is that the calculated probability of fixation for any given frame late in the sentence does not really correspond to a fixation relating to any particular linguistic event, as it did early on in the sentence. A better representation of the probability of fixation on N2 would be to align the data at the N2 onset and recalculate probability at each frame from this point, or to realign and measure the cumulative probability under the N2 curve after the N2 onset. The latter would take into account subjects' individual differences in delay in launching a saccade to a named object (John Trueswell, personal communication). For our purposes here, we are primarily interested in fixations on referents mentioned very early on in the target utterance (sentence 3), so our figures will continue to show alignment from sentence onset.

3.4.2 Fixation on referents of nouns vs. pronouns

In connected discourse speakers can use a full noun phrase or a reduced form such as a pronoun to refer to a discourse entity. As we discussed above, referents which are named in a spoken discourse are relevant to the task of following along, so we expect listeners to fixate each object whose referent is named. But

do pronouns carry the same weight as full noun phrases with regard to triggering a saccade to the named object?

Cooper observed a significant difference in fixations on images of referents mentioned in his discourse condition, in comparison to a control condition in which un-related images were exchanged for related images [Coop74, p. 96, Fig. 2]. However, this difference was not the same across all word-image relation category categories. Cooper found a significant difference among 'noncontextual' versus 'contextual' references. '(Direct) noncontextual' word-image relations were defined as cases in which the image was "an exact representation of the corresponding pronounced word, when this word was interpreted in isolation from the previous verbal context" (i.e. full NPs), while '(direct) contextual' word-image relations included those such as pronouns and other anaphoric expressions which require consideration of the previous context for proper interpretation [Coop74, p. 87]. Fixations on images in the (direct) noncontextual category were significantly greater than those in the (direct) contextual category. For example, an image of a lion was fixated more often upon mention of the full NP (e.g. *the lion*) than when a pronominal (e.g. *he*) was used to refer to it. This finding suggests that full NPs and pronominals may have a different probability of fixation in the present study as well.

Arnold et al.'s [AEBST00] study also examined the fixations on referents of pronouns in spoken discourse. Their experiment design was different from Cooper's in that the target utterance in each condition contained an unambiguous or ambiguous pronominal, and no comparison to a full noun phrase was included. Arnold et al. found that the probability of fixation on a given character increased dramatically around 200ms after the referent was mentioned in the discourse, even when a pronoun was used. This suggests that listeners do fixate visual information that is referred to using a reduced pronominal form. However, the design of the target utterances in Arnold et al.'s study was such that the subject was a pronominal and the predicate (the 'new' information) described some property or action attributed to the referent (Janet Eisenband, personal communication). Therefore, it is difficult to determine whether increased fixations observed shortly after uttering the pronoun were in response to the naming of the character, or in response to a search for a property or action attributed to that character. This point is not crucial to Arnold et al.'s results, but it is relevant for a hypothesis which links the perception of anaphoric expressions to eye fixations in discourses in which there may not be a visual property directly attributable to the referent, such as in our study.

The present experiment contained test stimuli such as 1 and 2 in (10) below (already described above), and similar filler stimuli such as 1' and 2'.

(10) 1. The zebra and the pig wanted to wash the car together.
2. The zebra put a bucket of soapy water next to the pig near the front of the car.

1'. The zebra asked the pig to help wash the car.
2'. He put a bucket of soapy water next to the pig near the front of the car.

Exact discourses are given in the Appendix. Sentence 1' is slightly different from sentence 1 in that N1 is a unique subject and N2 is introduced as an object. This allows felicitous reference to N1 using the pronoun *he* in 2', in contrast to sentence 2, in which a full noun phrase is necessary in this position. Fixation probabilities for 1' will not be included here, but were almost identical to those shown in Figure 4 for sentence 1. What is relevant are the fixations during sentence 2 vs. 2': since the only difference in the form of these utterances is the surface form of the NP referring to N1 (full NP vs. pronoun), we are interested in knowing if this results in different probabilities of fixating N1 in both cases.

Figure 6 shows the probability of fixations during the presentation of utterance 2': *He put a bucket of soapy water next to the pig near the front of the car*. Comparison with Figure 5 shows a marked lack of fixations on N1 (referred to by the pronoun *he* here), and also on N2 (e.g. the full NP *the pig*). The low probability of fixation on the full noun phrase N2 in this representation could be attributable to the competition and alignment factors described in Section 3.4.1 above. However, competition and alignment factors cannot account for the very low fixation probability of the pronominal N1 in Figure 6 (sentence 2'), as compared with the full noun phrase N1 in Figure 5 (sentence 2). One possible explanation for this difference lies in the relationship among utterances 1 and 2, and 1' and 2', respectively. In sentence 1, two characters are introduced as a conjoined subject, and one of them is subsequently picked out as subject

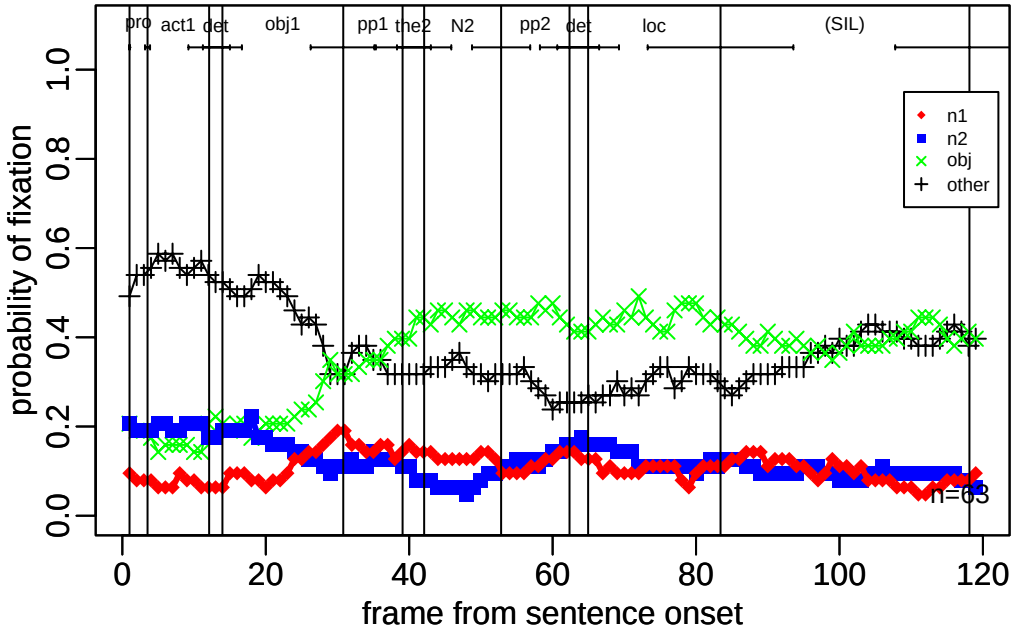

Figure 6: Probability of fixation on objects in visual display for each successive frame in sentence 2': the 'doing' type filler trials (e.g. $[He]_{pro}$ $[put]_{act1}$ $[the]_{det}$ $[bucket\ of\ soapy\ water]_{obj1}$ $[next\ to]_{pp1}$ [*the*] $[pig]_{N2}$ $[near]_{pp2}$ $[the]_{det}$ $[front\ of\ the\ car]_{loc}$). Vertical lines mark mean (with standard deviations) onsets of words in the utterance.

(N1) in the following sentence 2. That is, while N1 is already globally salient ('given') when sentence 2 is encountered, is not yet locally salient. This contrasts with 1' and 2', in which N1 is introduced as a unique subject in sentence 1', and is marked as locally salient by use of a pronominal in sentence 2'.[33] It may turn out that, in connected discourse, simply naming a referent is not sufficient for a saccade to be launched to the visual referent. Instead, fixation probabilities may be related to distinctions of global vs. local discourse salience in a systematic way. This is similar in spirit to Cooper's observations of the difference between 'noncontextual' and 'contextual' references, though framed in a more formal theory of discourse organization. The relation of fixations to differences in discourse salience is an interesting empirical question that warrants careful investigation by future studies.

Let us return briefly to the fixations on N2. As mentioned above, the low probability of fixation on the full noun phrase N2 in Figure 6 could be attributable to competition with other visual objects also mentioned (or inferred) late in the sentence, or to problems of alignment. However, if competition and alignment were the only contributing factors, the plot of N2 (blue squares) should look just like that shown in Figure 5, in which the N2 referent is also mentioned in a nearly identical position at the end of the sentence. In that graphic representation, we do observe some 'activation' of N2 late in the sentence, and this activation undoubtedly would be greater had the plots been aligned at the auditory onset of N2 instead of at the beginning of the utterance. In contrast, in Figure 6, there is a marked lack of *any* N2 activation.[34] Data from the 'telling' type of discourses, which were also included as fillers in our experiment, may shed light on this issue. Consider the utterance pair in (11).

(11) 1". The zebra asked the pig to help wash the car.
2". He told the pig to put a bucket of soapy water near the front of the car.

[33] See [GS86, GJW95] for discussions of global and local discourse salience.

[34] Note that N2 did bear a pitch accent in both the 2 and 2' variants, so a lack of prosodic salience could not explain the differences in fixation probabilities.

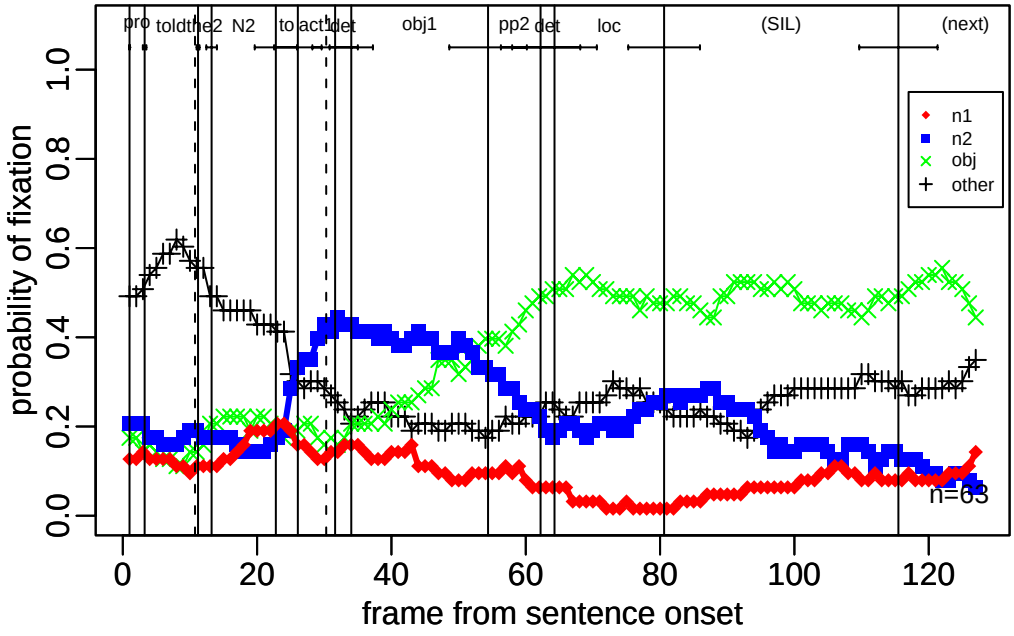

Figure 7: Probability of fixation on objects in visual display for each successive frame in sentence 2": the 'telling' type filler trials (e.g. $[He]_{pro}$ [*told*] [*the*] $[pig]_{N2}$ [*to*] $[put]_{act1}$ $[the]_{det}$ $[bucket\ of\ soapy\ water]_{obj1}$ $[near]_{pp2}$ $[the]_{det}$ $[front\ of\ the\ car]_{loc}$). Vertical lines mark mean (with standard deviations) onsets of words in the utterance.

In this utterance pair, sentence 1" is identical to 1' shown above. Sentence 2" differs from 2' in who is the do-er of the action (recall the stimuli used in Experiment 1). In sentence 2', N1 is the do-er of the action, while in 2", N2 is the do-er. How might this difference affect the probability of fixating each of the referents?

Figure 7 plots the fixation data from sentence 2": *He told the pig to put a bucket of soapy water near the front of the car*. The figure shows low probability of fixation of N1, similar to that in Figure 6. This is likely due to the fact that the use of the pronoun signals that this entity is already highly salient in the discourse. The striking difference between the two figures is the probability of fixation on N2: there are virtually no fixations on N2 in Figure 6, while there are significantly more in Figure 7. This difference cannot be due to the form of referring expression, since in both sentences N2 was uttered with a full NP. It also is probably not due to differences in prosodic salience, since N2 in 2' bears a (nuclear H*) pitch accent, and N2 in 2" bears a downstepped accent (!H*) or no pitch accent in the recorded utterances. Our prediction would be that the more 'reduced' pronunciation (i.e. a downstepped accent) would receive *fewer* looks in this discourse context. Instead, we observe significantly *more* fixations in this case. We suspect that the high probability of fixation of N2 in 2" is due to the fact that this referent is not the expected do-er of the action, as one might first assume due to the start of the utterance with reference to N1: *Then he ....* This shift of agency may be what is drawing looks to N2. This is, of course, another interesting empirical question.

In sum, our data suggest that the link between eye fixations and 'understanding' in spoken discourse is more complex than just the naming of referents. The form of the referring expression seems to play a crucial role, not just because of the differences in surface form, but because these differences are reflexes of different degrees of salience of the entity in the overall discourse. Another factor is the salience transitions between adjacent utterances: an entity which is less salient in a previous utterance may be promoted to a more salient role in the current utterance, thus attracting eye gaze. These are all issues that will be crucial to investigate further in subsequent studies of eye movements in spoken discourse understanding.

In the following section, we address another important factor: the effect of intonational tune on discourse interpretation and eye fixations.

3.4.3 Effect of accent on fixation behavior

On-line preferences

The main research question in this study is how prosodic prominence, or *pitch accent*, affects the (on-line) interpretation of pronominal forms. The preceding two sections set up the experimental context in which the accent results can be interpreted. Based on the discussion in Section 1.1, our hypothesis is that in the case of an unaccented pronoun, fixations should be centered on either N1 (the most salient character in the previous utterance), or on neither of the characters. The latter would be true if the use of a pronominal form did not draw fixations to its visual referent due to the fact that it is already highly salient in the discourse context (see Section 3.4.2).[35] In the case of the nuclear-accented pronoun, on the other hand, fixations should be directed to N2, since this is the 'unexpected' character. However, it is an empirical question whether there will be a marked decrease in fixations overall, due to the fact that the referring expression is a pronominal form. We do not expect to see the lack of fixation to the same degree as with the unaccented pronoun, since the function of the accent is to draw the listener's attention to the fact that the referent is "not the one you thought it should be". Therefore, we predict the patterns of fixation shown in Table 1 below.

referring expression	N1 fixations	N2 fixations
N1 full NP	many	none
unaccented pro	fewer/none	none
N2 full NP	none	many
accented pro	none	many

Table 1: Predicted eye fixation patterns on subject NPs in target sentence 3.

Figure 8 shows the probability of fixation during presentation of the target sentence 3: *Then the zebra/the pig/he/HE got out some sponges*. Each experimental condition is plotted separately: the full N1 subject condition is shown in the upper left, the full N2 in the lower left, the unaccented pronoun in the upper right, and the nuclear-accented pronoun condition is plotted in the lower right. The figure shows that when a full NP is used to refer to N1, the probability of fixation on N1 increases immediately after (or while) that referring expression is uttered (red diamonds). Interestingly, the probability of fixation of N1's associated object increases even before the expression referring to that object is uttered, indicating that listeners use their interpretation of who is doing the target action to anticipate which object will be acted upon (thin red line).[36] Similarly, when a full NP is used to refer to N2, we observe increased probability of fixation on N2 and his associated object, as predicted (blue squares and thin blue line, respectively).

Now, what happens when a pronoun is used? We predict that there will be fewer (relative to full N1) or no fixations on the visual referent of N1 when an unaccented pronoun is used, but that there will be a high probability of fixation on the object (the 'new' information in the utterance) which is associated with N1. This would indicate that the listener has taken the unaccented pronoun to refer to N1, even in the absence of fixations on the character himself. The upper right-hand plot in Figure 8 shows the actual fixation probabilities. We do observe an increase in fixation on N1 after the pronoun is uttered (red

[35] Remember that two (sets of) objects representing the 'new' information in the target sentence were included in the visual scene: one object was placed in the immediate vicinity of N1, and an identical object was placed in the immediate vicinity of N2. As described in Section 3.2.3, this was done so that, in the case where no fixations of a referent occur in response to a pronoun since the entity is already highly salient, we would still be able to determine which character the listener interpreted to be the antecedent of the pronoun. This of course is based on the assumption that the object which the agent will act upon will be that which is nearest to himself (i.e. N1 will act on the object closest to him, and N2 hill act on the object closest to HIM).

[36] The fact that such anticipation is so prevalent may possibly be due to the similarity among test and filler discourses in the experiment. Listeners may have somehow learned that the object in this third sentence was one of the two objects located near either character. It will be interesting to see if this anticipatory pattern holds in future experiments, in which the filler discourses really do distract from the discourse patterns of the test stimuli.

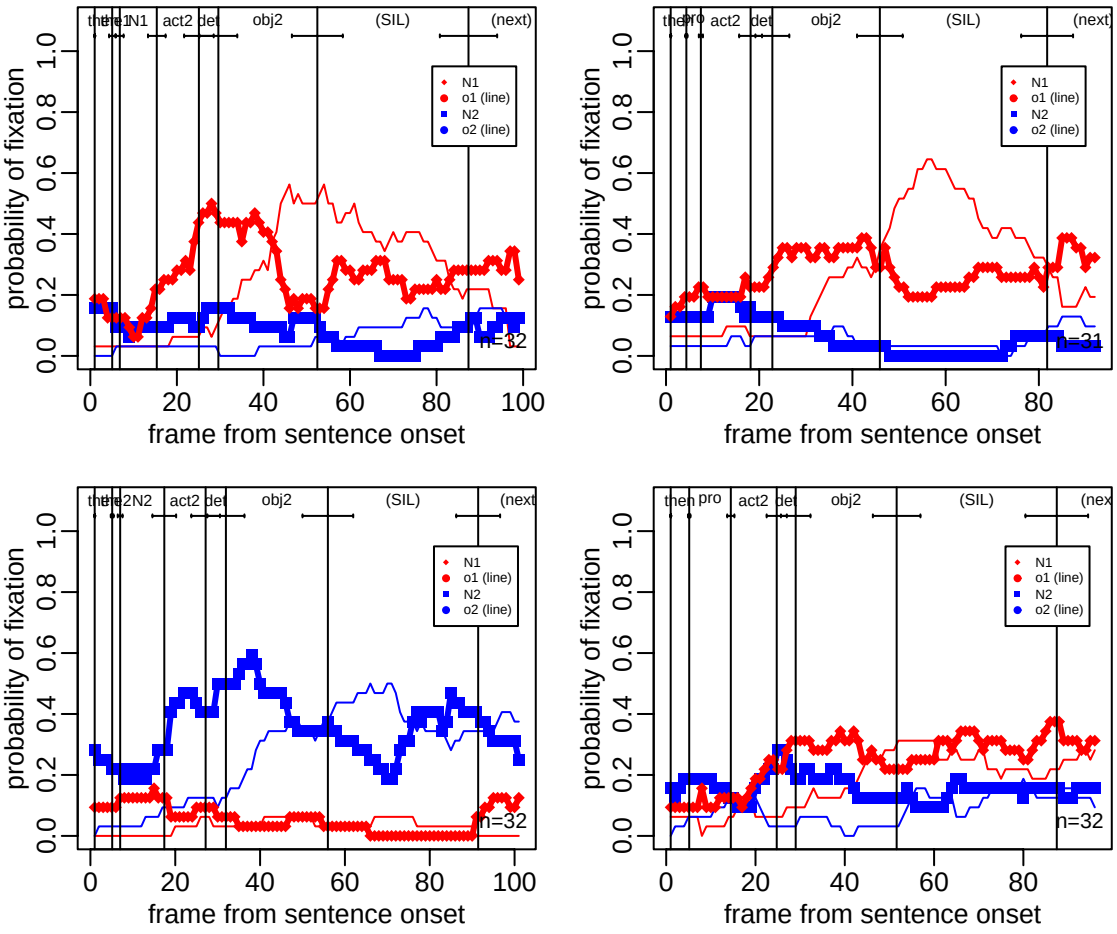

Figure 8: Probability of fixation on objects in visual display for each successive frame in sentence 3 (e.g. [*Then*] $[he]_{pro}$ $[got\ out]_{act2}$ $[some]_{det}$ $[sponges]_{obj2}$). Vertical lines mark mean (with standard deviations) onsets of words in the utterance. Each experimental condition is plotted separately: Full N1 (upper left), full N2 (lower left), unaccented pronoun (upper right), or nuclear-accented pronoun (lower right).

diamonds), and a marked increase in fixations on his associated object (thin red line). Crucially, fixations on N2 and his associated object are minimal. Determining whether or not the fixation probability of N1 in the pronoun vs. full NP conditions is significantly different would require a statistical analysis, which we have yet to conduct. An eyeball guesstimate suggests that the difference would not be significant in these data, which contrasts to the results reported in Section 3.4.2 above, in which the pronoun vs. full NP did elicit substantially different fixation patterns. At this point in time, we do not have an explanation for this difference in patterning, though it will be important to pursue in future studies.[37]

Fixation patterns in the nuclear-accented pronoun condition are markedly different. If prosodic information is not considered (on-line) in spoken language processing (which nearly all of the previous research suggests is *not* the case), then the accented pronoun plot in the lower right-hand corner of Figure 8 should exactly resemble that of the unaccented pronoun. Or, if prosodic information is used at a late stage in the interpretation process, we would predict that the initial pattern of fixations immediately after the pronoun is uttered would resemble those in the unaccented pronoun condition, and then at some (undetermined) point later in the utterance, fixations of N2 should increase substantially, relative to N1. If, on the other hand,

[37] Also note that the delay in the marked increase in fixations on the object associated with N1 (thin red line) appears to be greater in the pronoun condition than in the full N1 condition. This observation also warrants closer investigation in future studies.

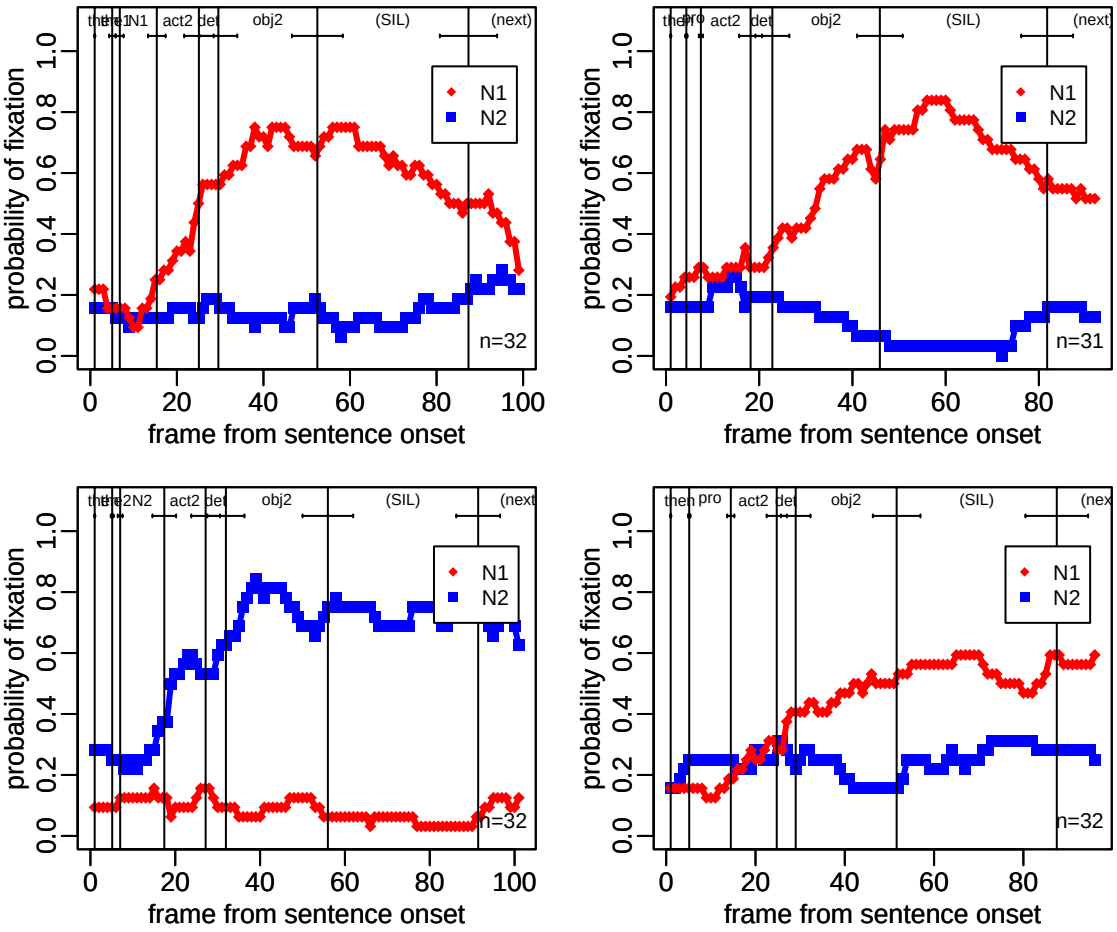

Figure 9: Probability of fixation on objects in visual display for each successive frame in sentence 3 (e.g. [*Then*] $[he]_{pro}$ $[got\ out]_{act2}$ $[some]_{det}$ $[sponges]_{obj2}$). Character and associated object categories have been collapsed. Solid vertical lines mark mean (with standard deviations) onsets of words in the utterance. Each experimental condition is plotted separately: Full N1 (upper left), full N2 (lower left), unaccented pronoun (upper right), or nuclear-accented pronoun (lower right).

prosodic information is used immediately in on-line processing, fixation patterns should largely resemble the plot of the full N2 condition. However, none of these three scenarios can account perfectly for the patterns observed in our data. The plot shows equal amounts of activation of both referents immediately after the pronoun is uttered, but with fixations on N2 tapering off slightly as the utterance progresses. Clearly, in this condition there is competition among the two visual referents, and among their associated objects, with a slight advantage going to the N1 interpretation. Listeners are confused.

Figure 9 plots these same data, but with categories for N1 and his associated object, and N2 and HIS associated object collapsed. The general fixation patterns are shown more clearly: there is a clear preference for N1 when a full N1 is uttered, for N2 with a full N2, for N1 with an unaccented pronoun, and only a slight preference for N1 over N2 when the accented pronoun is uttered. In the absence of a detailed statistical analysis, there appears to be a 2:1 preference for N1 to be the referent of the accented pronoun. This preference starts to appear soon after the offset of the verb, and continues throughout the entire utterance.

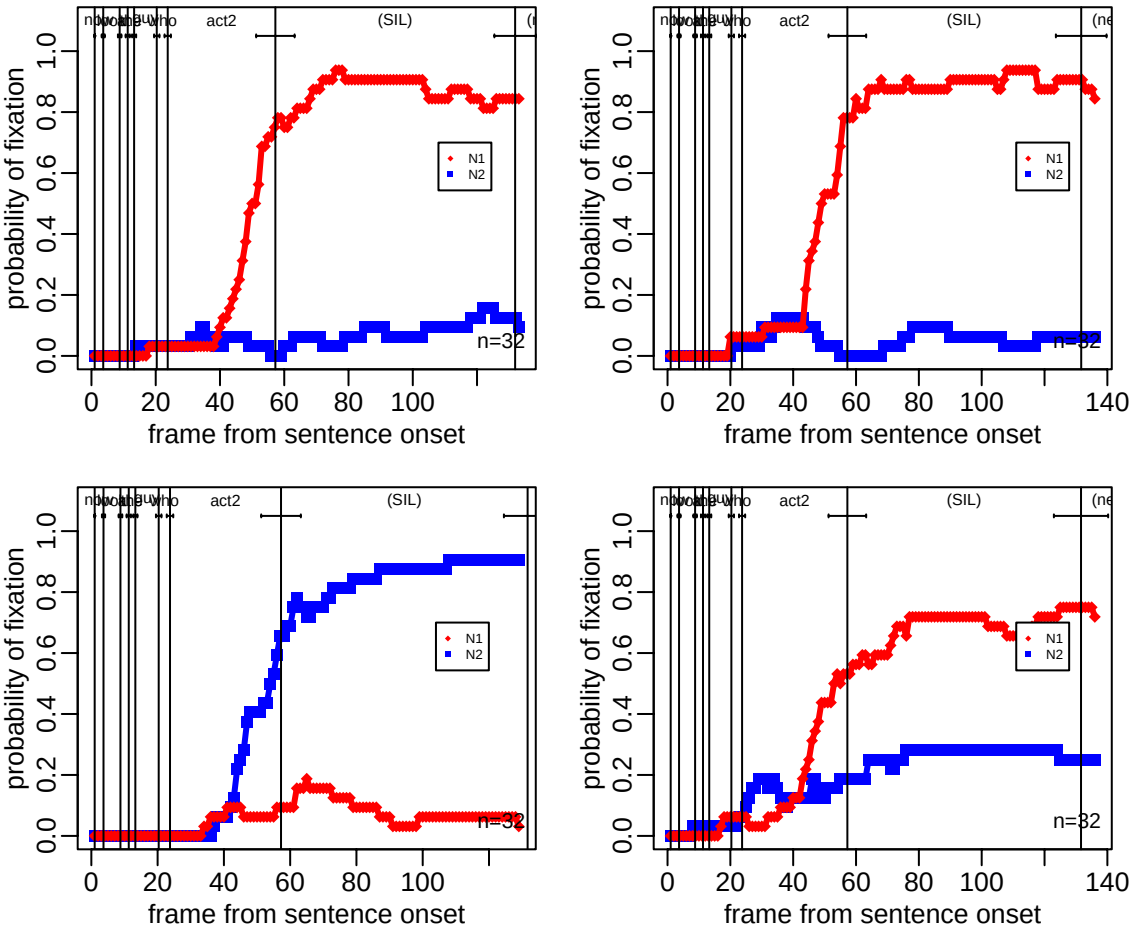

Figure 10: Probability of fixation on objects in visual display for each successive frame in the target instruction (e.g. [*Now*] [*look*] [*at*] [*the*] [*guy*] [*who*] $[got\ the\ sponges]_{act2}$). Solid vertical lines mark mean (with standard deviations) onsets of words in the utterance. Each experimental condition is plotted separately: Full N1 in sentence 3 (upper left), full N2 (lower left), unaccented pronoun (upper right), or nuclear-accented pronoun (lower right).

Off-line preferences

The previous discussion suggests that prosodic information is salient and relevant for immediate on-line interpretation of pronominal forms, but that the cue is somehow ambiguous and causes listener confusion, at least in these data. It is possible that this information can only be fully utilized in conjunction with other semantic information and pragmatic inferences which may occur after the entire utterance has been presented.[38] Therefore, eye fixations during our off-line 'look at' instruction task may shed more light on the time-course of interpretation. If prosodic information is fully used only at a very late stage, listeners' interpretations should be determinate in this off-line task, which occurred 3 sentences after the target sentence in the discourse. At this late stage, listeners should prefer (as measured by probability of eye fixations) the full N1 and unaccented pronoun to refer to N1, and the full N2 and accented pronoun to refer to N2.

Figure 10 plots eye fixations during this 'look at' task: *Now look at the guy who got the sponges*. It is clear that the predictions hold for the full N1, full N2, and unaccented pronoun conditions, but listeners are still 'confused' in the nuclear-accented pronoun condition. There is still a 2:1 preference for the ref-

[38] See the discussion in Section 1.1.3 above about Kameyama's claim that the presupposed constraint of contrast is discharged after the entire clause is parsed. We will return to this in Section 4.

subj	on-line		off-line
1	N2?	=	N2?
2	N1?	=	N1?
3	N1	=	N1
4	N1?	$\rightarrow$	N1
5	?	=	?
6	?	$\rightarrow$	N2?
7	N1	=	N1
8	?	$\rightarrow$	N1

Table 2: Summary of individual subject preferences in on-line and off-line interpretations of who did the action described by the target sentence.

erent of the accented pronoun to be N1 rather than N2. That is, interpretation of the accented pronoun is indeterminate.[39]

Subject analysis

Since both the off-line and on-line eye fixation data are data pooled over all subjects (and items), the question then becomes: Is each subject choosing N2 as the antecedent of the accented pronoun 1/3 of the time, or are 1/3 of all subjects choosing N2 all of the time? Clearly a subjects analysis is warranted, though it may be confounded by differences among items due to the small number of subjects used in this pilot experiment (2 per list). Instead, we choose to summarize qualitative differences in on-line interpretation behavior among subjects. This analysis is given in Table 2. Fixation preferences were informally judged as to whether the subject (i) fixated on N1 or N2 in all (four) trials of the nuclear-accented condition, (ii) fixated the same referent in 3 of 4 trials (N1? and N2?), or (iii) fixated equally on both referents (?). Of course, these informal tallies should be replaced with formal subjects analyses in subsequent experiments.

The tallies show a clear effect of subject, with some people interpreting the nuclear-accented pronoun as referring to N2 (e.g. subjects 1 & 6), others interpreting it as referring to N1 (e.g. subjects 3 & 7), and still others uncertain (e.g. subject 5). In addition, the data suggest that listeners may change their 'vote' in the off-line judgment. For example, subject 6 showed uncertainty on-line while parsing the target sentence, but leaned toward N2 in the off-line task. Other subjects did the reverse: subjects 4 and 8 shifted from uncertainty on-line to a firm N1 judgment off-line.

The nature of the interpretation uncertainty on-line and apparent changing of vote in some cases off-line might be illuminated by subject responses to the verbal debriefing session following the experiment session. When asked if they knew what the experiment was about, most subjects responded that they thought it was about "what we looked at". Some subjects noticed that the experiment investigated "who we thought did what". When asked explicitly about the pronoun ambiguity and about the cases of 'emphatic' (nuclear-accented) pronouns, subjects were divided in their opinions. Some subjects (e.g. 1 & 6) reported that they thought the accented pronoun referred to "the other guy", as predicted by our hypothesis. The eye fixation data for these two subjects reflect this interpretation. Some subjects reported that they were sure in the unaccented pronoun case (i.e. N1 is the antecedent), but were "unsure" or "confused" in the accented pronoun case. More interestingly, some subjects reported that although they thought the accented pronoun should refer to "the other guy", they just tried to "go by the grammar". In fact, a number of subjects reported that they learned in school that "you shouldn't start a new sentence with a pronoun unless you are talking about the previous subject", or that "English grammar requires a pronoun to refer to 'the same guy' unless the other guy is introduced first with a full noun phrase" (e.g. subject 8).[40] These reports are extremely worrisome. They suggest that prescriptive notions about when a pronoun should be used *in writing* may be

[39] In a separate pilot study, we we used the 'thematic' H*L-H% accent instead of the L+H*L- accent. All other details were the same, though fewer subjects were run. Fixation patterns both on-line and off-line were highly comparable to the data for L+H*L- reported here.

[40] All of their reports are paraphrased here.

influencing subject responses, even when listening to intoned speech. Of course, the 'rules' that they are citing to describe written pronoun use would hold for unaccented spoken pronouns as well. These biases may come into play in the off-line judgment task, which is the only task in the experiment that the subjects consciously do. But could it be responsible for the uncertainty in on-line interpretation? This is still an open question. Since the experimental setting is a formal laboratory context, in which subjects look at a computer screen and hear constructed stories about scenes which seem like children's books (in fact, we tell them that we also do this experiment with kids), this social context alone may prime them to use 'proper' grammar both on-line and off-line. Or, it may be the case that it's only when subjects are specifically asked to make a judgment about 'who did what' that they invoke their prescriptive knowledge base. We may never know which, if either, is the case. However, since listeners are usually not at all conscious of the many rapid eye movements they make while viewing a scene, we suspect that subjects probably did not have access to this prescriptive filter in on-line comprehension (as measured by eye fixations). The fact that subjects reported that at first they thought the referent was the other guy, but then just "went by the grammar", suggests that this is so.

If not prescriptive biases, then what might be causing the indeterminacy in interpretation on-line? Another possibility is that the use of a nuclear-accented pronoun in these particular discourse contexts is just not felicitous, or at least not sufficient to point to a single determinate antecedent. As mentioned in Section 1, the literature on accented pronouns for the most part does not describe the range of discourse contexts in which such references are (or are not) felicitous. Rather, most of the studies describe intuitions about coreference only in very parallel clause sequences like *John hit Bill and then HE hit George*, presented in isolation (e.g. [Gleit61, AJ70, Lak71, Oeh81, Sol83, Sol84, Smyth94, BST98]). In more recent work on pronoun resolution, researchers have generalized the use and interpretation of accented pronouns beyond just parallel structures (e.g. [Cahn, Naka93, Terk93, Cahn95, Prev95, Prev96, Naka97a, Naka97b, Kame99]). It was by these accounts that we initially formed the hypothesis that a nuclear-accented pronoun would serve to cue a shift in interpretation in discourses in which there is a basis for contrast. However, based on the discourses used in our experiment, this hypothesis was not fully supported by our on-line eye movement or off-line judgment data.[41]

What might be the cause of the discrepancy between previous theoretical accounts and our experimental results? One possibility is that something about our tasks (following along on-line and making judgments off-line) was such that they were not sensitive enough to test the hypotheses. This is unlikely, but still possible. Another possibility is that determinate interpretation of accented pronouns as switching reference is indeed only observed in parallel structures, whatever the definition of 'parallel' might be (more on this definition later). To test this hypothesis — that accented pronouns cue shifts in interpretation in parallel structures — we included a few examples of strictly parallel *John hit Bill and then HE hit George*-like examples as filler discourses in our experiment, just as a sanity check. Results from these discourse types are presented below.

3.4.4 The infamous "parallel structures"

A small number of parallel structures describing animals hitting each other were included as fillers in our experiment (see Set 5 in the Appendix). The example in (12) shows the structure of the parallel discourse contexts.

(12) 0. The animals were playing out near the barn when something unexpected happened.
1. The lion started going ballistic.
2. He hit the alligator with a long wooden rake,
3a. Then he hit the duck. ('unaccented pronoun' = un)
3b. Then HE hit the duck. ('nuclear-accented pronoun' = nuc)
4. A big fight ensued and it was a terrible scene.

Sentence 0 introduces the animals as a group and sets up the context of an unexpected event. Sentence 1 picks out N1 (here, the *lion*) as the salient entity, then sentence 2 continues to talk about what this salient entity did, using a subject pronoun to refer to N1 and a full NP to refer to N2. The prepositional phrase

[41] See Section 4 for more discussion of the contexts in which accented pronouns are or are not felicitous.

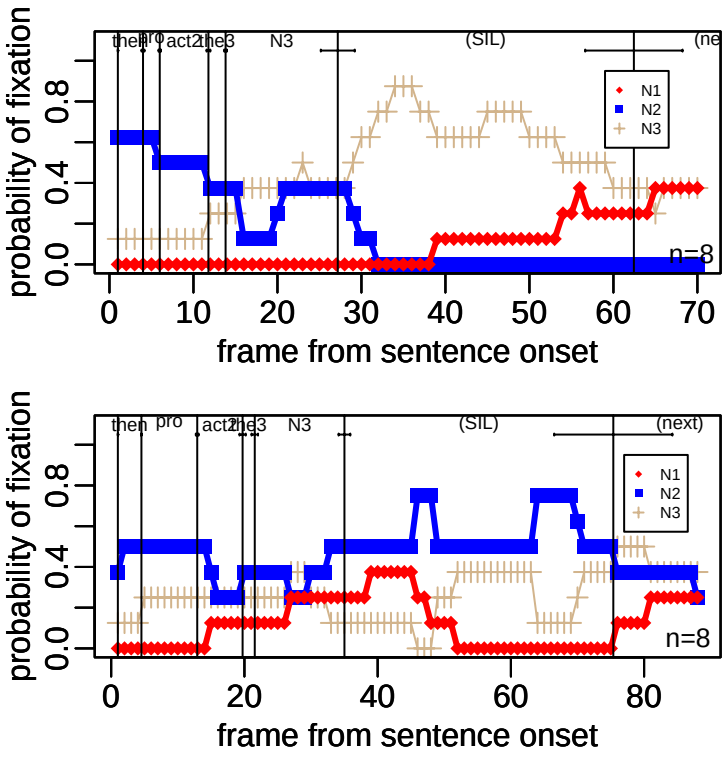

Figure 11: Probability of fixation on objects in visual display for each successive frame in sentence 3 of the parallel structures (e.g. [*Then*] $[he]_{pro}$ $[hit]_{act2}$ [*the*] $[duck]_{N3}$). Solid vertical lines mark mean (with standard deviations) onsets of words in the utterance. Each experimental condition is plotted separately: unaccented pronoun (top) or nuclear-accented pronoun (bottom).

(e.g. *with a long wooden rake*) was included in sentence 2 to lead eye fixations away from N2 to a neutral zone in the scene. Sentence 3 is the target utterance, containing either an unaccented or nuclear-accented subject pronoun (the two full NP subject conditions were not included), and a full object NP referring to N3. According to the hypothesis that an accented pronoun cues a shift in attention to a discourse entity that is not currently the most salient, and that such a shift in attention is reflected in increased fixation probability on that non-salient referent, we can make the following predictions. In the unaccented case (12.3a), listeners should either fixate N1 upon hearing the subject pronoun, or possibly not fixate any character at all (due to the fact that N1 is already highly salient, as described in Section 3.4.2). In either case, they are not expected to fixate on N2 after hearing the pronoun. Then, upon hearing the object NP, fixations should move to N3. In contrast, in the accented pronoun case (12.3b), we predict that there will be an increased probability of fixation on N2 upon hearing the subject pronoun *HE*, then fixations should move to N3. In this condition, N1 should not be considered.

Figure 11 shows the fixation probabilities during the target sentence 3 in the parallel examples: *Then he/HE hit the duck*. At the beginning of the target utterance in the unaccented case (top panel), subjects are fixating N2. This is actually carry-over from the previous utterance, where N2 was mentioned in object position. Ideally, subjects should be looking at the instrumental object (e.g. the *rake*) at the end of sentence 2, but in fact they hardly considered that object.[42] After N3 is uttered, there is a marked increase in fixation probability of N3 (the 'new' information), as expected. In this condition, there are little or no fixations on N1, indicating that this referent is either (i) already highly salient in the discourse, or (ii) not considered as an antecedent of the target *he*. We suspect that the former is the case. In the accented pronoun condition

[42]This is fixation behavior is curious, given the linking hypothesis outlined in Section 3.4, and the fact that the *rake* here is 'new' information. Future studies should examine this more closely.

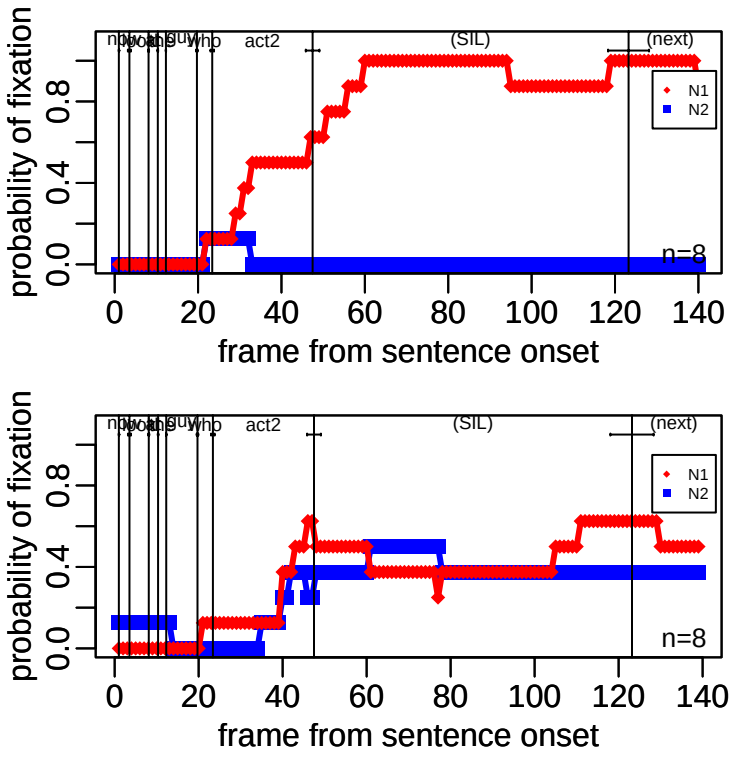

Figure 12: Probability of fixation on objects in visual display for each successive frame in the target instruction of the parallel structures (e.g. [*Now*] [*look*] [*at*] [*the*] [*guy*] [*who*] $[hit\ the\ duck]_{act2}$). Solid vertical lines mark mean (with standard deviations) onsets of words in the utterance. Each experimental condition is plotted separately: unaccented pronoun (top) or nuclear-accented pronoun (bottom).

(bottom panel), in contrast, the probability of fixating N2 is high throughout the entire utterance. Fixations near the utterance onset are most likely due to the carry-over effect also seen in the unaccented case, but these should fall off rapidly if N2 is no longer considered. Instead, subjects continue to fixate N2. In addition, the increase in N1 fixation probability after *HE* is uttered indicates that N1 is also being considered early on, but this activation soon falls off, and N2 remains activated. In fact, N2 competes with N3 during the silent interval after the utterance has ended. When viewing subjects' eye movements in real time, it was common to see rapid saccades back and forth from N2 to N3 and back to N2 during this interval. Such rapid and continued shifts in fixation location results in increased probabilities of fixation for both referents in such a graphic representation. These data suggest that listeners consider N1 as the antecedent of the unaccented pronoun and N2 as the antecedent of the nuclear-accented pronoun (albeit with competition from N1 early on) on-line while listening to discourses containing parallel syntactic structures. Do these preferences remain in the off-line 'look at' judgment task?

Figure 12 plots eye fixations during this 'look at' task: *Now look at the guy who hit the duck*. The graphs show that, while listeners unambiguously take N1 to be the antecedent of the unaccented pronoun *he*, there is now confusion about who the antecedent of *HE* is. There is an equal number of 'votes' for N1 as there are for N2. This result is mysterious, given that the on-line eye fixations showed a strong preference for N2 as the antecedent. One possible explanation for this change of vote off-line is, again, prescriptive biases. Subjects may be able to access their prescriptive knowledge base to make off-line judgments, which might then override their (unbiased) on-line preferences.[43]

[43] In the separate pilot study in which we used the 'thematic' H*L-H% accent, we observed switched reference to N2 on-line and also (unambiguously) off-line, in the parallel structures. This contrasts with the ambiguous off-line judgments shown here in the L+H*L- case. This suggests that the H*L-H% tune may enhance the shift effect in off-line judgments, although this deserves further

Table 3 shows informal tallies of individual subject preferences, for both the narrative ('joint collaborative action') experimental stimuli presented in Section 3.4.3, and for the parallel structures discussed here. The tallies highlight two main trends in the parallel structures: (i) in their on-line judgments, subjects either prefer N2 or are confused about the appropriate antecedent, while (ii) in their off-line judgments, as many as 4 of 8 subjects change their vote to N1, despite the fact that they entertained N2 on-line. This suggests that additional information is coming into play to influence their off-line judgments, and based on subject reports during the debriefing session, we suspect that this information may be prescriptive knowledge about 'proper' pronoun use.[44]

	NARRATIVE			PARALLEL		
subj	on-line		off-line	on-line		off-line
1	N2?	=	N2?	N2	=	N2
2	N1?	=	N1?	?	$\rightarrow$	N2
3	N1	=	N1	N2	$\rightarrow$	N1
4	N1?	$\rightarrow$	N1	N2	=	N2
5	?	=	?	N2	$\rightarrow$	N1
6	?	$\rightarrow$	N2?	?	$\rightarrow$	N1
7	N1	=	N1	?	=	?
8	?	$\rightarrow$	N1	?	$\rightarrow$	N1

Table 3: Summary of individual subject preferences in on-line and off-line interpretations in the main test stimuli (repeated from Table 2) and the parallel stimuli.

The only other study that we are aware of which has examined on-line interpretation of accented pronouns is a cross-modal naming study conducted by Balogh and colleagues [BST98]. In their experiment, they auditorily presented subjects with a sequence of parallel clauses, such as: *The cowboy pushed the robber into the chairs by the bar and the waiter pushed him/HIM into the poker table by the staircase*, and had subjects read aloud a written probe word which appeared on the screen either 800ms before the target pronoun, or right at the pronoun offset. They found that reaction time to naming a probe related to the object NP of the first clause (here, the *robber*) was faster than in an unrelated control condition, for *both* accented as well as unaccented pronouns, but only when the probe appeared at the pronoun offset. They concluded from this that the grammatically-parallel referent (here, the object NP) is accessed immediately upon hearing the pronoun, and that 'contrastive stress' on the pronoun does not interfere with this process. This finding is both consistent with, and contradictory to, our on-line eye fixation data. On the one hand, we also observed that the 'default' (grammatically parallel) referent is considered immediately after the accented pronoun is uttered. However, we also observed activation of the non-parallel referent (the 'other guy') in these cases. In fact, there is competition between the two early on. In Section 4 we will discuss this initial competition in more detail.

3.4.5 When accented pronouns appear not to switch reference

In this section, we present a final observation from our experiment regarding cases in which accented pronouns appear not to shift the center of attention. In Section 1.1.4 we discussed cases in which accent does not switch reference because there is only a single entity in the salient subset. These are not the cases we will describe here. Rather, we will examine cases in which there are indeed two salient referents in the immediate context, but the accent still does not switch reference from the most salient (i.e. highest-ranked) one. Consider the discourse given in (13).

careful investigation.

[44] Another possible explanation for this change of vote is that the difficulty subjects had in getting N2 to be the antecedent of accented *HE* in the narrative 'joint collaborative action' test stimuli (which constituted the bulk of the experiment) might have contaminated the off-line preferences in the parallel structures. That is, if the switched reference due to the accent is not felicitous in a majority of the stimuli, then this may weaken the effect on the parallel examples as well. We are currently running an experiment in which only parallel structures are tested, along with a far greater number of distractor discourses. This should rule out any possibility of cross-contamination.

(13) 1. The zebra asked the pig to help wash the car.
2. He told the pig to put a bucket of soapy water near the front of the car.
3a. Then he got out some sponges.
3b. Then HE got out some sponges.
4. And together they started washing the hood and the fenders.

Sentence 1 introduces both N1 and N2 into the discourse context in a joint collaborative action. Sentence 2 then realizes N1 with a subject pronoun, and N2 with a full NP object (which is also coindexed with the subject trace in the following complement clause). The target sentence then refers to one of these salient characters with either an unaccented pronoun (13.3a) or an accented pronoun (13.3b). Our introspective judgment is that, in this context, the pronoun in sentence 3 refers to N1 (the subject of the preceding matrix clause) regardless of whether it bears a pitch accent or not. The function of the accent in this case is not to shift interpretation from the default antecedent, but rather to cue an explicit contrast to N2's contribution to the joint action (described in sentence 2). Do listener judgments confirm this intuition?

In order to test this experimentally, we included in our list of stimuli a subset of discourses of the type described in (13), in which N1 tells N2 to do some action. These 'telling' discourse types were also matched with 'doing' types, in which N1 does the action described in sentence 2. Examples of the exact context of the 'doing' and 'telling' types can be found in Sets 2 and 3 in the Appendix. Fixation patterns during presentation of sentence 2 in both types were described in detail in Section 3.4.2 above. Now we turn our attention to the pronoun interpretation in the target sentence 3. For ease of presentation, we will show only the off-line judgment data here (i.e. *Now look at the guy who got the sponges*).

The fixations during the target sentence for the 'doing' type will not be shown here since they pattern very similarly to the main narrative stimuli already presented in Figures 8 and 9: there is a determinate preference for N1 in the unaccented case, and a more ambiguous 2:1 preference for N1 over N2 in the accented case.[45] Figure 13 plots the probability of fixations in the off-line judgment task for the 'telling' discourse types, in which N2 is the do-er of the action described in the complement clause. In this context, the judgments are strikingly different. Listeners now prefer accented *HE* to refer determinately to N1 (consistent with our introspective judgments), while unaccented *he* is now the one showing ambiguity. In this discourse context, listeners show a tendency to prefer N1 as the antecedent of the unaccented pronoun, though the activation of N2 is substantial.

There may be a number of explanations for this patterning. One possibility could be that somehow N2 is ranked higher than N1 in the salient subset of entities in U_{i-1}, thus making N1 the most salient after re-ranking (due to the accent). This could account for the robust N1 preferences in the accented case, but it predicts that there should be a stronger preference for N2 in the unaccented case. Since in the Centering Theory literature it remains unclear how entities are ranked in complex utterances (see [Kame98, Milt]), this possibility remains an interesting open research question.[46]

Another possibility is that the accent on the pronoun cues the listener to search for a proposition in the preceding discourse which describes a contrasting contribution to the overall joint collaborative action. Since U_{i-1} describes N2's contribution (regardless of the fact that N1 is the matrix subject), this proposition most readily fits the description. This results in the interpretation of U_i as N1's (contrasting) contribution. We will return to this issue in the general discussion in Section 4 below.

[45] It is important to point out that the discourses in the main narrative data subset (Set 1 in Appendix, see Section 3.4.3) all began with a conjoined NP subject in sentence 1, then used a full N1 as subject in sentence 2. In contrast, this 'doing' data subset (Set 2 in Appendix) introduced N1 as subject and N2 as object in sentence 1, then referred to N1 using an unaccented subject pronoun and to N2 using a full NP in sentence 2. If anything, we would predict that listeners would have more difficultly resolving the referent of the target *HE* (in sentence 3) in the latter discourse type, in which the center of attention is already well-established by use of the pronoun *he* in sentence 2. Using a pronominal in sentence 3 (albeit accented) to now shift the center of attention may be less felicitous in this case. However, we can only speculate about this at this point. Our data show that both discourse types exhibit remarkably similar fixation patterns.

[46] Eleni Miltsakaki has suggested that interpretation of the target pronoun in sentence 3 will crucially depend on whether the listener perceives the sub-discourse opened by the verb *tell* to be completed or not. If the sub-discourse is closed at the end of sentence 2 (possibly due to a terminal L-L% fall accompanied by final F0 and amplitude lowering?), then the target *he* is likely to refer to N1, and *HE* to N2. However, if the target utterance is perceived as a continuation of the 'telling' sub-discourse (e.g. *N1 told N2 to* [*do X then do Y then do Z*]), then the target *he* after *then* may be interpreted as referring to N2, and *HE* to N1. The eye fixation data suggests that the latter scenario may be the case, though further experimental investigation is warranted.

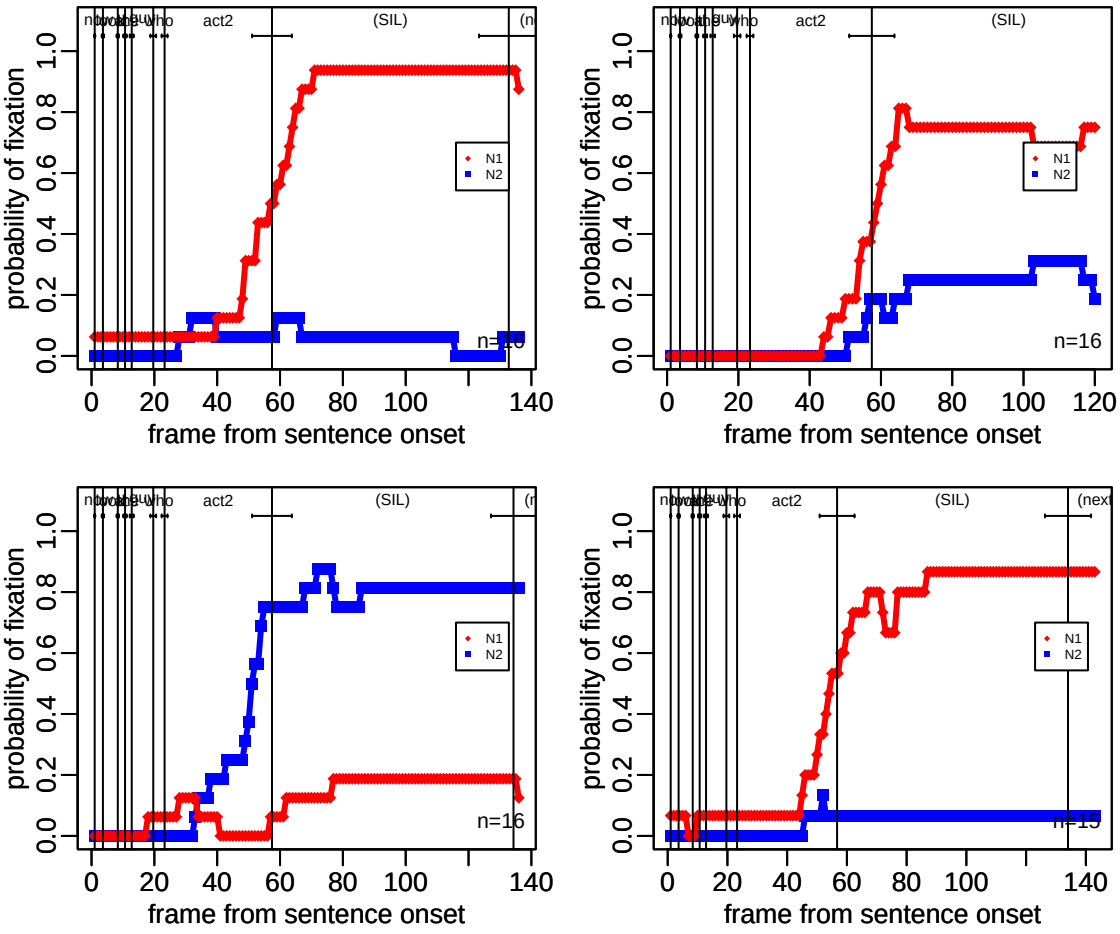

Figure 13: Probability of fixation on objects in visual display for each successive frame in the target instruction in the 'telling' discourse types (e.g. [*Now*] [*look*] [*at*] [*the*] [*guy*] [*who*] $[got\ the\ sponges]_{act2}$). Solid vertical lines mark mean (with standard deviations) onsets of words in the utterance. Each experimental condition is plotted separately: Full N1 in sentence 3 (upper left), full N2 (lower left), unaccented pronoun (upper right), or nuclear-accented pronoun (lower right).

4 Discussion and preliminary proposal

The main focus of this experiment was the investigation of on-line and off-line interpretation preferences of unaccented and nuclear-accented pronominal expressions in connected spoken discourse. Our main findings are the following:

- UNACCENTED PRONOUNS: The interpretation of unaccented pronouns is determinate, and occurs rapidly after the pronoun is encountered in the discourse. That is, unaccented pronouns are unambiguously taken to refer to the most salient entity (N1) of the previous utterance, and such interpretation is reflected by increased probability of eye fixations to N1 on-line immediately after the pronoun is uttered. This preference is maintained in off-line judgments as well. This result is consistent with the eye-tracking results reported by Cooper [Coop74] and Arnold et al. [AEBST00], and other on-line and off-line measures reported by various authors.[47]

[47] But see Section 3.4.2 for discussion of the lack of fixations on pronominal referents in some cases.

- NUCLEAR-ACCENTED PRONOUNS: The interpretation of accented pronouns is more complicated. Our data show that accent alone is not sufficient to switch reference to a less salient entity, contrary to the general proposals offered by attention-driven theories of pronoun interpretation (e.g. [Cahn, Cahn95, Terk93, Naka93, Naka97a, Naka97b, Kame99], see extended discussion of Kameyama's proposal in Section 1.1.3). Rather, we find that interpretation is influenced by the discourse context. In contexts in which adjacent utterances are in a parallel relation, listeners can interpret the accented pronoun as switching reference. However, in contexts in which adjacent utterances are in a (non-parallel) narrative relation, the ability of the accented pronoun to switch reference is less clear. In both contexts, we observe competition among both salient referents early on just after the pronoun is encountered. However, as verbal and subsequent propositional information is encountered, a preference for one antecedent over the other emerges, and is dependent on the type of context in which the accented pronoun occurs.

In the following sections, we will outline our proposal detailing the time-course of accented pronoun interpretation in discourse context. We discuss (i) the contribution of 'parallelism', and (ii) the interaction among inferred discourse coherence relations and (narrow focus) pitch accent in interpretation. We also describe the predictions our proposal makes about the time-course of incremental interpretation of accented and unaccented (subject) pronouns.

4.1 The role of 'parallelism'

Previous studies have noted the contribution of parallelism in identifying and generating contrast in discourse. Theune [Theu99] has shown experimentally that it is not just the presence of alternative items (à la [Rooth92, Prev95, Prev96]) that results in preference for contrastive accent, but rather it is the existence of this set *in addition to parallelism* that determines contrast. This observation is consistent with our data: the narrow focus contrastive accent on *HE* served to switch reference only in the contexts in which parallelism could be identified.

This leads us to the question of what the exact nature of 'parallelism' is, and how listeners are able to identify such contexts. This is still an open research question which is currently under much debate. Some researchers argue for a strict syntactically-structured definition of parallelism (e.g. [Smyth94]), while others argue for semantic parallelism (e.g. [vDeem99, Pul97, Theu97]), or parallelism as a discourse coherence relation (e.g. [Kehl01]). We are not able to review each proposal here, but in the following sections we will briefly describe how parallelism comes into play in two very different accounts of pronoun interpretation: Smyth's syntactic priming account [Smyth94], and Kehler's discourse coherence account [Kehl01].

4.1.1 Smyth's syntactic account

In an extensive study of the role of 'parallelism' in pronoun interpretation, Smyth proposes a *feature match hypothesis*, which states that "pronoun resolution is a feature-match process whereby the 'best' antecedent is that which shares the most features with the pronoun" [Smyth94, p. 220]. The features considered by Smyth to play a role include morphological information (e.g. gender/number), grammatical role (e.g. subject vs. object), and thematic role (e.g. agent vs. patient). He cites *syntactic priming* as the means by which a strong feature match may aid pronoun interpretation: "the syntactic 'frame' of the first clause in a sentence will remain active in memory if the following clause has exactly the same structure, and its activation state affects the accessibility of its nodes during pronoun resolution" [Smyth94, p. 220]. The degree to which the syntactic structure of the first clause primes the second will depend on the extent to which the grammatical features match. Smyth reports findings from his study that "even with the minimal inter-clause differences brought about by the addition of a single adjunct to one clause, the proportion of parallel assignment dropped significantly" [Smyth94, p. 220].

How well does this account predict the observed interpretation differences in 'parallel' vs. 'non-parallel' (henceforth, 'narrative') constructions in our experiment? Based on the grammatical features identified by Smyth, it is not clear how our two contexts result in drastically different predictions with respect to feature matching. Consider the structure of our parallel sequences, given in (14).

(14)	∅	the lion	hit	the alligator	with a rake
		$[NP]_{subj}$	*verb*	$[NP]_{dir-obj}$	$[prep\ NP]_{instr-adjunct}$
	then	he	hit	the duck	
		$[NP]_{subj}$	*verb*	$[NP]_{dir-obj}$	∅

Both clauses consist of a subject (agent), verb, and direct object (patient). However, the first clause also has an additional prepositional phrase adjunct, which is predicted by Smyth to be enough to detract from the strong parallelism effect through feature matching. Despite this mismatch, we observe robust effects of interpretation shift in our accented pronoun condition for this discourse type. What about the structure of our narrative discourses, such as in (15)?

(15)	∅	the zebra	put	the bucket	next to the pig	near the car
		$[NP]_{subj}$	*verb*	$[NP]_{dir-obj}$	$[prep\ NP]_{loc-arg}$	$[prep\ NP]_{loc-adjunct}$
	then	he	got out	some sponges		
		$[NP]_{subj}$	*verb*	$[NP]_{dir-obj}$	∅	∅

In these discourses, both clauses also consist of a subject (agent), verb, and direct object (patient). In addition, the first clause has a prepositional phrase argument (required by the verb *put*), and a prepositional phrase adjunct. According to Smyth's claims, these two additional PPs in the first clause may be sufficient to prevent the parallelism effect. This may be the reason why we see a weaker effect of interpretation shift in narrative discourses (15), in comparison with parallel discourses (14).

4.1.2 Kehler's discourse coherence account

In contrast to Smyth's account based on syntactic priming, Kehler [Kehl01] proposes an alternative account of how parallelism comes into play in pronoun interpretation. Kehler claims that interpretation falls out as a side-effect of listeners' underlying desire to *establish coherence* across utterances in discourse. He proposes three main coherence relations that are used, among them the OCCASION (i.e. 'narrative') and RESEMBLANCE (i.e. 'parallel') relations are of interest to us here. In the occasion relation, listeners make the inference that the speaker has used a pair of constituents to describe a single situation localized in space and unfolding in time (this relation is commonly observed in narrative sequences). In the resemblance relation, listeners make the inference that the speaker has used a pair of constituents to place two propositions into correspondence so as to reveal important commonalities and differences between them.

Under this account, listeners will infer occasion among adjacent utterances in both the parallel and narrative discourses, repeated in (14) and (15) below. That is, listeners take the two conjuncts to describe a single fight in (14), and a single event of washing the car in (15). This inference is supported by the discourse connective *then*, which signals the occasion relationship between its matrix clause and a proposition recovered from context (see [WKSJ99] for more details). However, only the parallel context supports the inference of resemblance. That is, Kehler suggests that the similarity in features (presumably both syntactic and semantic) will be used in "identifying sets of parallel entities and relations as arguments to the coherence relation, and then attempting to identify points of similarity and contrast among each set" [Kehl01, p. 153]. Particular significance will be attributed to the commonalities (the act of *hitting*, and in the unaccented case *who hit*) and the differences (*who was hit*, and in the accented case *who hit*) between the two events in the fight.

(14) PARALLEL:
... He [the lion] hit the alligator with a long wooden rake. Then he hit the duck ...

(15) NARRATIVE:
... The zebra put a bucket of soapy water next to the pig near the front of the car. Then he got out some sponges ...

Given such inferences of occasion and resemblance relations, pronoun interpretation falls out as a side-effect of the inference that supports structural relationships in discourse. Listeners' preferences for pronoun resolution covary with the coherence relation they infer to link the clause into the discourse structure. For example, linking clauses together by an occasion relationship triggers general attentional preferences suitable for extended descriptions of situations. A specific model for these preferences might lie in the discourse centering approach proposed by Grosz and colleagues (e.g. [GS86, GJW95]) described in Section 1.1.2 above, in which a pronoun in utterance U_i is taken to refer to the most salient entity in the local attentional state of U_{i-1}, which is generally the subject NP of U_{i-1}. In contrast, linking clauses together by resemblance relationships triggers preferences for resolutions that can help establish the commonalities and differences between successive clauses. Such preferences recall the *parallel function strategy* of Sheldon [Shel74], Solan [Sol83], Smyth [Smyth94], and others, outlined in Section 1.1.1.

4.2 Accented pronoun interpretation based on coherence relations

Accented pronouns provide additional information to constrain interpretation. According to Rooth [Rooth92], (narrow focus) pitch accents are licensed by certain kinds of semantic operators. Semantically, these operators partition the content of a sentence into a background B applied to a focus F; the constituent that expresses F then receives appropriate accentuation. Pragmatically, these operators presuppose a proposition C from the context (see extended discussion in Section 1.1.3). The operators signal a resemblance relation that identifies F as a point of difference between $B(F)$ and C.

How would the search for this presupposed contrasting proposition work in the different discourse contexts which have described in this paper? In the parallel constructions, the situation is quite straightforward. The accented pronoun reflects a focus F on the referent X of *HE*; the background is *hit the duck*. Thus, to interpret the accent we must find a proposition C for which X in X *hit the duck* is a point of difference. When the verb in U_i is encountered (which is identical across clauses), the listener has mounting evidence to infer a resemblance relation, and thus the contrasting proposition C can most easily be found in the previous clause, without need for further accommodation. Therefore, N2 (i.e. *the alligator*) is evoked as the referent of the accented pronoun.

In the narrative discourses, the situation is somewhat different. Again, the accentuation on *HE* triggers the search for a contrasting proposition somewhere in the context. The sequence in (15) *could* be compatible with a resemblance relation (since both describe a contribution to washing the car), though evidence that this resemblance is intended to structure the discourse is weaker. Instead, the listener may prefer an occasion relation, which involves a strong subject (N1) preference (as described above). In this case, listeners may not be able to resolve the presupposition of contrast to U_{i-1} (which would result in *HE*=N2), but may instead prefer to accommodate the presupposition outside the immediate discourse context. This conflict between (i) taking the default referent and accommodating (=N1), vs. (ii) choosing a dispreferred relation (resemblance) and resolving the presupposition locally to U_{i-1} (=N2), results in indeterminacy in interpretation, which is exactly what we observed in the eye fixation data for this condition.[48]

In the case of the telling discourses, there is also an inferred occasion relation (due to the connective *then*), and also only weak support for a resemblance relation, just as in the narrative case. However, in this context either (i) taking the default referent and accommodating, or (ii) resolving the presupposition of contrast locally to U_{i-1}, will result in the same interpretation: *HE*=N1. That is, contrast between U_{i-1} and U_i does not conflict with the default inferred occasion relation between the two utterances. Therefore, since both possibilities converge on the same antecedent, there is no inherent ambiguity in these cases. Listener judgments show that N1 is uniquely preferred as the antecedent of *HE* in this context.[49]

4.3 The time-course of interpretation

The preceding discussion outlined a proposal describing how accented pronoun interpretation is dependent upon (i) the inferred discourse coherence relation and (ii) the ability of listeners to resolve the presupposed

[48] This is compatible with Kameyama's claim that "the infelicity of the stressed *HE* is due to the difficulty in discharging the presupposed focus constraint" [Kame99, p. 309]. Here, the 'difficulty' in discharging the presupposition results from the ambiguity about what to take as the contrasting proposition C.

[49] For more details on resolving accented pronouns based on coherence relations, see also the discussion presented in [VSNT].

constraint of contrast locally in the discourse. Now we turn our attention to the time-course of interpretation, focusing only on the interpretation of subject pronouns.[50] At what point can listeners make inferences about discourse coherence, and at what point can they (uniquely) interpret an accented pronoun? Since we see pronoun resolution as a side-effect of inference about the coherence relation holding among utterances, which interacts with intonational cues, we predict that listeners may need to wait for propositional information lending evidence about coherence before resolution is achieved.

Let us walk step-by-step through the proposed interpretation process, considering what information the listener has at each point in time.

John hit Bill ...

Upon hearing the utterance U_{i-1}, listeners already have quite a bit of information, and can start to make predictions about what a subsequent pronoun (if there is one) will refer to. They know that the discourse is currently about *John* (especially if U_{i-2} was also about *John*), and the salience ranking of entities in U_{n-1} predicts that it is likely that the discourse will continue to be about *John*.

John hit Bill. Then ...

Upon hearing the discourse connective *then*, listeners can infer occasion between utterances U_{i-1} and U_i, due to the semantics of *then* (see e.g. [WKSJ99, Milt]). This triggers general attentional preferences suitable for extended descriptions of situations, and thus enhances the preference that the discourse will continue to be about *John*.

John hit Bill. Then he ...
John hit Bill. Then HE ...

Upon hearing the intoned pronoun, a wealth of new information now becomes available. The the lexical form of the pronoun tells listeners that it is a subject NP, and the fact that an anaphoric form is used cues listeners to search for its referent somewhere in the previous discourse context. This is true for both unaccented and accented forms. If the pronoun is unaccented, listeners are able to make a guess that the relevant antecedent is the most salient entity in the context which is consistent with the occasion relation. Unique resolution to *John* can occur at this point.

In the accented case, in contrast, the information that listeners have is different. They not only know that it is a subject anaphor, but also that there is an additional presupposition of contrast which must be resolved as well. However, it is not clear at this point how the presupposition will be resolved. Will listeners infer contrast with an entity in the immediately preceding utterance U_{i-1}, or will they prefer to accommodate? In our eye fixation data, we observe competition among both salient referents at this point, regardless of the discourse context type. Let us now focus only on the accented case.

John hit Bill. Then HE hit ...

Upon hearing the verb, there is strong evidence in favor of a resemblance (i.e. parallel) relation, due to the verb being identical across clauses. At this point, listeners are able to make a guess that the speaker's intention was to put these two propositions into correspondence so as to reveal important commonalities and differences between them. Given this inference, listeners can then resolve the presupposition of contrast locally to U_{i-1} (resulting in *HE*=N2) without accommodation, as described in Section 4.2 above. Additional propositional information encountered in the utterance (*Then HE hit George ...*) will enhance this inferred resemblance, and hence give support to the chosen interpretation.

John hit Bill. Then HE made ...

What about cases in which the verb information does not give strong support for a resemblance relation? In this constructed example, the verb (*made*) is not identical across clauses, so no strong evidence for resemblance can come from that. Of course, the speaker could be *intending* to cue resemblance (for example, by continuing with something like *Then HE made Bill hit George*) – the issue is (i) *whether* and (ii) *when* listeners can identify this intention. In the absence of strong cues for resemblance (be they lexical

[50] Incremental processing of (accented) object pronouns may proceed quite differently. We will return to this issue in the report of our ongoing study which investigates the on-line interpretation of both subject and object accented pronouns.

or prosodic or whatever), listeners may prefer to interpret an occasion relation between the two utterances, and the presupposed contrast need not be resolved locally. For example, if the speaker continues with *Then HE made a funny face*, then listeners may prefer to take *HE* to refer to the default referent (N1) and accommodate the presupposition of contrast. Our eye fixation data suggest that this ambiguity about how the presupposition should be resolved results in mixed preferences: fixations on both N1 and N2 occur, with slight preference for N1.

In conclusion, this study examined on-line and off-line interpretation of unaccented and nuclear-accented subject pronouns in various discourse contexts. Our findings suggest that understanding accented pronouns is in fact quite a bit more complicated than just *John hit Bill and then HE hit George*. That is, accent alone is not sufficient to switch reference to a less salient entity. We presented data suggesting that (i) the type of inferred discourse coherence relation, and (ii) the ability to locally resolve the presupposition of contrast evoked by the accent, influences the interpretation of accented pronouns. In addition, our data tell us something about the time-course of incremental interpretation of utterances with accented subject pronouns. We find that both potential antecedents are evoked immediately upon hearing the accented pronoun. A preference for one referent over the other only emerges once subsequent propositional information is encountered which lends support for the inferred discourse relation.

Acknowledgements

This is a report of pilot eye-tracking experiments conducted while the first author was a postdoctoral research fellow in the VILLAGE lab at Rutgers Center for Cognitive Science during 2000-2001. The VILLAGE lab is funded by NSF research instrumentation award 9818322. This research was funded by an NIH Training fellowship award 1-T32-MH-19975-03 to Jennifer Venditti and a Rutgers University ISATC grant "Describing action for human-computer communication" awarded to Suzanne Stevenson, administered by Matthew Stone. We would like to thank Doug DeCarlo, Eleni Miltsakaki, Irina Sekerina, Karin Stromswold, John Trueswell, and the members of the Rutgers CogSci/CompSci VILLAGE lab and the Gleitman/Trueswell psycholinguistics lab at Penn for generous help and comments on this research.

References

[AEBST00] Jennifer E. Arnold, Janet G. Eisenband, Sarah Brown-Schmidt, and John C. Trueswell. The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76:B13–B26, 2000.

[AJ70] Adrian Akmajian and Ray Jackendoff. Coreferentiality and stress. *Linguistic Inquiry*, 1(1):124–126, 1970.

[AMT98] Paul D. Allopenna, James S. Magnuson, and Michael K. Tanenhaus. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38:419–439, 1998.

[BE94] Mary E. Beckman and Gayle Ayers Elam. Guidelines for ToBI labelling. Unpublished manuscript, Ohio State University. Version 3.0 March 1997. [http://ling.ohio-state.edu/Phonetics/etobi_homepage.html], 1994.

[Beck91] Wolfgang Becker. Saccades. In R. H. S. Carpenter, editor, *Vision and Visual Dysfunction: Eye Movements*, pages 95–137. MacMillan Press, 1991.

[Bol61] Dwight Bolinger. Contrastive accent and contrastive stress. *Language*, 37:83–96, 1961.

[BST98] Jennifer E. Balogh, David Swinney, and Zachary Tigue. Real-time processing of pronouns with contrastive stress. Poster presented at the 11th Annual CUNY Conference on Human Sentence Processing, 1998.

[Cahn] Janet Cahn. The effect of intonation on pronoun referent resolution. Unpublished manuscript.

[Cahn95] Janet Cahn. The effect of pitch accenting on pronoun referent resolution. In *Proc. of the Association for Computational Linguistics (ACL)*, pages 290–293, Cambridge, Massachusetts, 1995.

[Chafe76] Wallace Chafe. Givenness, contrastiveness, definiteness, subjects, and topics. In Charles N. Li, editor, *Subject and Topic*, pages 27–55. Academic Press, 1976.

[Coop74] Roger M. Cooper. The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6:84–107, 1974.

[GJW95] Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225, 1995.

[Gleit61] Lila R. Gleitman. Pronominals and stress in English conjunctions. *Language Learning*, 11:157–169, 1961.

[GS86] Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[Kame86] Megumi Kameyama. A property–sharing constraint in centering. In *Proc. of the Association for Computational Linguistics (ACL)*, pages 200–206, 1986.

[Kame98] Megumi Kameyama. Intrasentential centering. In Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors, *Centering Theory in Discourse*, pages 89–112. Clarendon Press, 1998.

[Kame99] Megumi Kameyama. Stressed and unstressed pronouns: complementary preferences. In Peter Bosch and Rob van der Sandt, editors, *Focus: Linguistic, Cognitive, and Computational Perspectives*, pages 306–321. Cambridge University Press, 1999.

[Kehl01] Andrew Kehler. *Coherence, Reference, and the Theory of Grammar*. CSLI Publications, 2001.

[Ladd80] D. R. Ladd. *The Structure of Intonational Meaning: Evidence from English*. Indiana University Press, 1980.

[Lak71] George Lakoff. Presupposition and relative well-formedness. In Danny D. Steinberg and Leon A. Jakobovits, editors, *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*, pages 329–340. Cambridge University Press, 1971.

[Milt] Eleni Miltsakaki. *Attention Structure in Discourse: A Cross-lingustic Investigation*. PhD thesis, University of Pennsylvania, in progress.

[Naka93] Christine H. Nakatani. Accenting on pronouns and proper names in spontaneous narrative. In *ESCA Workshop on Prosody*, volume 41 of *Lund Working Papers*, pages 164–167, Lund, Sweden, 1993.

[Naka97a] Christine H. Nakatani. *The computational processing of intonational prominence: A functional prosody perspective*. PhD thesis, Harvard University, 1997.

[Naka97b] Christine H. Nakatani. Discourse structural constraints on accent in narrative. In Jan P. H. van Santen, Richard W. Sproat, Joseph P. Olive, and Julia Hirschberg, editors, *Progress in Speech Synthesis*, pages 139–156. Springer–Verlag, 1997.

[Oeh81] Richard T. Oehrle. Common problems in the theory of anaphora and the theory of discourse. In Herman Parret, Marina Sbisà, and Jef Verschueren, editors, *Possibilities and Limitations of Pragmatics*, pages 509–530. John Benjamins, 1981.

[Prev95] Scott Prevost. *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation*. PhD thesis, University of Pennsylvania, 1995.

[Prev96] Scott Prevost. Modeling contrast in the generation and synthesis of spoken language. In *Proc. of the International Conference on Spoken Language Processing (ICSLP)*, 1996.

[Pul97] Stephen Pulman. Higher order unification and the interpretation of focus. *Linguistics and Philosophy*, 20:73–115, 1997.

[Rob96] Craige Roberts. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Ohio State University Working Papers in Linguistics*, 49:91–136, 1996.

[Rooth92] Mats Rooth. A theory of focus interpretation. *Natural Language Semantics*, 1(1):75–116, 1992.

[Shel74] A. Sheldon. The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13:272–281, 1974.

[Smyth92] Ron Smyth. Multiple feature matching in pronoun resolution: A new look at parallel function. In *Proc. of the International Conference on Spoken Language Processing (ICSLP)*, pages 145–148, Banff, Canada, 1992.

[Smyth94] Ron Smyth. Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, 23(3):197–229, 1994.

[Sol83] Lawrence Solan. *Pronominal Reference: Child Language and the Theory of Grammar*. D. Reidel Publishing Company, 1983.

[Sol84] Lawrence Solan. Focus and levels of representation. *Linguistic Inquiry*, 15:174–178, 1984.

[Terk93] Jaques Terken. Accessbility, prominence, pronouns, and accent. Paper presented at the 1993 Centering Workshop held at the University of Pennsylvania, 1993.

[Theu97] Mariët Theune. GoalGetter: Predicting contrastive accent in data–to–speech generation. In *Papers from the 7th Computational Linguistics in the Netherlands Meeting*, pages 177–190, Eindhoven, 1997.

[Theu99] Mariët Theune. Parallelism, coherence, and contrastive accent. In *Proc. of the European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 555–558, Budapest, 1999.

[TKH+99] Jan Theeuwes, Arthur F. Kramer, Sowon Hahn, David E. Irwin, and Gregory J. Zelinsky. Influence of attentional capture on oculomotor control. *Journal of Experimental Psychology: Human Perception and Performance*, 25(6):1595–1608, 1999.

[TMDC00] Michael K. Tanenhaus, James S. Magnuson, Delphine Dahan, and Craig Chambers. Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, 29(6):557–580, 2000.

[TSKES95] Michael K. Tanenhaus, Michael Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268:1632–1634, 1995.

[TT95] Michael K. Tanenhaus and John C. Trueswell. Sentence comprehension. In J. L. Miller and P. D. Eimas, editors, *Speech, Language, and Communication: Handbook of Perception and Cognition*, volume 11, pages 217–262. Academic Press, 1995.

[vDeem99] Kees van Deemter. Contrastive stress, contrariety, and focus. In Peter Bosch and Rob van der Sandt, editors, *Focus: Linguistic, Cognitive, and Computational Perspectives*, pages 3–17. Cambridge University Press, 1999.

[VSNT] Jennifer J. Venditti, Matthew Stone, Preetham Nanda, and Paul Tepper. Discourse constraints on the interpretation of nuclear-accented pronouns. In *Proceedings of the International Workshop on Speech Prosody*, Aix-en-Provence, (submitted).

[WJP98] Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors. *Centering Theory in Discourse*. Clarendon Press, 1998.

[WKSJ99] Bonnie Webber, Alistair Knott, Matthew Stone, and Aravind Joshi. Discourse relations: A structural and presuppositional account using lexicalised TAG. In *Association for Computational Linguistics*, pages 41–48, 1999.

Appendix: Discourse stimuli

SET 1: main narrative stimuli (‘joint collaborative action’)

1
The zebra and the pig wanted to wash the car together.
The zebra put a bucket of soapy water next to the pig near the front of the car.
Then he got out some sponges.
And together they started washing the hood and the fenders.
<Now look at the left headlight of the car.>
<Now look at the guy who got the sponges.>
<Now look at the guy wearing the red hat.>

2
The duck and the raccoon wanted to paint the shed together.
The duck set a couple buckets of paint near the raccoon by the front of the shed.
Then he grabbed some brushes.
And together they managed to put on the first coat in an hour.
<Now look at the window on the shed.>
<Now look at the guy who got the brushes.>
<Now look at the purple paint in the buckets.>

3
The rabbit and the horse wanted to set up for the birthday party.
The rabbit placed the birthday cake near the horse on the table.
Then he stuck in some candles.
And together they strung up balloons all around the room.
<Now look at the leftmost leg of the table.>
<Now look at the guy who stuck in the candles.>
<Now look at the balloon with the “A” on it.>

4
The dog and the zebra wanted to refinish an old antique table together.
The dog spread out some newspapers near the zebra by an open window.
Then he got a can of varnish.
And together they moved the table onto the newspapers and went about applying the first coat.
<Now look at the green in the window curtain.>
<Now look at the guy who got the can of varnish.>
<Now look at the blue sander next to the table.>

5
The frog and the beaver wanted to set up for a game of chess.
The frog pushed a small table over next to the beaver near the window.
Then he got out some playing pieces.
And together they laid out the board and set up for the game.
<Now look at the leg of the table.>
<Now look at the guy who got out the chess pieces.>
<Now look at the guy wearing the red shirt.>

6

The cow and the rhino wanted to rearrange their office.
The cow put the rolled-up carpet near the rhino next to the filing cabinet.
Then he got out a furniture dolly.
And together they moved the desk and the cabinet across the room.
<Now look at the open cabinet drawer.>
<Now look at the guy who got the dolly.>
<Now look at the yellow vest.>

7

The bear and the rabbit wanted to build more shelves for their bookcase.
The bear laid a couple of spare boards near the rabbit on the sawhorses.
Then he picked up some nails.
And together they cut the boards to length and nailed them into place.
<Now look at the handle of the saw.>
<Now look at the guy who picked up the nails.>
<Now look at the top shelf in the bookcase.>

8

The elephant and the fox wanted to do some work around the yard.
The elephant raked leaves into a pile near the fox under the maple tree.
Then he picked up some fallen branches.
And together they loaded the leaves and the branches into the wheelbarrow.
<Now look at the trunk of the tree.>
<Now look at the guy who picked up the branches.>
<Now look at the guy wearing the shirt with buttons.>

9

The horse and the skunk wanted to paint out in the yard together.
The horse set up their easel near the skunk in front of the flower garden.
Then he got out some brushes and paints.
And together they painted a beautiful spring scene.
<Now look at the yellow flowers in the garden.>
<Now look at the guy who got the brushes and paints.>
<Now look at the guy with his hand on his hip.>

10

The hippo and the bear wanted to wash dishes together.
The hippo piled up the dirty dishes next to the bear on the counter.
Then he put on some rubber gloves.
And together they washed and rinsed every last one.
<Now look at the faucet on the sink.>
<Now look at the guy who put on the gloves.>
<Now look at the guy with the red tongue.>

11

The dog and the kangaroo wanted to set up for their game of horseshoes.
The dog set up a stake near the kangaroo underneath the tree.
Then he picked up a bunch of horseshoes.
And together they took turns practicing shots until the others arrived.
<Now look at the branches of the tree.>
<Now look at the guy who got the horseshoes.>
<Now look at the guy with the long tail.>

12
The hippo and the skunk wanted to have a barbecue together.
The hippo set a bunch of hotdogs near the skunk next to the campfire.
Then he gathered up some sticks.
And together they skewered the meat and started grilling away.
<Now look at the flames of the fire.>
<Now look at the guy who gathered the sticks.>
<Now look at the stakes holding down the tent.>

13
The fox and the pig wanted to chop some firewood together.
The fox set a large log next to the pig on the tree stump.
Then he got out an axe.
And together they chopped up the wood and stacked it next to the cabin.
<Now look at the base of the tree stump.>
<Now look at the guy who got the axe.>
<Now look at the guy with the white fur.>

14
The zebra and the raccoon wanted to catch some fish for dinner.
The zebra set a couple of fishing poles near the raccoon against the cooler.
Then he got out some worms.
And together they fastened them to the hooks and cast their lines.
<Now look at the latch on the cooler.>
<Now look at the guy who got the worms.>
<Now look at the water in the lake.>

15
The dog and the alligator wanted to pick some apples together.
The dog propped up a ladder near the alligator against the tree.
Then he got a big basket.
And together they loaded all the apples they could reach into it.
<Now look at the apples in the tree.>
<Now look at the guy who got the basket.>
<Now look at the guy wearing the blue shirt.>

16
The monkey and the flamingo wanted to hang the wash out to dry.
The monkey set down the laundry basket near the flamingo under the clothesline.
Then he picked up some clothespins.
And together they hung every single thing in the basket on the line.
<Now look at the jeans on the clothesline.>
<Now look at the guy who picked up the clothespins.>
<Now look at the guy with the big ears.>

SET 2: ‘doing’-type fillers (N1 is do-er of action in sentence 2)

(numbering is not consecutive)

18

The monkey asked the cow to help decorate for the holidays.
He hung up a string of lights near the cow above the fireplace.
Then he put some candles around the room.
And together they made paper snowflakes to paste on the windows.
<Now look at the grill on the fireplace.>
<Now look at the guy who set the candles around the room.>
<Now look at the guy holding out his hand.>

20

The skunk asked the flamingo to help repot their giant cactus outside on the patio.
He set the cactus down near the flamingo next to the watering hose.
Then he got out some newspapers.
And together they spread them out under the cactus and carefully transferred it into a larger pot.
<Now look at the nozzle of the hose.>
<Now look at the guy who got the newspapers.>
<Now look at the guy with skinny legs.>

23

The pig asked the elephant to help put up some fence posts.
He set a shovel near the elephant next to the pile of posts.
Then he started marking out locations for the holes.
And together they dug down about a foot and set the posts into place.
<Now look at the ends of the posts.>
<Now look at the guy who marked the locations.>
<Now look at the red siding of the barn.>

24

The cat asked the duck to help carry their old dresser up to the attic.
He laid down the drawers near the duck next to the bed.
Then he grabbed the base of the dresser.
And together they lifted it up and headed towards the attic.
<Now look at the green blanket on the bed.>
<Now look at the guy who grabbed the base of the dresser.>
<Now look at the pink curtains.>

26

The beaver asked the lion to help clean the living room.
He moved the coffee table next to the lion underneath the window.
Then he gathered up some books from the floor.
And together they vacuumed and dusted the whole place.
<Now look at the television antenna.>
<Now look at the guy who gathered up the books.>
<Now look at the guy with the big front teeth.>

29
The fox asked the dog to help get their kite down from the tree.
He set up a ladder next to the dog against the tree.
Then he picked up some long poles.
And together they managed to poke the kite free.
<Now look at the kite in the tree.>
<Now look at the guy who picked up the poles.>
<Now look at the guy wearing the shirt with the collar.>

30
The raccoon asked the hippo to help pick up their playroom.
He set the toy chest next to the hippo near the chair.
Then he gathered up some toys.
And together they loaded them one by one into the chest.
<Now look at the bear on the chair.>
<Now look at the guy who gathered up the toys.>
<Now look at the guy with white toenails.>

32
The horse asked the mouse to help build a sandcastle.
He put a shovel and a bucket near the mouse beside the sandbox.
Then he put on some rubber boots.
And together they jumped into the sandbox and made a huge castle.
<Now look at the sand in the sandbox.>
<Now look at the guy who put on the boots.>
<Now look at the guy with the big ears.>

SET 3: 'telling'-type fillers (N2 is do-er of action in sentence 2)

(numbering is not consecutive)

17
The lion asked the monkey to help tend the fire.
He told the monkey to put some logs next to the fireplace.
Then he lifted off the firescreen.
And together they loaded the logs carefully onto the fire.
<Now look at the books on the mantle.>
<Now look at the guy who lifted off the firescreen.>
<Now look at the guy with the yellow fur.>

19
The alligator asked the turtle to help fix the old lawnmower.
He told the turtle to put the tools they'd need on the workbench.
Then he got a can of oil.
And together they set to work trying to get it started.
<Now look at the tools hung on the workbench.>
<Now look at the guy who got the oil can.>
<Now look at the guy with the long tail.>

21
The flamingo asked the kangaroo to help pack for their picnic in the park.
He told the kangaroo to put a bunch of food on the table.
Then he got out a picnic basket.
And together they loaded it full of food for their lunch.
<Now look at the cloth on the counter.>
<Now look at the guy who got out the basket.>
<Now look at the two bottles of cola.>

22
The rhino asked the cat to help pack for their trip to the beach.
He told the cat to put their old suitcase on top of the bed.
Then he gathered up all of their beach gear.
And together they they stuffed every last thing into the suitcase.
<Now look at the pillow on the bed.>
<Now look at the guy who gathered up the beach gear.>
<Now look at the blue swimming shorts.>

25
The turtle asked the mouse to help set up for their soccer game.
He told the mouse to kick the soccer ball out next to the fence.
Then he got some cones to use as a goal.
And together they took turns practicing penalty shots until the game began.
<Now look at the rightmost fence post.>
<Now look at the guy who got the cones.>
<Now look at the guy wearing the long-sleeve shirt.>

27
The frog asked the cow to help make a stew.
He told the cow to put a large cooking pot on the stove.
Then he gathered up the ingredients.
And together they filled the pot to the brim.
<Now look at the burners on the stove.>
<Now look at the guy who gathered the ingredients.>
<Now look at the handle of the refrigerator.>

28
The mouse asked the frog to help clean up in the garage.
He told the frog to set a couple of empty boxes next to the shelves.
Then he gathered up some of the tools.
And together they packed them up and hoisted the box up onto the top of the shelf.
<Now look at the hammer on the shelves.>
<Now look at the guy who gathered up the tools.>
<Now look at the guy who has his hand near his tail.>

31
The kangaroo asked the alligator to help plant vegetables in their garden.
He told the alligator to dig a bunch of holes in a straight row.
Then he sprinkled in some seeds.
And together they poured a large bucket of water over the top.
<Now look at the handle of the spade.>
<Now look at the guy who sprinkled in the seeds.>
<Now look at the red shirt.>

SET 4: Discourse digression fillers (not discussed)

33
The elephant asked the bear to help trim branches off the old oak tree.
He set down the gardening tools near the bear next to the wheelbarrow.
He used to work in the landscaping business,
So he had a lot of experience and knew just how to trim a big tree like this.
<Now look at the handle of the wheelbarrow.>
<Now look at the guy who used to be a landscaper.>
<Now look at the red handle on the clippers.>

34
The turtle asked the cat to paint together in the park.
He set down their easel near the cat next to an empty bench.
He used to make his living as an artist,
So he has very particular preferences about how to set up for painting.
<Now look at the boards on the bench.>
<Now look at the guy who used to work as an artist.>
<Now look at the guy with the brown shell.>

SET 5: Parallel structure fillers

35
The animals were playing near the barn when something unexpected happened.
The lion started going ballistic.
He hit the alligator with a long wooden rake,
Then he hit the duck.
A big fit ensued and it was a terrible scene.
<Now look at the shovel against the fence.>
<Now look at the guy who hit the duck.>
<Now look at the red siding of the barn.>

36
The animals were playing out in the lawn when something unexpected happened.
The rhino started getting really upset.
He hit the beaver with a toy truck,
Then he hit the rabbit.
Finally someone had to come break up their fight.
<Now look at the stairs on the slide.>
<Now look at the guy who hit the rabbit.>
<Now look at the yellow shirt.>