

Discourse Constraints on the Interpretation of Nuclear-accented Pronouns

Jennifer J. Venditti*, Matthew Stone†, Preetham Nanda† and Paul Tepper†

*University of Pennsylvania Institute for Research in Cognitive Science

†Rutgers Center for Cognitive Science and Dept. of Computer Science

jjv@unagi.cis.upenn.edu, {mdstone, pnanda81, paultep}@cs.rutgers.edu

Abstract

This paper reports the results from an experimental study of on-line interpretation of nuclear-accented (subject) pronouns in English. We present data suggesting that (i) the type of inferred discourse coherence relation, and (ii) the ability to resolve the presupposition of contrast evoked by the accent locally, 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.

1. Introduction

Pronouns uttered with intonational prominence (i.e. ‘pitch accent’) are interpreted differently from those uttered without such prominence. Consider the example 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)

Despite the potential ambiguity, native intuitions are unambiguous: the pronoun uttered without an accent (1a) evokes *John*, while the accented pronoun (1b) evokes *Bill*. Of course, other non-pronominal referring expressions will not switch reference when given intonational prominence. This difference has led a number of researchers to hypothesize a special interaction in processing between pronouns and accent [1, 2, 3, 4, 5, 6].

In this paper, we account for the interpretation of accented pronouns simply in terms of independently-motivated principles of pronoun resolution and intonational meaning. Following Kehler [7], we model pronoun resolution as a side-effect of establishing a coherent structure for discourse. Meanwhile, following Rooth [8], we take (narrow focus) accent to convey the presupposition of suitable alternative information in the discourse. We combine these principles drawing on the insight of Webber et al. [9] that inferences that achieve coherent discourse structure are strictly local, while a much wider range of inferences are available for resolving presupposition.

We take an experimental approach to documenting our account: we use eye-tracking to investigate the on-line incremental process of interpretation of narrow focus nuclear-accented pronouns (henceforth ‘accented pronouns’). We substantiate the prediction of our account that accent is not itself sufficient to switch the reference of a pronoun; the referent (if any) of an accented pronoun depends both on the coherence relation the listener infers between the utterance and the ongoing discourse, and on the alternative to which they link the accent. We further find that the time-course of pronoun interpretation is con-

sistent with our account: the referent which listeners infer appears to reflect their incrementally-accumulated evidence about the overall discourse interpretation.

2. Discourse relations, structure and presupposition

We start from the observation that understanding a discourse involves reconstructing inferential relationships that connect the meanings of its constituent parts. Our conception and terminology for these relationships follows [7]. In this paper, we are particularly concerned with the two kinds of relationships that we see in the interpretations of (1). The first relation is OCCASION, the inference that the speaker has used a pair of constituents to describe a single situation localized in space and unfolding in time. In inferring occasion in (1), we take the two conjuncts to describe a single fight. The second relation is RESEMBLANCE, 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. In inferring resemblance in (1), we attribute particular significance to the commonalities (the act of *hitting*, and in (1a) *who hit*) and the differences (*who was hit*, and in (1b) *who hit*) between the two events in the fight.

There are two ways for discourse to convey such relationships [9]. The first way is through STRUCTURE. Discourse coheres hierarchically, resulting in a tree structure; listeners must infer a suitable relationship between sister constituents within this tree. In our analysis, we are committed that resemblance between clauses, as in (1b) particularly, often represents a structural inference. The second way is through PRESUPPOSITION. Coherent discourse frequently includes adverbials that signal a relation between their matrix clause and some suitable proposition in the discourse context. The requirement to find such a proposition is called a PRESUPPOSITION; RESOLUTION is the name for the process of finding a matching proposition for a presupposition in the context (and thereby determining the relation signaled by a discourse adverbial). Many types of presupposition can also be accommodated, or taken up without comment as though they had been implicit in the discourse all along. Webber et al. [9] argue that *then*, as used in (1), functions as a discourse adverbial; *then* signals the occasion relationship between its matrix clause and a proposition recovered from context.

Following [7], we take pronoun resolution as a side-effect of the inference that supports structural relationships in discourse. Listeners’ preferences for pronoun resolution thus 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. [10, 11]), 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 Solan [4], Smyth [5], and others—a general heuristic by which listeners and 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.

Accent, however, conveys discourse relations by presupposition [8]. For Rooth, (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. The operators signal a resemblance relation that identifies F as a point of difference between $B(F)$ and C .

Understanding discourse is inference to the best explanation [12]. The listener follows as conservative a strategy as possible to determine how to infer coherence and resolve presuppositions. Specifically, the listener aims to recover a coherent interpretation for the discourse while respecting anaphoric preferences, avoiding accommodation, and avoiding inconsistency (other things being equal).

These principles suffice to account for the interpretation of accented pronouns and unaccented pronouns alike. Consider (1). (1a), with an unaccented pronoun in subject position, is a case where resemblance and occasion preferences coincide to predict $he=John$. The presupposition of *then* is also straightforward to resolve on this interpretation.

In (1b) however, accent provides a further constraint on interpretation. The accented pronoun reflects a focus F on the referent X of *HE*; the background is *hit George*. Thus, to interpret the accent we must find a proposition C for which X in X *hit George* is a point of difference. Without accommodating, we can only take C to be the fact *John hit Bill* and this requires X to be different from *John*. (Compare Akmajian and Jackendoff’s observation that “contrastive stress on either a pronoun or noun will prohibit coreference” [1, p. 124].)

Let us then assume that the listener infers a structural resemblance relation to link together the clauses in (1b). For the discourse to describe two contrasting things that the same individual did, it would require interpreting $HE=John$: accentuation rules this out. Thus, the discourse must be describing the resemblance between things that two contrasting individuals did. The resolution of $HE=Bill$ thus represents the most salient consistent interpretation of (1b).

Most theories of accent, parallelism and pronouns are designed to capture straightforward examples such as (1). In this paper, we focus on three further predictions of our theory which distinguish it from previous proposals.

- We predict that in some cases (though not all, e.g. (1a)), listeners may have to wait for evidence about the coherence relation before inferring the referent of a pronoun. These cases pose a problem for theories such as [6, 13, 14] that link accentuation of pronouns directly to salience ranking.

- We predict that the preference for switched reference only arises when the presupposed contrasting proposition C describes the default referent; however, C need not do so. These cases also pose problems for approaches such as [4, 5] that link accentuation of pronouns directly to strategies for achieving coherence (such as the parallel function strategy).
- Since identification of referents proceeds in tandem with inference of coherence relations, we predict that listeners may sometimes face an uncertain tradeoff in resolving an accented pronoun: whether to accommodate a presupposition or to infer a dispreferred coherence relation. These ambiguities raise general problems for theories that only address one aspect of discourse interpretation.

In the rest of this paper, we present empirical evidence consistent with each of these predictions.

3. On-line interpretation with the resemblance relation

Given that accented pronouns have been shown to switch reference in off-line judgments of sequences in which resemblance is inferred (as in (1b)), in this study we are interested in examining the incremental nature of this interpretation in on-line processing. We use eye-tracking to monitor which referents in a visual scene listeners are considering at any given point in a spoken utterance. Eye-tracking has been used successfully to track spoken language comprehension by a number of authors (e.g. [15, 16]). Arnold and colleagues [17] have employed this method to investigate the incremental interpretation of (unaccented) gender-ambiguous and unambiguous pronouns in discourse. They found that gender information is used by listeners early on to identify a unique referent, immediately after the pronoun is uttered. How then might intonational information be used in the early stages of processing? Since we see pronoun resolution as a side-effect of inference about the coherence relation holding among utterances, which may interact with intonational cues, we predict that listeners may need to wait for propositional information lending evidence about coherence before resolution is achieved.

In our experiment, subjects were presented with a visual scene (e.g. Figure 1) depicting all male (animal) characters involved in some joint action.¹ Subjects viewed the scene while listening to a short discourse about the scene (e.g. (2)), and were asked to ‘follow along’ (though no explanation of what it means to ‘follow along’ was given). 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, and this composite was recorded along with the simultaneous audio information onto a digital video tape with a sampling rate of 30 fps. The data were then manually analyzed by coding which object in the scene the subject was fixating at each successive frame in the utterance.²

We designed discourses (e.g. (2)) in which a resemblance relation is inferred between U_{i-1} (2.2) and U_i (2.3a/b), similar to (1) above.

¹Animals were used as characters in order to avoid potential biases due to stereotypes associated with roles such as *the doctor* or *the patient*, etc. All animals were introduced as males at the start of the experiment.

²See [18] for full details of the experiment design and procedure.

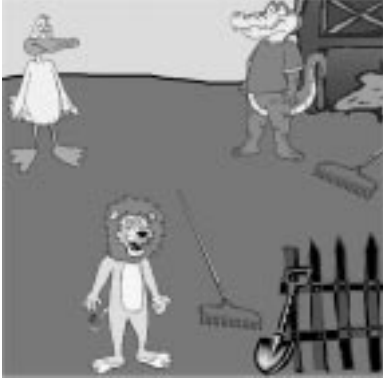


Figure 1: Visual scene paired with the discourse in (2).

- (2) 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. (UNACC PRO)*
 - 3b. *Then HE [L+H* L-] hit the duck. (NUC-ACC PRO)*
 4. *A big fight ensued and it was a terrible scene.*

Figure 2 shows fixation probabilities during the auditory presentation of the target utterance containing the nuclear-accented pronoun (2.3b). At each successive frame in the utterance, the probability that a given object is fixated was calculated across subjects and discourse items. All plots are time-aligned at the utterance onset. Only the three characters are plotted here for ease of presentation: N1 (*lion*), N2 (*alligator*), and N3 (*duck*). At the onset of the utterance, we find that subjects tend to be fixating on the N2 character (a carry-over effect from the previous utterance).³ Upon hearing the accented pronoun, N2 remains highly activated (i.e. ‘fixated’), and fixations on N1 also increase.⁴ As additional propositional information is encountered in the speech stream, preference for N2 is maintained, while fixations on N1 soon drop off.

The fixation data suggest that upon hearing the nuclear-accented pronoun, listeners evoke the contrast set containing both N1 and N2. At this point interpretation is still indeterminate. The listener must wait until subsequent propositional information is encountered to infer the speaker’s intended coherence relation. At the verb (which is identical across clauses), the listener has mounting evidence to infer a resemblance relation, and thus the contrasting proposition *C* (for which *X* in *X hit the duck* is a point of difference) can most easily be found in the previous clause, without need for further accommodation. Therefore, N2 is evoked as the referent of the accented pronoun.

³The mention of the instrument at the end of utterance (2.2) was intended to draw fixations away from N2 before *HE* is mentioned. This apparently was not totally successful. Therefore, it is impossible to know the exact point at which N2 starts to be considered as a potential referent of *HE*. We are currently running a follow-up experiment to correct this problem.

⁴A number of studies have observed an approximately 200ms delay (6 frames) between the speech cue and the launch of a saccade to the visual object to which the speech refers (e.g. [16], among many others). This is attributed to the time it takes for the oculomotor system to program a saccade, and should be kept in mind when viewing the graphs here.

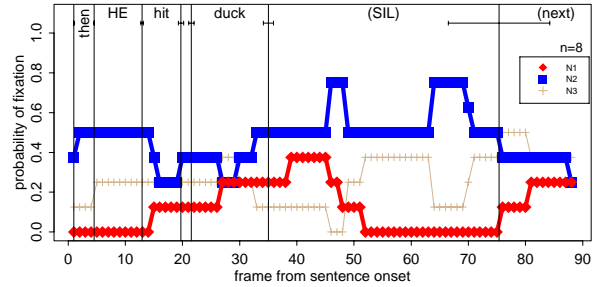


Figure 2: Probability of eye fixations on objects in scene while hearing utterance (2.3b): e.g. *Then HE hit the duck.*

4. Ambiguities of coherence

There are cases in which there is ambiguity about whether the contrasting proposition can be found in the immediately preceding clause. Take for example the discourse in (3), paired with the scene in Figure 3.

- (3) 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 [L+H* L-] got out some sponges.*
 4. *And together they started washing the hood and the fenders.*



Figure 3: Visual scene paired with the discourse in (3).

Here, (3.2) and (3.3b) could be compatible with a resemblance relation, since both describe a contribution to washing the car. However, the evidence that this resemblance is intended to structure the discourse is weaker. Instead, the listener may prefer an occasion relation in this case. Our account predicts that listeners may not necessarily resolve the presupposition of contrast to U_{i-1} in these cases (which would result in $HE=N2$), since the occasion relation involves a strong subject (N1) preference. Listeners may instead prefer to accommodate the presupposition outside the immediate discourse context.

Figure 4 shows the fixation data for the utterance containing the accented pronoun (3.3b). We have collapsed fixations on each character and his associated object (*sponges*) for ease of presentation. Again we observe that both N1 and N2 are evoked immediately upon hearing the nuclear-accented pronoun. However, in contrast to the resemblance case in (2), here a preference for N1 emerges in the fixation data as the verb and subsequent

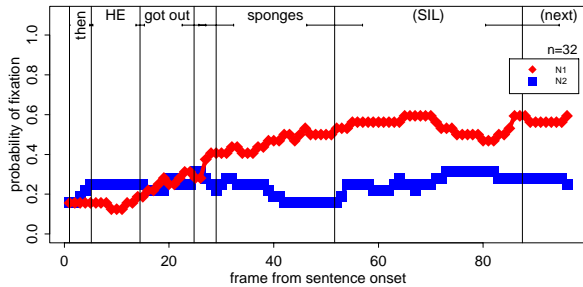


Figure 4: Probability of eye fixations on objects in scene while hearing utterance (3.3b): e.g. *Then HE got out some sponges.*

information is encountered. We suggest that this preference for N1 (albeit with some competition from N2; a 2:1 ratio) is due to the potential ambiguity in resolving the presupposition of contrast to U_{i-1} in the occasion relation. That is, listeners may take U_{i-1} to be the contrasting proposition, or they may be able to accommodate another contrasting proposition from somewhere outside the immediate discourse context. The lack of strong evidence for a resemblance relation in these cases contributes to the ambiguity, making it difficult for listeners to take *HE* as referring uniquely to N2; it may in fact refer to the salient N1 if the contrast is found elsewhere in the discourse. Thus, our account predicts that interpretation may remain indeterminate in such contexts.⁵

5. Identification of contrast

The above discussion suggests that accented pronoun interpretation will depend on an interaction between inferred coherence relations and the ability to resolve the presupposition of contrast evoked by the narrow focus accent locally. To further test this prediction, our experiment also included discourses in which contrast between U_{i-1} and U_i would not conflict with the default inferred occasion relation between them. Consider the variation on (3.2) given in (4.2').

- (4) 2'. *The zebra told the pig to put a bucket of soapy water near the front of the car.*
 3b. *Then HE [L+H* L-] got out some sponges.*

The fixation data in this condition (not included here) show that listeners prefer N1 as the antecedent of *HE*, with no competition from N2. Under the current proposal, this strong preference is due to the fact that, since (4.2') describes a contribution of N2, either resolution of the presupposition of contrast locally to U_{i-1} or accommodation will support the default interpretation (N1). Therefore, since both possibilities converge on the same antecedent, there is no inherent ambiguity in these cases.

6. Conclusion

Data from our experiment reveal two important aspects of accented pronoun interpretation. First, the fixation data suggest that uttering a pronoun with a (narrow focus) nuclear L+H* accent evokes a contrast set containing both salient referents in the local attentional state. Fixations on both N1 and N2 are observed immediately after the pronoun is uttered, across discourse conditions. Second, accent alone is not sufficient to

switch reference to a less salient entity. Rather, the type of inferred coherence relation and the ability of the listener to resolve the presupposition of contrast locally has a significant effect on interpretation. This interaction of the meaning of the accent with the inferred coherence relation determines whether not the accent will serve to switch reference, or will cause ambiguity in interpretation.

7. References

- [1] Akmajian, A.; Jackendoff, R., 1970. Coreferentiality and stress. *Linguistic Inquiry* 1(1), 124–126.
- [2] Lakoff, G., 1971. Presupposition and relative well-formedness. In *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*, Steinberg, D.D.; Jakobovits, L.A., (eds.). Cambridge University Press, 329–340.
- [3] Oehrle, R.T., 1981. Common problems in the theory of anaphora and the theory of discourse. In *Possibilities and Limitations of Pragmatics*, Parret, H., et al., (eds.). John Benjamins, 509–530.
- [4] Solan, L., 1983. *Pronominal Reference: Child Language and the Theory of Grammar*. D. Reidel Publishers.
- [5] Smyth, R., 1994. Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research* 23(3), 197–229.
- [6] Kameyama, M., 1999. Stressed and unstressed pronouns: Complementary preferences. In *Focus: Linguistic, Cognitive, and Computational Perspectives*, Bosch, P.; van der Sandt, R., (eds.). Cambridge University Press, 306–321.
- [7] Kehler, A., 2001. *Coherence, Reference and the Theory of Grammar*. CSLI Publications.
- [8] Rooth, M., 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1), 75–116.
- [9] Webber, B.; Knott, A.; Stone, M.; Joshi, A., 1999. Discourse relations: A structural and presuppositional account using lexicalised TAG. In *Assoc. for Computational Linguistics*, 41–48.
- [10] Grosz, B.J.; Sidner, C.L., 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3), 175–204.
- [11] Grosz, B.J.; Joshi, A.K.; Weinstein, S., 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2), 203–225.
- [12] Hobbs, J.; Stickel, M.; Appelt, D.; Martin, P., 1993. Interpretation as abduction. *Artificial Intelligence* 63, 69–142.
- [13] Cahn, J., 1995. The effect of pitch accenting on pronoun referent resolution. In *Assoc. for Computational Linguistics*, 290–293.
- [14] Nakatani, C.H., 1997. Discourse structural constraints on accent in narrative. In *Progress in Speech Synthesis*, van Santen, J.P.H., et al., (eds.), Springer-Verlag, 139–156.
- [15] Cooper, R.M., 1974. 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.
- [16] Tanenhaus, M.K.; Spivey-Knowlton, M.; Eberhard, K.M.; Sedivy, J.C., 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634.
- [17] Arnold, J.E.; Eisenband, J.G.; Brown-Schmidt, S.; Trueswell, J.C., 2000. The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition* 76, B13–B26.
- [18] Venditti, J.J.; Stone, M.; Nanda, P.; Tepper, P., 2001. Toward an account of accented pronoun interpretation in discourse context: Evidence from eye-tracking. Technical Report, Rutgers Center for Cognitive Science.

⁵We observe the same 2:1 ratio in preferences in off-line judgments of these discourses as well.