ELSEVIER

Cognition 71 (1999) 109–147

COGNITION

Achieving incremental semantic interpretation through contextual representation

Julie C. Sedivy[a,*], Michael K. Tanenhaus[b], Craig G. Chambers[b], Gregory N. Carlson[b]

[a]*Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI 02912, USA*
[b]*University of Rochester, Rochester, NY, USA*

Received 3 September 1998; accepted 9 April 1999

Abstract

While much work has been done investigating the role of context in the incremental processing of syntactic indeterminacies, relatively little is known about online semantic interpretation. The experiments in this article made use of the eye-tracking paradigm with spoken language and visual contexts in order to examine how, and when listeners make use of contextually-defined contrast in interpreting simple prenominal adjectives. Experiment 1 focused on intersective adjectives. Experiment 1A provided further evidence that intersective adjectives are processed incrementally. Experiment 1B compared response times to follow instructions such as '*Pick up the blue comb*' under conditions where there were two blue objects (e.g. a blue pen and a blue comb), but only one of these objects had a contrasting member in the display. Responses were faster to objects with a contrasting member, establishing that the listeners initially assume a contrastive interpretation for intersective adjectives. Experiments 2 and 3 focused on vague scalar adjectives examining the time course with which listeners establish contrast for scalar adjectives such as *tall* using information provided by the head noun (e.g. *glass*) and information provided by the visual context. Use of head-based information was examined by manipulating the typicality of the target object (e.g. whether it was a good or poor example of a *tall glass*. Use of context-dependent contrast was examined by either having only a single glass in the display (the no contrast condition) or a contrasting object (e.g. a smaller glass). The pattern of results indicated that listeners interpreted the scalar adjective incrementally taking into account context-specific contrast prior to encountering the head. Moreover, the presence of a contrasting object, sharply reduced, and in some conditions completely eliminated, typicality effects. The results suggest a language processing system in which semantic interpretation, as well as syntactic processing, is conducted incrementally, with early integration of contextual information.

Keywords: Language processing; Semantic interpretation; Prenominal adjectives

* Corresponding author. Tel.: +1-401-863-3383; fax: +1-401-863-2255;
E-mail address: julie_sedivy@brown.edu (J.C. Sedivy)

0010-0277/99/$ - see front matter
PII: S0010-0277(99)00025-6

1. Introduction

There is growing evidence that language processing is highly incremental, that is, processing occurs with very little lag following the input (e.g. Marslen-Wilson, 1975; Frazier, 1987; Steedman, 1989). Such evidence stands in contrast to early studies which suggested that meaningful syntactic and semantic processing was delayed until a certain amount of information had accrued, perhaps as late as until a clausal boundary had been encountered and identified (Carroll and Bever, 1978). The property of incrementality has important implications for the human language processing system. While it has obvious advantages, such as avoiding burdening short-term memory with unstructured linguistic material, incremental processing results in a specific problem. In particular, a high degree of incrementality gives rise to extensive temporary indeterminacies that must be resolved by the processing system, where utterances that may be unambiguous in their entirety are temporarily ambiguous at some point in the input. This problem has been the focus of a large body of work in sentence processing, with various hypotheses forwarded to explain the mechanisms used by the processing system to handle widespread temporary ambiguity. Most of this work has focused on a wide range of temporary syntactic ambiguities, as illustrated by the now famous example below:

(1) The horse raced past the barn fell.

(2) The horse raced past the barn and fell.

These two sentences are characterized by very different syntactic structures, but are superficially identical until the word following *barn*. Thus, if processing is highly incremental, decisions regarding the structure of the input string must be considered before the point in the string where there is sufficient information to rule out all but the correct alternative. The fact that sentence (1) creates much more processing difficulty than sentence (2) is generally attributed to the consequences of incorrect early decisions on the part of the processing system.

However, the indeterminacy problem is not limited to strings with ambiguous syntactic structure. In addition to computing the correct syntactic structure for a linguistic string, the processing system must ultimately be able to compute a meaningful interpretation of the string. We assume that an important part of semantic interpretation consists of a mapping from linguistic expressions to entities in the real world, or more accurately, some partial representation of the world (i.e. a *model*). The specific characteristics of the model are crucial for the resolution of reference. To see how problems of referential indeterminacy may arise, consider the following simple example: suppose two people are jointly engaged in a task involving building blocks, and one person utters the request: ‘Please hand me the large red rectangular block.’ If interpretation of the referential expression *the large red rectangular block* is undertaken incrementally, the properties of the other blocks in the array will determine the degree of indeterminacy involved in processing the instruction. For instance, if the array of blocks contains several large blocks, but only one block that is both large and red, then it is possible to isolate the target object in the display

following the word *red*. However, if the array contains multiple blocks that are both large and red, but a unique block that is large, red *and* rectangular in shape, then the instruction is indeterminate with respect to multiple blocks at *the large red...*, with the indeterminacy resolvable only at the following word, *rectangular*.

Empirical evidence for precisely this level of word-by-word incrementality in mapping language to the model comes from a series of experiments reported in Eberhard et al. (1995). These studies used an experimental paradigm in which subjects were given spoken instructions to manipulate a set of real objects in a workspace, while their eye movements to the objects in the visual display were monitored throughout the instruction. The identity of the objects in the model was manipulated in such a way as to vary the point in the speech stream where the referential expression became unambiguous (see Fig. 1). For instance, an example instruction might be 'Touch the plain red square.' In the early disambiguation condition, the visual array of objects presented to a subject consisted of three objects marked with a star, and a single object with no marking. The mid-disambiguation condition had a display consisting of four plain objects, only one of which was red. Finally, in the late disambiguation condition, the visual array consisted of four plain blocks, two of which were red. Of the red blocks, one was square in shape, and the second was rectangular.

Thus, by manipulating the displays, it was possible to alter the point in the input string which allowed for the identification of a unique referent compatible with the instruction, with disambiguation occurring at the first adjective (*plain*) for the early condition, the second adjective (*red*) in the mid condition, and only at the final noun (*rectangle*) in the late condition. Analysis of the eye movement record showed eye movements occurring generally well before the end of the referential expression. More interestingly, the eye movements were closely time-locked to the point in the speech stream where it became possible to pick out a unique object from among the alternatives in the display. When the point of disambiguation was identified for each of the conditions, it was found that eye movements were launched generally within 0.5 s of the beginning of the disambiguating word. It is estimated that the programming of an eye movement actually begins roughly 200 ms before it is launched

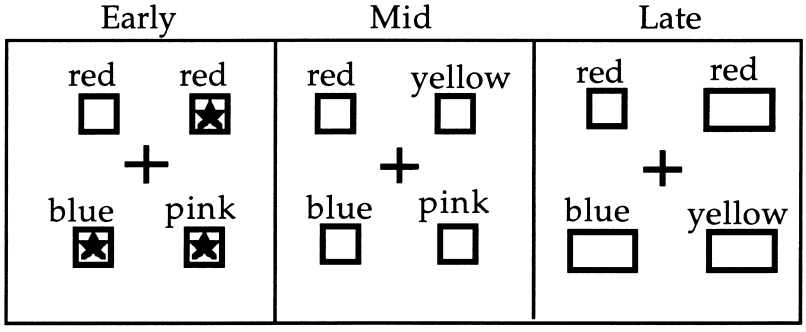

Fig. 1. Example displays from conditions manipulating the point at which a spoken instruction becomes unambiguous with respect to its referent. The accompanying instruction to this example was 'Touch the plain red square'.

(Matin et al., 1993). Thus, subjects were typically initiating saccades within 300 ms of the onset of the disambiguating word, often before the end of that word.

This experiment provides evidence that, like the processing of structural representations for a linguistic string, the process of establishing reference is incremental, resulting in local indeterminacies. It appears that subjects actively consider all the referents that are compatible with the linguistic input at a particular point in time, continuously narrowing the set of possible referents until it is possible to identify a singleton set. Clearly, the information provided in the visual model is of primary importance, with the resolution of reference involving a continuous integration of the linguistic information together with information pertaining to the model.

The Eberhard et al. study provides direct evidence for a view of language processing in which incremental referential processing is central, a view described by Altmann and Steedman (1988) below:

> The process of incremental evaluation involves having available representations of 'partially evaluated' referents. These are simply the members of the set of referents which satisfy the available constraints. This set gradually becomes more and more refined as the analysis proceeds, until just the candidate referent remains (Altmann and Steedman, 1988, p. 196).

In addition to claiming that, like syntactic processing, referential processing is highly incremental, proponents of what has come to be known as the Referential Theory of sentence processing (e.g. Crain and Steedman, 1985; Altmann and Steedman, 1988) have made the controversial claim that referential processing is implicated in the resolution of local syntactic ambiguities, such as the ambiguity illustrated in sentences (1) and (2), reproduced below:

(2a) The horse raced past the barn fell.

(2b) The horse raced past the barn and fell.

Psycholinguistic studies of such ambiguities have demonstrated a clear preference for the structure in (2) over the one in sentence (1). This preference has frequently been interpreted as a preference for the syntactically simpler option when more than one structure is possible for a particular input string (Rayner et al., 1983). However, Crain and Steedman (1985) and Altmann and Steedman (1988) have argued that syntactic differences are confounded with crucial referential differences. Specifically, they point out that the more difficult sentence (1), which involves a reduced relative clause modifying the noun horse, results in the complex subject noun phrase *The horse raced past the barn*. In contrast, sentence (2) has the simple subject noun phrase *The horse*. Steedman et al. have argued that complex noun phrases such as the one in (1) presuppose a richer representation of entities in the discourse model than simple noun phrases. That is, modification of a definite noun phrase presupposes the existence in the model of an entity or set of entities that is of the same category as the head noun, but that contrasts with respect to the property encoded by the adjective.

Thus, a complex noun phrase such as *The horse raced past the barn* presupposes the existence of two or more horses, only one of which has the property of being

raced past the barn. The simple noun phrase *the horse*, on the other hand, merely requires the instantiation of a single entity that has the property of being a horse. Empirical evidence has shown that by manipulating the context of an utterance, and therefore the model that is instantiated prior to the temporarily ambiguous string, it is possible to shift preference for simple versus complex referential expressions (Crain and Steedman, 1985; Altmann and Steedman, 1988; Altmann et al., 1992; Altmann et al., 1994; Britt, 1994) (but cf. Mitchell et al., 1992). Until recently, it has been possible to demonstrate support for incremental referential processing only indirectly, by examining, as the studies cited above have done, cases where referential factors correlate with syntactic ambiguities, and observing effects of referential manipulations on syntactic preferences. The Eberhard et al. (1995) study illustrates an experimental paradigm in which referential processing can be investigated more directly, by observing the entities in the visual model that elicit eye movements as the utterance unfolds. This paradigm has been used to corroborate the results of studies investigating the syntactic consequences of referential factors. Tanenhaus et al. (1995) report a study using spoken utterances with the head-mounted eyetracking paradigm, showing no evidence of difficulty with a temporarily ambiguous instruction when the visual model supports the more complex referential expression (and hence the 'dispreferred' syntactic structure).

Results such as these provide compelling support for a theory of language processing which accords a central role to continuous referential processes. However, when one surveys a broader range of linguistic expressions, there is reason to suspect that the process of mapping expressions to a model should in fact display limited, rather than continuous incrementality, a point argued by Clifton and Ferreira (1989):

> We doubt that Altmann and Steedman's suggestion will prove to be adequate. It may be attractive to think in terms of progressively narrowing sets of referents for NPs with possible post-nominal modifiers. However, referential narrowing is far less plausible for other syntactic categories. To make just one argument, consider adjectival modifiers. The interpretation of an adjective (e.g. 'red') generally depends upon its head noun (compare 'red hair' and 'red truck'). As an extreme case, consider the adjectives 'good' and 'big' and 'fake'. These adjectives do not permit narrowing of the set of referents prior to the receipt of their head noun (Clifton and Ferreira, 1989, p. 86).

The semantic dependency of a linguistic unit on some other, later-occurring linguistic expression is by no means exceptional in natural language. It occurs not only with expressions that identify properties, or sets of entities, such as *red* or *tall*, but also with expressions that convey relations that hold between two entities. Consider, for instance, the difficulty in precisely establishing the relation expressed by the preposition independently of the noun that follows it in the phrases *on the table/on Thursday*, and *in the box/in the grass*. Similar problems are encountered with many verbs, where the same verb can encode somewhat different relations, depending in part upon the object of the verb: *John loved the Englishwoman/John loved the Viennese torte*, and *Suzanne sent the letter/Suzanne sent the messenger*. Examples such as these highlight the difficulties that would be encountered by a

processing system in which semantic interpretation is carried out on an incremental word-by-word basis. One might conclude in all of these cases, that semantic interpretation should be delayed until some point in the phrase, perhaps at the head noun for adjectives, and at the complement noun for prepositions or verbs.

In this article, we will pursue an alternative hypothesis, namely, that there are no principled limits on the degree of incrementality of semantic processing. We will argue that in cases of local indeterminacy, information from the context of the utterance can be used to pin down the meaning of a linguistic expression. Our experimental investigations here focus on the interpretation of adjectives, for two reasons: First, the problem of referential indeterminacy has been argued to be fairly extensive for adjectives, and second, existing literature provides some insight into possible contextually-based mechanisms that might be exploited by a processing system in interpreting incrementally.

1.1. The semantics of adjectives and implications for incrementality

Formal semantic accounts involving adjectives illustrate the difficulty of providing a unified characterization of adjectival meaning (see Kamp and Partee, 1995, for an excellent overview of semantic issues in the formal representation of adjectives). Most important for considerations of incremental semantic processing are the varying degrees to which adjectives display a stable core meaning of their own. Kamp and Partee provide a rough classification of adjectives into three broad categories. The simplest category includes adjectives that have the most stable core meanings, and has frequently been referred to as the class of *intersective* adjectives. Under a classical model-theoretic view of semantic representation, the meaning of such adjectives (as well as nouns and other one-place predicates such as intransitive verbs), can be identified with the set of entities in the model that bear the particular property encoded by the predicate. Thus, the meaning of the word *red* corresponds to the entities in the model that have the property of being red. Using this notion of predicate meanings, it is easy to see how simple expressions can compose into more complex expressions. The terms *plain*, *red* and *square* each denote sets of entities bearing a particular property, and the meaning of the complex expression *plain red square* is simply the intersection of all of these sets, hence the label *intersective*.

However, it is well known that many adjectives do not, in fact, conform to the simple compositional analysis described above. Treating modified adjective-noun pairs as an intersection of predicates only works in those cases where the adjective has a stable and independent core meaning. While this is arguably true of many adjectives, such as *red*, or *striped*, a variety of adjectives appears to lack an invariant meaning. It has often been observed, for instance, that many adjectives depend upon the noun that they modify for their meaning. Consider for example, the sentences below:

(3a) Tom is a good priest. He is caring and empathic.

(3b) Tom is a good lawyer. He is ruthless and perseverent.

The adjective *good* conveys a very different set of attributes depending upon whether it is modifying *priest* or *lawyer*. It is impossible to pick out a set of entities in the model that correspond to the predicate *good*, independently of the sets of entities corresponding to *priest* and *lawyer*. In fact, it has been argued that adjectives such as *good* can only be evaluated with respect to the set denoted by the head noun such that the set denoted by the phrase *good N* is necessarily a subset of the set denoted by *N*. For this reason, adjectives such as this are frequently called *subsective* adjectives, as compared to intersective adjectives like red. The difference in the independence of adjectival meanings can also be illustrated by the following reasoning test, where the argument in (4a) is valid for intersective adjectives, but not for the subsective adjective in (4b):

(4a) Ben is a bald man. Ben is an accountant.

Therefore, Ben is a bald accountant.

(4b) Tom is a good priest. Tom is a fisherman.

Therefore, Tom is a good fisherman.

A third class of adjectives, termed *non-subsective* by Kamp and Partee, refers to those adjectives that are typically restricted in such a way that the phrase *adj. N* can not be a subset of the set denoted by *N*. An example of such an adjective is *fake*, where, for instance, a *fake gun* is typically not among the set of entities denoted by *gun* at all.

Considering the flexibility of meaning that is associated with many adjectives, a question arises with respect to the online processing of adjective-noun pairs: what are the implications of semantic elasticity, and particularly, the dependence of adjectival meaning upon the head noun, for an incremental processing system? The eyetracking study by Eberhard et al. (1995) provides compelling evidence that there is no general architectural constraint against incremental processing of adjectives prior to the head noun. However, the adjectives used in that study were generally of the sort for which it is relatively easy to identify a meaning independent of the modified noun, such as color adjectives. Incremental processing of subsective adjectives would presumably depend largely on immediate accessibility to information pertaining to the head noun.

To make matters more interesting, adjectives are frequently not only dependent on the head noun for their meaning, but also on aspects of the context of utterance. Compare, for instance, the meaning of the phrase *really tall snowman* in the following sentences, taken from Kamp and Partee (1995):

(5a) My 2-year-old son built a really tall snowman yesterday.

(5b) The D.U. fraternity brothers built a really tall snowman last weekend.

In fact, the fluidity of adjectival meanings in context has led some researchers to suggest that so-called subsective adjectives are not subsective at all, but simply heavily context-dependent. That is, rather than having their meanings fixed with

respect to the head noun, they introduce free parameters which are fixed relative to some salient aspect of the context, of which the set of entities introduced by the head noun is simply one factor. Consider, for instance, the following text, in which the phrase *good linguist* is used (Pollard and Sag, 1994 p. 330):

(6) The Linguistics Department has an important volleyball game coming up against the Philosophy Department. I see the Phils have recruited Julius to play with them, which means we are in real trouble unless we can find a good linguist to add to our team in time for the game.

The standard of goodness here seems to be determined by the contextually relevant parameter of goodness-as-a-volleyball-player, rather than with respect to the head noun *linguist*. That is, the quality of linguistic scholarship appears to be wholly irrelevant to the interpretation of *good linguist* in this case.

As pointed out by Kamp and Partee, many adjectives do not fit clearly into one category as opposed to another. All adjectives appear to exhibit some degree of susceptibility to shifts in meaning due either to the head noun they are modifying or the context of use, though the degree of sensitivity may differ. Instability of adjectival meaning can be observed even for adjectives that are generally considered to fall squarely into the intersective category, such as color adjectives, as is shown by the different meanings of *red* in the phrases *red car*, *red hair*, and *red cabbage*. Such shifts in meaning have been established experimentally for color adjectives (Halff et al., 1976) as well as for adjectives that display greater vagueness and context-sensitivity in general, such as scalar adjectives like *tall* (Maloney and Gelman, 1987).

If it is indeed the case that context has far-reaching implications for apparently different classes of adjectives, it will become crucial for theories of meaning to begin to specify the mechanisms for contextual influence on meaning. One such attempt is made by Bierwisch (1987) in a formal analysis of scalar adjectives. According to Bierwisch, scalar adjectives must always be understood with respect to some relevant comparison class. The meaning of a scalar adjective is characterized as a relation which assigns an entity to a value on some dimensional scale. The value on the scale can be specified numerically, or a range of values can be set relative to some norm that is fixed with respect to the comparison class. Thus, the meaning of the sentence in (7a) can be paraphrased as in (7b):

(7a) Hans is tall.

(7b) The value for height that corresponds to Hans is greater than some norm for a relevant comparison class.

The fluidity of the meanings of scalar adjectives comes from the various possibilities for establishing the relevant comparison class. One of the most common ways of fixing the comparison class is with respect to the class of entities denoted by the head

noun. It can also be set to correspond to a subset of entities denoted by the head noun (as well as a set of entities that is broader than the set denoted by the head noun). Under this view, there is no real distinction between cases where the value for scalar adjectives is set relative to the head noun, and cases where it is contextually determined. Presumably, however, there are at least somewhat systematic correlations between linguistic form and the method for fixing the comparison class. Some of the linguistic factors discussed by Bierwisch include whether a noun phrase has particular or generic reference, and whether it involves modification that is restrictive or non-restrictive.

One of the implications of this general approach is that it should be possible to fix a value for the scalar adjective as soon as some comparison class becomes available; because this need not be accomplished strictly with respect to the head noun, there is no principled reason why the interpretation should not be incremental. Thus, given sufficient relevant contextual information, it should be possible to fix a value for the scalar adjective prior to encountering the head noun.

The emphasis on the contrastive nature of adjectives relative to some comparison class is particularly appealing given evidence from the language processing literature for sensitivity to contextually-available contrast in online processing. Most of the work has focused on the contextual implications of other nominal modifiers such as relative clauses, and prepositional phrases. As discussed in the introductory section above, these studies have tested the hypothesis that modifiers of nouns convey contrastive information, that is, a modified NP such as *The horse raced past the barn* is most naturally used in contexts where the modifying phrase *raced past the barn* provides information that contrasts the referent of the modified noun phrase with some other entity or entities of the same category (e.g. *horse*). Studies manipulating the referential context have generally focused on changing the entities that are introduced into the discourse prior to the target sentence, such that some contexts provide a contrasting entity for the modified noun, and thereby support the contrastive use of the modifier, while other contexts do not.

Contextual manipulations of this sort have frequently been shown to affect the online parsing preferences for temporarily ambiguous sentences involving reduced relative clauses and prepositional phrases (Altmann and Steedman, 1988; Altmann et al., 1992; Altmann et al., 1994; Britt, 1994; Spivey-Knowlton and Tanenhaus, 1994) (but see Ferreira and Clifton, 1986; Britt, 1994; Murray and Liversedge, 1994). Similar contextual manipulations, but in a visual context, have been demonstrated to have effects on processing syntactic ambiguities involving ambiguously attached prepositional phrases in Tanenhaus et al. (1995). Furthermore, model-based effects have been demonstrated sentence internally by manipulating whether the critical NP was introduced by the definite article *the*, or the indefinite article *a*, capitalizing upon the presuppositional differences of definite and indefinite NPs (Spivey-Knowlton and Sedivy, 1995; Schelstraete, 1996).

Adjectives appear to have a contrastive function similar to that of other modifying phrases. The contrastive use of adjectives is illustrated in the following example, taken from an actual piece of film dialogue, in which a character refers to a group of

women as 'angry feminists', and subsequently comments on his use of that term by saying:

(8) And I say angry feminists like there's some other kind.

(*The Last Supper*, 1995, Tristar Pictures).

Here, the character seems to be aware that by using an adjective, he has evoked a contrasting set of feminists who have some property distinct from being angry, and then denies that such a contrast actually exists.

Experimental evidence for the interpretation of adjectives as contrastive exists as well, primarily manifest in interactions with so-called focus-sensitive expressions such as *only*, which have been argued to require contrast for interpretation (Ni et al., 1996). In addition, there is evidence from language acquisition indicating that children use their knowledge of the contrastive use of adjectives to constrain the learning of new words (Gelman and Markman, 1985).

The contrastive function of adjectival modifiers is very similar to the presupposition of contrast that is required for the interpretation of sentences with focus. For instance, a sentence such as *The STUDENTS in this department are happy*, with contrastive stress on *STUDENTS*, seems to invoke a contrast between the happiness of students on the one hand, and the happiness of other departmental members such as faculty or staff, on the other hand. There is experimental evidence that, for sentences marked intonationally with focus, the computation of sets of contrasting entities in a context is accomplished rapidly enough to be useful to online processing during the course of interpreting a referential indeterminacy. In an eyetracking study by Sedivy et al. (1995), it was established that contrastive information conveyed by intonational focus could be used to narrow the set of alternative referents in the visual display. The experiment used displays consisting of sets of four colored cardboard shapes. Two of these shapes differed minimally from each other with respect to size (e.g. a large and small red circle). Instructions all contained referential phrases with the following sequence: size adjective, color adjective, noun. They varied as to whether they were uttered with neutral intonation or had intonational focus on the size adjective (e.g. 'Touch the LARGE red circle'). The displays were constructed such that, on the basis of the content of the referential phrase alone, the point at which a unique referent could be picked out was at the color adjective. However, subjects were able to make use of the contrastive information associated with the focused adjective to narrow down the set of possible referents to just the set of entities containing the target and its contrast, resulting in an earlier point of disambiguation. Thus, eye movements to the target were launched significantly faster when the instruction contained contrastive focus intonation than when it did not.

Experimental evidence along these lines illustrates the accessibility of referentially-based contextual information, and lends force to the hypothesis that incremental interpretation of adjectives that lack an independent denotation can be achieved by means of integrating information about contextually-defined contrast. The experiments in this paper represent an empirical evaluation of such a mechan-

ism. The goal of this article was to examine evidence for incrementality and use of contextually-defined constraints for both intersective and non-intersective adjectives. Our contextual manipulations made use of contextually defined contrast, which is described briefly in the next section.

The experiments in this article made use of the eyetracking paradigm with spoken language and visual contexts in order to examine how, and when, listeners make use of contextually-defined contrast in interpreting prenominal adjectives. Experiments 1A and 1B focused on intersective adjectives. Experiment 1A provided further evidence that intersective adjectives are processed incrementally, replicating the basic findings of Eberhard et al. (1995). Experiment 1B compared response times to follow instructions such as 'Pick up the blue comb' under conditions where there were two blue objects (e.g. a blue pen and a blue comb), but only one of these objects had a contrasting member in the display. Responses were faster to objects with a contrasting member, establishing that the listeners initially assume a contrastive interpretation for intersective adjectives.

Experiments 2 and 3 focused on vague scalar adjectives examining the time course with which listeners establish contrast for scalar adjectives such as *tall* using information provided by the head noun (e.g. *glass*) and information provided by the visual context. Use of head-based information was examined by manipulating the typicality of the target object (e.g. whether it was a good or poor example of a *tall glass*). Use of context-dependent contrast was examined by either having only a single glass in the display (the no contrast condition) or a contrasting object (e.g. a smaller glass). The pattern of results indicated that listeners interpreted the scalar adjective incrementally, taking into account context-specific contrast prior to encountering the head. Moreover, the presence of a contrasting object sharply reduced, and in some conditions completely eliminated, typicality effects.

2. Experiment 1: evidence for the use of contextual contrast in interpreting adjectives online

Experiment 1 was designed to investigate two specific questions: first, the experiment investigated whether evidence for incremental processing would be obtained with less predictable experimental materials than those used by the previous Eberhard et al. (1995) eyetracking study. A second goal of the experiment was to establish the degree to which the contrastiveness of adjectival modifiers can be observed both with and without focus intonation. In the following discussion, these two questions will be treated separately as Experiments 1A and 1B, respectively, though the materials were presented together as part of the same experimental session.

2.1. Experiment 1A

With respect to the question of incrementality, results in Eberhard et al. (1995) indicate that linguistic expressions are evaluated against a model on an incremental

word-by-word basis (and perhaps finer). Experiment 1A represents an attempt to replicate this result using materials whose form is considerably less predictable than the materials used by Eberhard et al. Specifically, the Eberhard et al. study used instructions that were with respect to general form and content, with each referential phrase encoding marking, color and shape, in that order (e.g. 'Touch the plain yellow square.') In the current experiment, target instructions included only one adjective, which might refer to any one of a number of salient properties such as color, shape or material. In addition, the instructions were embedded within a set of filler instructions which included either a noun modified by an adjective (encoding color, shape, size or material) or a bare unmodified noun. On occasion, the same object appeared in numerous trials, with varying labels associated with it. Thus, for any target object, it was impossible to predict solely on the basis of the stimuli used in the experiment what the content of the referential phrase would be.

2.1.1. Subjects

Twelve undergraduate students from the University of Rochester participated as subjects in Experiment 1. Subjects were recruited by means of announcements posted on the university campus, and verbal announcements made in Cognitive Science courses, and were paid $7 per session. All subjects were native, monolingual speakers of English, and either had normal uncorrected vision, or wore soft contact lenses.

2.1.2. Materials and design

Experimental materials included ten target instructions involving a referential expression that included an adjectival modifier. Half of the experimental items included a color adjective, and half included an adjective referring to the material the object was made from, or the shape of the object. The displays were constructed such that for half of the target instructions, the point of disambiguation (i.e. the point at which there was sufficient lexical information to identify a single object as the target referent) was after the adjective. For instance, for an instruction such as 'Touch the blue pen', the display contained the target item, a yellow rubber duck, a red notebook and a pink comb. We will refer to these trials as belonging to the early disambiguation condition. For the remaining half of the displays (the late disambiguation condition), the point of disambiguation was at the noun, such that for the same instruction, the display now contained the target blue pen, a blue bowl, the yellow duck and the red notebook. Thus, each display in the late disambiguation condition contained an object that shared the property denoted by the adjective in the instruction. We will refer to this object as the competitor object.

Subjects heard each target instruction only once, with displays rotated such that half the subjects saw one of the possible displays, and the remaining half saw the second possible display. The target instructions were followed by two distractor instructions referring to other objects in the same display. In addition, the target trials were interspersed with 20 experimental trials from Experiment 1B (to be described below), and 10 distractor trials. Four practice trials preceded the set of 40 trials, each involving a different display of four objects. Some of the same objects

appeared in a number of different displays, in different arrays with other objects; however, care was taken to ensure that for critical trials, the target object had never been referred to in a previous trial, to ensure that subjects did not develop any expectations about how these objects would be described. There were a total of 44 display changes, with three instructions per display, for a total of 132 instructions. Of the 132 instructions, 68 involved an adjectival modifier, 56 involved a bare unmodified noun, and eight involved some other type of referential expression such as a pronominal form (e.g. 'Touch it again') or a noun phrase with a quantifying predeterminer (e.g. 'Touch one of the utensils'). The adjectival modifiers in the critical and filler trials referred either to color, material, shape, some scalar dimension (e.g. 'tall glass') or another salient property of the object (e.g. 'stuffed dog').

2.1.3. Procedure

Subjects were given spoken instructions to touch various objects on a horizontal workspace. The instructions were read aloud by the experimenter from a script. Display changes took approximately 20 s, and subjects were permitted to watch the display as it was being changed. Every display contained a centrally located fixation cross. Each trial began with a request for the subject to look at the cross, and subjects were instructed to rest their eyes on the central cross between instructions. This was done so that eye movements to the target objects could be measured from a default position that was equidistant to all of the objects in the display. Subjects were told simply to perform the instructions as naturally as possible.

While the subject followed instructions to touch objects in the workspace, eye movement data were recorded using a lightweight Applied Scientific Laboratories (ASL) head-mounted video-based tracking system. The camera provided an infrared image of the eye at 60 Hz, and determined monocular eye position by monitoring the locations of the center of the pupil and the cornea reflection. A scene camera was mounted on the side of the helmet, providing an image of the subject's field of view. Calibration was carefully monitored throughout each trial, and minor adjustments were occasionally made between trials. A VCR record was made for each experimental trial, consisting of the instructions spoken by the experimenter into a microphone, as well as the subject's moment-by-moment gaze fixation superimposed over the scene camera image. Because the scene camera was mounted onto the helmet itself, and moved with the subject's head, the VCR record took into account any head movements made by the subject, allowing for unrestricted head and body movements throughout the experiment. The configuration of the equipment and displays is illustrated in Fig. 2.

The timing of eye movements relative to information in the speech stream was computed as follows: eye movement data for trials in which the initial eye fixation was to the correct object was analyzed from the video tapes by identifying the beginnings and ends of critical words in the speech stream for each trial, and noting the time lapse between the critical speech points and the onset of an eye movement to the intended object. Eye movement latencies were measured from the onset of the noun in the target instruction.

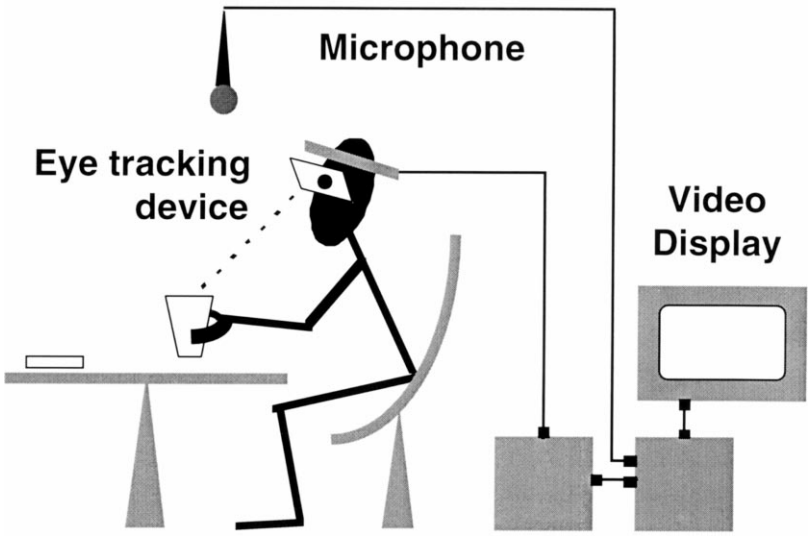

Fig. 2. An illustration of the configuration of the eyetracking equipment. Both eye image and scene image were taken in by camera mounted onto the headband. The CPU computed and superimposed the eye fixation over the scene image, with the resulting video data recorded by the VCR and displayed on the monitor. Experimental instructions were recorded via microphone directly onto the videotape by means of a frame-accurate editing VCR, which synchronized video and audio signals.

2.1.4. Results

Table 1 shows the mean eye movement latencies for the early disambiguation and late disambiguation conditions. Analysis of variance revealed the difference in eye movement latencies to be statistically reliable both in analyses by subjects ($F_1(1,11) = 11.58$, $P < 0.01$) and by items ($F_2(1,9) = 5.83$, $P < 0.05$).

On occasion, subjects would fixate an object other than the target prior to looking at the target object. Table 1 also shows the percentage of trials which include a look to an object other than the target at any point before the subject reached for the object, as well as the percentage of trials which include a look to either the competitor object for displays that had a competitor (i.e. the late disambiguation conditions) or the object in the same location in displays that did not have a competitor object (i.e. the early disambiguation condition). An analysis of variance was performed, and indicated that although the total percentage of trials including a look to

Table 1
Eye movement data for Experiment 1A, showing eye movement latencies as computed from the onset of the head noun, and the percentage of trials that included an eye movement to a non-target object

Condition	Eye movement latency (ms)	Looks to competitor or control object (%) Total looks to non-target objects (%)
Early	378	4.4416.26
Late	460	18.3321.66

any object other than the target was not reliably different for the two conditions, there were significantly more looks to the competitor object in the late disambiguation condition than there were to the object in the same location for the early disambiguation condition. This difference was reliable both by subjects ($F_1(1,11) = 8.31$, $P < 0.05$) and by items ($F_2(1,9) = 11.89$, $P < 0.01$).

2.2. Discussion

These results replicate the Eberhard et al. (1995) findings of incremental referential interpretation where linguistic expressions are continuously interpreted with respect to sets of entities available in a visual model. This is evident not only from the eye movement latencies, but the pattern of looks to non-target objects as well. Upon hearing the adjective, subjects considered as possible referents the set of objects in the display that bear the property denoted by the adjective. In the early disambiguation condition, this set is a singleton set, and hence the target object is identified as the referent at an earlier point than the late disambiguation condition. Additional, direct evidence for the activation of sets of alternatives comes from the fact that looks to non-target objects were dispersed among the objects in the display when there is no object bearing the property picked out by the adjective, but are concentrated on the competitor object for displays that do have such an object.

2.3. Experiment 1B

Experiment 1B was conducted in conjunction with Experiment 1A, and was designed to probe the contrastive information associated with adjectival modifiers. Previous experimentation with focus intonation (Sedivy et al., 1995) suggests that subjects are sensitive to the presence of an appropriate contrast when contrast is indicated intonationally. In this experiment, we used both modifiers with focus stress, and modifiers that were not marked with focus stress in an attempt to determine whether focus is necessary to induce a contrastive interpretation for modifiers.

2.3.1. Subjects

The subjects were the same as those for Experiment 1A, as both experiments were administered within the same experimental session.

2.3.2. Materials and design

Fig. 3 shows a sample display that was characteristic of all of the experimental displays. In all cases, the displays contained two objects of the same category, that differed with respect to a salient property. The property that varied was either color, shape or material. In addition, each display had an object that was of a different category than this minimal pair, but that shared a salient property with one member of this pair. In Fig. 3, for instance, there is a minimal pair consisting of two combs differing only in color (i.e. a pink comb and a yellow comb), an object that was of the same color as one of these combs (i.e. a yellow bowl), and a fourth object that was unrelated in any way to the other three (i.e. a metal knife).

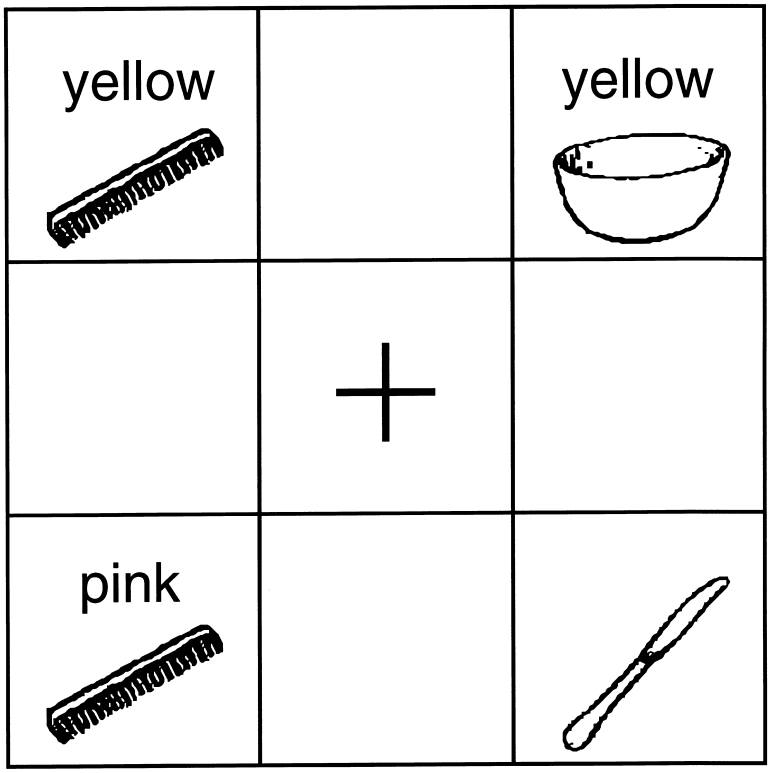

Fig. 3. A sample display for Experiment 1B, using common objects. Target objects were either the yellow comb (in the contrasting pair) or the yellow bowl (outside of the contrasting pair). For each instruction, the competitor object was the other object in the display that was yellow (e.g. for the critical instruction 'Now touch the yellow bowl', the competitor object was the yellow comb).

The first instruction always contained reference to one of the objects in the minimal pair (e.g. the pink comb). The second instruction, which was the critical instruction, involved a noun modified by a single adjective, and referred either to the object that contrasted with the first (e.g. the yellow comb), or the other object in the display that shared a crucial property with the contrasting object (e.g. the yellow bowl). This manipulation involving the referent was crossed with a manipulation

Table 2
Set of example instructions accompanying displays such as the one in Fig. 3, used for Experiment 1B

Condition	First instruction	Target instruction
Focus-contrast referent	Touch the pink comb.	Now touch the YELLOW comb.
Focus-non-contrast referent	Touch the pink comb.	Now touch the YELLOW bowl.
No focus-contrast referent	Touch the pink comb.	Now touch the yellow comb.
No focus-non-contrast referent	Touch the pink comb.	Now touch the yellow bowl.

involving contrastive stress, such that half of the critical instructions were produced with stress on the adjective (corresponding to a L + H* accent under the notational system of Pierrehumbert (1990)), and the other half had neutral intonation, with nuclear stress (H* accent) on the noun. Table 2 exemplifies the experimental manipulations that were carried out.

There were five items in each cell, resulting in 20 critical instructions. A third filler instruction accompanied each critical display. In addition, the materials from Experiment 1A, along with ten filler trials were interspersed with the critical trials. Four lists were constructed such that each subject heard only one set of instructions for each critical display.

It is important to note that the experiment was designed in such a way as to eliminate any internal bias towards the contrasting object for the critical trials. That is, the critical instruction referred equally frequently to the contrasting object, and the object that was not a member if the contrasting pair, but shared a property with one of its members. In addition, modifiers were used without any contrast present in the display at all, as was the case for all of the critical trials in Experiment 1A, and many of the filler trials as well. Thus, the experiment involved trials in which modifiers were used to signal contrastive information, trials in which no relevant contrast was present and trials in which a contrast was present in the display, but the modifier was not used in a way that related to the contrast between objects.

2.3.3. Procedure

The procedure was identical to Experiment 1A, as these two studies were carried out simultaneously.

2.4. Results

The results from both eye movement latencies and looks to non-target objects indicated a robust bias to interpret the modified nouns contrastively. However, there was little or no evidence that contrastive stress heightened the contrast effect as compared with instructions spoken with neutral intonation. Table 3 shows mean eye movement latencies to the target object as measured from the onset of the noun. The latencies are, in general, considerably shorter than those in Experiment 1A, which may be due to the fact that the target instructions in Experiment 1A always occurred as the first instruction accompanying each display, whereas in this experiment, they occurred as the second instruction. Display changes took approximately 20 s. Sub-

Table 3
Eye movement latencies in ms for Experiment 1B, as computed from the onset of the head noun

Condition	Contrast referent (ms)	Non-contrast referent (ms)
Focus	270	445
No focus	281	459

jects were permitted to watch the display as it was being changed. The data were submitted to 2 × 2 (referent by stress) repeated measures ANOVAs by subjects (F_1) and by items (F_2). Results of these analyses yielded a robust main effect of referent, such that instructions involving the contrasting object resulted in faster looks to the target object than instructions involving the competitor object. This difference was reliable in the analyses by subjects ($F_1(1,11) = 66.285$, $P < 0.001$) and by items ($F_2(1,19) = 78.869$, $P < 0.001$). There was no main effect of stress and no significant interaction of referent and stress.

Table 4 shows the percentage of trials that included a look to an object other than the target, indicating both percentage of trials including a look to any non-target object, and trials including a look to the competitor object (i.e. the object in the display that shared the property indicated by the adjective in the critical instruction). Note that when the target referent was a member of the contrasting pair, the competitor object was an object that was not a member of this pair; however, when the target referent was not a contrasting object, then the competitor object was a member of the contrasting pair. As can by seen in the table, virtually all looks to non-target objects were to the competitor object, and instructions in which the target object was not a member of the contrasting pair elicited a far greater number of looks to the competitor object, suggesting that the interpretation of the modifier favors a contrastive reading.

2 × 2 (referent by stress) ANOVAs (both by subjects and by items) were performed with the percentage of trials including a look to the competitor object as the dependent variable. The main effect of referent was significant by subjects $F_1(1,11) = 36.89$, $P < 0.001$) and by items ($F_2(1,19) = 17.83$, $P < 0.001$). There was no main effect of stress. Although the presence of contrastive stress resulted in a numerical increase in looks to the competitor object only when the target object was not the contrasting object, the interaction between stress and referent was not significant.

2.5. Discussion

Experiments 1A and 1B show evidence for incremental semantic processing with respect to a visually-available set of potential referents. In particular, these data indicate that nouns that are modified by adjectives are interpreted incrementally. In addition, there is evidence from Experiment 1B suggesting that people have immediate access to the contrastive function associated with adjectival modifiers,

Table 4
Percentage of trials in Experiment 1B that included a look to the 'competitor' object (i.e. the object in the display that shared the same property denoted by the adjective in the target instruction)

Condition	Referent: contrast member (%)	Referent: non-contrast member (%)
Stress	8	43
No stress	12	29

and that this information is used to resolve temporary referential indeterminacies in online processing. The immediate use of linguistically encoded contrastive information is consistent with previous work showing immediate effects of intonationally marked contrast (e.g. Sedivy et al., 1995). However, the results of Experiment 1B are somewhat surprising in their lack of an effect of contrastive stress over and above the use of an adjectival modifier. Although there is some suggestion of an effect of contrastive stress on looks to non-target objects, this effect is not statistically reliable, in sharp contrast the to effects of referent type. In fact, it is likely that the combined strength of the effect of referent and the weakness of any effect of stress reflects a ceiling effect of contrastive interpretation with adjectival modifiers in general. There are several reasons why the adjectives in this study, in the absence of contrastive stress, may have exerted a stronger effect of contrast than previous experiments involving contrastive stress. First, the Sedivy et al. (1995) experiments used adjectival modification in all of the instructions, both critical and filler trials. In fact, all of the instructions were of the same form, namely a size adjective, followed by a color adjective, and finally the head noun denoting the shape of the object. In the current experiment, however, many of the instructions did not involve modification, and used simple unmodified nouns as referential expressions. Furthermore, the adjectival modifiers picked out a variety of types of properties, not just color and shape. Thus, modification in and of itself, may have been viewed as more informative by the subjects. In addition, in this experiment, not only did the critical instruction involve modification, and hence, a contrast presupposition, but the first instruction did as well. Thus, by virtue of using the modified expression *pink comb* in the first instruction, attention may have been drawn to the contrasting yellow comb as well, even prior to the second instruction, resulting in extremely fast eye movement latencies to this object relative to the yellow bowl.

The results of these experiments provide compelling evidence for a processing model in which linguistic expressions are undergoing continuous, moment-by-moment semantic interpretation, with immediate mapping onto a referential model. It is important to note, however, that these experiments demonstrate incremental model-based processing for the simplest type of adjectives, namely intersective adjectives. As discussed earlier, intersective adjectives lend themselves the most favorably to an algorithm that can compute semantic meanings incrementally on a word-by-word basis. That is, it is possible to evaluate the set of entities that are denoted by the adjective *yellow*, or *rectangular* independently of knowledge of the noun that is being modified. However, this is not the case for many types of adjectives, as pointed out by Clifton and Ferreira (1989). Thus, it is impossible to discern whether the expression *tall* applies to an object in the absence of knowledge pertaining to the identity of the object that is being so described. Clearly, it is possible for one of two different entities of the same height to be described as *tall*, and the other as *short*, as would be the case, for instance, with a child and basketball player of the same height. The existence of such adjectives suggests the possibility of a limit to the degree of incrementality of semantic processing. The experiments below address these issues in detail.

3. Experiment 2: contrast, typicality and scalar adjectives

Experiment 1 presented evidence for the incremental processing of adjectival modifiers, and the rapid availability of contextually-bound contrast sets in the interpretation of referential phrases involving adjectival modification. The experiments in this section investigate the use of contrastive knowledge in the process of interpreting adjectives that are vague in their denotation.

Scalar adjectives such as *tall*, *thin*, etc. have no central value, in contrast to adjectives such as *red* or *round*. As a consequence, if scalar adjectives are to be interpreted incrementally, the interpretation must be more complex, and involve the determination of a comparison class.

In this section, we explore the hypothesis that interpretation of adjectives is incremental even for the most problematic cases, where the adjective itself fails to have an invariant or stable meaning, but is highly dependent upon either the head noun, or some aspects of the context fixing a value on a scale. In such cases, evidence of incrementality is dependent upon the immediate use of head-based or contextual information. Experiment 2 assesses the relative contribution of stored representations associated with the head noun, and representations of the visual context in the incremental interpretation of scalar adjectives.

3.1. Norming data

Experiment 2 was conducted using real objects in a visual display. In order to determine the appropriateness of the target adjectives for a particular item with respect to its general category, a rating task was administered in which target objects were shown to subjects in isolation. Subjects were asked to indicate whether the object was best described by means of a noun modified by a target adjective (e.g. *a tall glass*), by means of a bare unmodified noun (e.g. *a glass*), or by means of an adjective that was on the opposite pole of the scale evoked by the target adjective (e.g. *a short glass*). The instructions for the rating task were as follows: ‘For every object that you are shown, indicate which of the three options is the most appropriate way to describe the object. If you feel that none of the descriptions provided are appropriate choose option (d) other, and write down what you feel would be a good description for that object.’ Twenty-five triplets of target objects were chosen, with the intent that each of the three choices provided on the ratings questionnaire would be represented by one of the objects. Objects described by means of the target adjective represented good tokens of the targeted adjective-noun pair, objects described by means of a bare noun were chosen as poor tokens for the adjective-noun pair, and objects described by means of the adjective opposite to the target were chosen as objects intended to contrast with the target objects in the visual context. Of the 25 triplets presented, 20 objects that received the most consistent ratings were chosen to be included in the experimental displays. Ratings for these 20 objects are presented in Table 5.

Table 5
Percentage of responses chosen for each option when shown good tokens, poor tokens and contrasting objects for the targeted adjective-noun expressions in the rating study

Shown	Target adjective (%)	No adjective (%)	Opposite adjective (%)	Other (%)
Good token	92.5	3.3	0	4.2
Poor token	19.4	69.3	10.5	0.8
Contrasting object	0	20.8	75.2	4.0

3.2. Subjects

Subjects were 24 members of the university community who were recruited by means of posted announcements, and were paid $7 for participating. All were monolingual native speakers of English and had either good uncorrected vision or wore soft contact lenses. None of the participating subjects had taken part in the rating study.

3.3. Materials and design

The ratings data indicate that it is possible to consistently categorize the target objects and their contrasts with respect to the appropriateness of modification by means of the target adjective. Based on these ratings, a set of experimental visual displays were constructed, such that for half of the displays the target object reflected a good fit with a description that involved a scalar adjective (good token), and for the remaining half, the target object reflected a poor fit (poor token). We will refer to this manipulation as the typicality manipulation, as it involves the typicality of an object with respect to the category corresponding to the complex noun phrase (e.g. the category of tall glasses). In addition, experimental displays were systematically varied such that half of the trials included a contrasting object which had been rated as being best described by means of an adjective that was opposite in meaning to the target adjective. The remaining half of the trials did not include such a contrast, but instead included an unrelated object for which the scale evoked by the adjective was completely irrelevant. There were a total of four objects in each display. In addition to the target and the contrast/distractor, each display also included a competitor object, that is, an object for which the scale evoked by the adjective was relevant (e.g. the competitor for *tall glass* was a pitcher). In absolute values, the competitor object was always further along on the scale evoked by the adjective than the target object, but was rated as being best described by means of an unmodified noun. For instance, the competitor pitcher was taller than either of the target glasses, but not tall with respect to pitchers in general. The fourth object in the display was an unrelated distractor item. An example display, involving contrast, is shown in Fig. 4.

A total of 20 experimental displays were used, consisting of objects on a horizontal workspace. Subjects were instructed to pick up particular objects and move them to a new location on the board (e.g. ‘Pick up the tall glass and put it below the

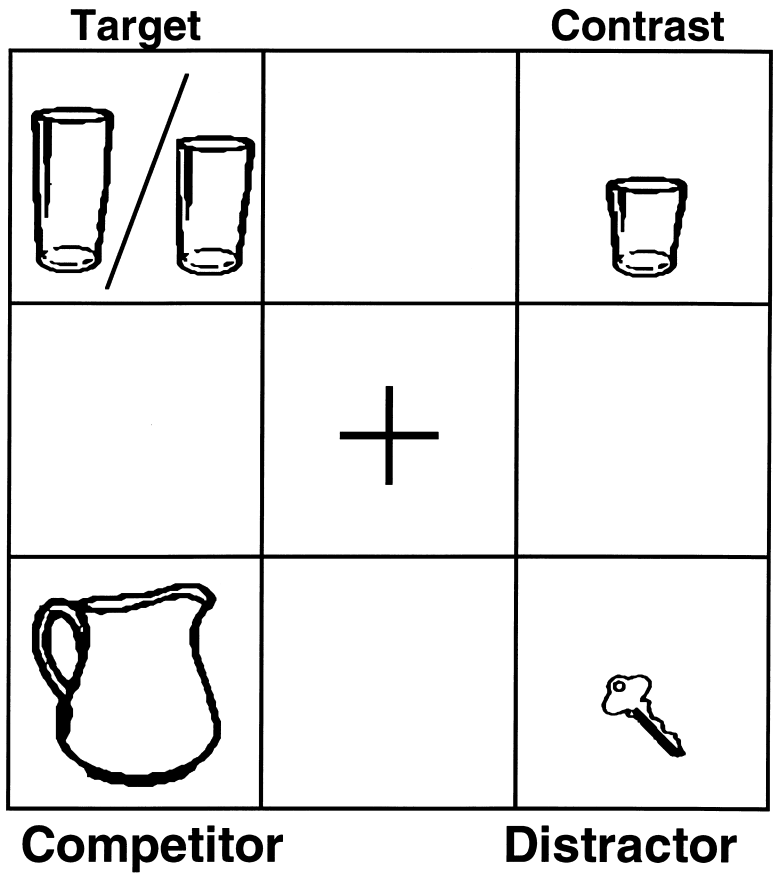

Fig. 4. An example of a display in the contrast condition for Experiment 2. In the no contrast condition, a second distractor (e.g. a file folder) occurred in place of the contrast object.

pitcher.') The critical instruction was always the first instruction accompanying the display and was followed by two filler instructions. In addition, 22 distractor trials accompanied by three instructions each were used. Ten of the distractor trials included a pair of contrasting objects (e.g. square and rectangular boxes), and were accompanied by instructions that involved adjectival modification, but with reference to an object outside of this contrasting pair (e.g. 'Pick up the square block'). This was done to prevent subjects from inducing a contrastive interpretation of adjectives based on the probabilistic use of adjectives internal to the experiment. In addition, many, but not all, of the filler trials used adjectival modification for displays that had no contrasting sets at all.

3.4. Procedure

The same head-mounted eyetracking apparatus was used as for Experiments 1A

and 1B, with similar procedures. However, in addition to the manipulations involving the display, a third manipulation was introduced to determine the impact of the degree of familiarity the subject had with the display, yielding a $2 \times 2 \times 2$ experimental design with contrast and typicality as a within-subjects factors, and display time as a between-subjects factor. Half of the subjects (i.e. those in the long display time condition) were permitted to freely view the displays as the objects were being placed on the display board, resulting in a high degree of familiarity with the simple display. Immediately prior to the first (i.e. critical) instruction for each display, subjects were asked to fixate on a central cross on the board to ensure that eye movements were launched from a position that was equidistant to all objects in the display. The remaining half of the subjects (in the short display time condition) were instructed to keep their eyes closed while the displays were being set up, so that they were unable to identify the objects in the display. The experimenter cued these latter subjects to open their eyes immediately prior to the instruction by uttering the word 'open'. Subjects were instructed to wait until they heard the cue word before opening their eyes, and to look towards the central fixation cross immediately upon opening their eyes. Eye movements were monitored by an assistant, who was able to determine whether subjects complied with the instruction to avoid looking at the displays between trials. If a subject did look at the objects between displays prior to hearing the cue at the beginning of the trial, that trial was skipped. Overall, subjects were extremely cooperative, and were able to avoid looking between trials almost all of the time.

Display time was manipulated because it was hypothesized that the relative effects of typicality and contrast would differ as a function of familiarity with the model. This is because an effect of typicality is related to the use of stored norms for object categories, whereas an effect of contrast is related to the presence of entities in the model that satisfy presuppositions associated with modifiers. Thus, we might expect contrast effects to be stronger, and typicality effects to be weaker in a task that allows subjects plenty of time to become familiar with the model.

3.5. Results and discussion

Results were scored by noting the onsets and offsets of critical words in the experimental instructions, and the timing of eye movement launches relative to these points. Data are reported below both in terms of latencies of eye movements to the target object, and false hits to objects other than the target.

Latencies for all first looks to the target were calculated from the onset of the disambiguating head noun, and are shown in Table 6. A $2 \times 2 \times 2$ ANOVA was conducted, with typicality and contrast as within-subjects factors, and display time as a between-subjects factor. There was a main effect of display time, with shorter latencies overall for the version where subjects were permitted to preview the displays prior to the instruction ($F_1(1,21) = 7.61$, $P < 0.05$; $F_2(1,19) = 33.64$, $P < 0.001$). A main effect of contrast was also observed, with displays that included a contrasting object resulting in shorter latencies than displays that did not include a contrasting object ($F_1(1,21) = 11.62$, $P < 0.01$; $F_2(1,19) = 4.19$, $P < 0.06$).

Table 6
Eye movement latencies in ms for Experiment 2, as measured from the onset of the head noun

Long display time	With contrast (ms)	Without contrast (ms)
Token		
Good token	408	485
Poor token	392	557
Short display time		
Good token	659	621
Poor token	703	812
Combined means		
Good token	538	556
Poor token	554	690

Finally, there was a main effect of typicality in the predicted direction as well, with good tokens yielding shorter latencies than poor tokens ($F_1(1,21) = 4.58$, $P < 0.05$; $F_2(1,19) = 8.46$, $P < 0.01$). In addition, the interaction of contrast and typicality was marginal by subjects, though not by items ($F_1(1,21) = 3.3$, $P = 0.08$; $F_2(1,19) = 2.67$, $P > 0.1$), such that the typicality effect was stronger for displays without contrast than for displays with contrast. The interaction of display time, typicality and contrast was not significant.

In addition to latencies, the percentage of trials that included a look to objects other than the target were calculated. Table 7 indicates the proportion of trials that included a look to the competitor object, the contrasting object (or the distractor object in the same location for displays that did not include a contrasting object) and to the fourth, unrelated distractor object. Separate analyses of variance were computed using the percentage of trials including looks to the competitor object, and to

Table 7
Percentage of trials in Experiment 2 that included a look to an object in the display other than the target object

Look	Competitor (%)	Contrast/Distractor (%)	Distractor (%)
Long display time			
Contrast-good token	5.42	25.42	6.25
Contrast-poor token	10.42	37.5	3.33
No contrast-good token	37.36	8.75	15.28
No contrast-poor token	37.08	5.42	7.5
Short display time			
Contrast-good token	35.0	53.75	24.17
Contrast-poor token	22.5	63.75	23.33
No contrast-good token	33.33	12.92	24.58
No contrast-poor token	67.08	25.0	26.67
Combined means			
Contrast-good token	20.21	39.58	15.21
Contrast-poor token	16.46	50.63	13.33
No contrast-good token	35.35	10.83	19.93
No contrast-poor token	52.08	15.21	17.08

the contrasting object as dependent measures. 2 × 2 × 2 ANOVAs for looks to the competitor object with display time as a between-subjects factor, and contrast and typicality as within-subjects factors, revealed a mean effect of display time, with a greater number of looks to the competitor object for the version in which subjects were not permitted to preview the displays ($F_1(1,22) = 6.34$, $P < 0.05$; $F_2(1,19) = 13.62$, $P < 0.01$). There was also a main effect of contrast, with displays involving contrast resulting in fewer looks to the competitor object ($F_1(1,22) = 32.66$, $P < 0.001$; $F_2(1,19) = 26.69$, $P < 0.001$). The effect of typicality was in the predicted direction, and marginally reliable by subjects only ($F_1(1,22) = 3.18$, $P < 0.09$; $F_2(1,19) = 2.25$, $P > 0.1$). An interaction of contrast and typicality was found ($F_1(1,22) = 8.73$, $P < 0.01$; $F_2(1,19) = 8.93$, $P < 0.01$), such that displays without contrast showed more marked effects of typicality than displays with contrast. In addition, the three-way interaction of display time, contrast and typicality was also significant ($F_1(1,22) = 13.81$, $P = 0.001$; $F_2(1,19) = 19.38$, $P < 0.001$).

Statistical analysis using the proportion of trials with looks to the contrasting object as a dependent measure (or the control object in the same location for displays without contrast) was also conducted. As with looks to the competitor, a main effect of display time was found ($F_1(1,22) = 8.44$, $P < 0.01$; $F_2(1,19) = 36.31$, $P < 0.001$), reflecting a general pattern where subjects looked at non-target objects more frequently when they were less familiar with the displays. A main effect of contrast was also found ($F_1(1,22) = 101.21$, $P < 0.001$; $F_2(1,19) = 65.96$, $P < 0.001$), with more trials including a look to the relevant object when it contrasted to the target object in the instruction than when it was an unrelated control in the same location. The effect of typicality was marginally significant by subjects and items ($F_1(1,22) = 4.28$, $P < 0.06$; $F_2(1,19) = 3.58$, $P < 0.08$). There was also an interaction of contrast and display time ($F_1(1,22) = 5.84$, $P < 0.05$; $F_2(1,19) = 5.99$, $P < 0.05$), with contrast exerting a stronger effect when subjects were more familiar with the visual displays.

Fig. 5 provides a more detailed look at the eye movement data over time by showing the proportions of trials that include fixations to the different objects in the display for each coding frame (frames at 30/s) throughout the trial. There are a number of interesting observations to draw from these graphs. First, the difference in latencies due to contrast was preserved in this presentation format. For displays with contrast, there was a steep rise in fixations upon the target object beginning immediately at, or before, the onset of the disambiguating head noun, suggesting that eye movements are beginning to be programmed sometime during the adjective. This rise occurred visibly later for displays without contrast. Second, the incidence of looks to the competitor object was much higher for displays without contrast, and the point at which looks to the target begin to diverge from looks to the competitor was considerably later.

This information has been presented in terms of proportions of trials including a look to the competitor at any point in the trial. However, of special interest here is the evidence that the timing of looks to the competitor was strikingly different from the timing of looks to the contrasting objects. Looks to the competitor peaked very early, at or before the offset of the head noun, with a steep rise occurring around the

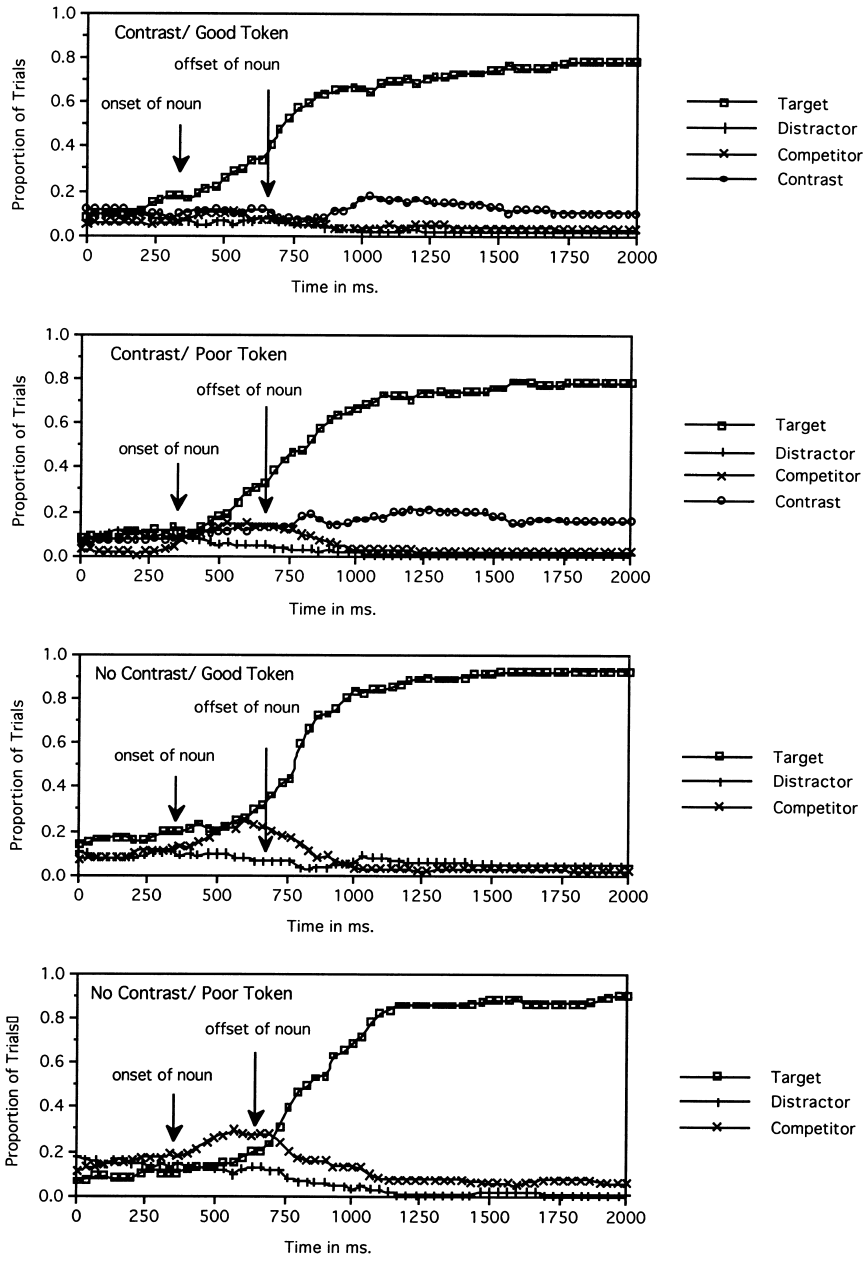

Fig. 5. Time course of eye movement data for Experiment 2 showing the proportion of trials that contain a look to each object in the display over time. Point 0 ms corresponds to the onset of the adjective in the instruction.

onset of the noun, and subsided relatively quickly. This is consistent with the interpretation that subjects were processing the meanings of adjectives incrementally. Eye movements to the competitor object were being programmed primarily during the ambiguous region of the modified phrase, that is during the adjective, and looks dropped off quickly as the disambiguating head noun unfolded.

The earliness of looks to the competitor suggests two things. First it suggests that subjects were not waiting until the head noun to begin processing adjectival meanings, but were beginning to assign an interpretation immediately. Second, the information available in the displays regarding the presence of a contrasting object was used extremely early as well, evident in the low occurrence of looks to the competitor for displays with a contrast. Thus, information about the various objects in the display was being used in conjunction with knowledge of the contrastive function of the adjective as the subject heard the adjective itself. The timing of looks to the contrast suggest that these objects, on the other hand, were not being considered as possible targets early on in the modified expression. Looks to the contrasting object occurred somewhat later than looks to the competitor, and did not drop off as sharply. This may reflect either a process of confirming the value for the adjective with respect to the relevant object for comparison, or simply an attempt to visually discriminate between two objects of the same category (e.g. the tall glass vs. the short glass) after hearing the head noun.

What emerges from the eye movement data is clear evidence that subjects are sensitive to the contrastive use of the adjectives, and that this information is used incrementally, in such a way as to affect early interpretation of the vague scalar adjective. This is evident in faster eye movement latencies to the target object, the lack of a competitor effect in displays containing a contrasting object and the prevalence of looks to the contrasting object when the display had one. In comparison, the effects of typicality are surprisingly subtle, particularly given the stability of judgments in the rating task. They emerge most clearly in displays with no contrasting object present, and are strikingly weak for displays with contrast.

4. Experiment 3

Experiment 2 indicates a strong tendency for nouns modified by adjectives to be understood as referring to one of a contrasting pair of objects. Furthermore, this information appears to be used extremely rapidly to reduce the degree of referential indeterminacy involved in incremental interpretation. Although the source of the effect of contrast with modifiers is incompletely understood, as will be addressed in Section 5, it appears to be closely related to the phenomenon of packaging utterances in such a way as to signal old or presupposed information, and new information. It is clear that not all uses of a modifier are contrastive. Fox and Thompson (1990) identify various functions associated with relative clause modifiers, and argue that the status of the head noun as either old or new in the discourse has an impact on the observed function. Further support for the relationship between information packaging and modification comes from Spivey-Knowlton and Sedivy (1995), who show

a correlation in corpus data between the definiteness of a noun phrase (where definite noun phrases typically signal old information) and the use of a prepositional phrase as a modifier of the noun, as well as online experiments showing an increased tendency to understand an ambiguous prepositional phrase as a nominal modifier when the noun is definite than when it is indefinite (see also Schelstraete, 1996). If contrastive uses of modifiers are indeed related to the status of the head noun as old or new, then several aspects of Experiment 2 are likely to have produced a maximal effect of contrast. First, the modified noun phrases in the target instructions were definite. Second, the entire task was set up in such a way as to presuppose the existence of the target entity in the discourse model. That is, it is only felicitous to instruct someone to pick up a particular object and move it in a context where such an object actually exists, and is perceptually or cognitively accessible to both the speaker and the hearer.

Experiment 3 was designed to test the degree to which the contrastive interpretation of adjectives is dependent upon the presupposition of existence and accessibility of the entity being described. Experiment 3 differed from Experiment 2 in two aspects. First, the target noun phrases were indefinite, rather than definite. Second, the task involved a verification task requiring subjects to answer questions such as 'Is there a tall glass?' rather than a task in which subjects were instructed to manipulate objects. This created an experimental situation in which the existence of an object fitting the target description was left completely open, rather than presupposed. The goal of this experiment was to probe for the relative contribution of contrast and typicality to incremental semantic processing in an experimental situation which removed the presuppositions associated with the target utterances in Experiment 2.

4.1. Subjects

Twenty-two members of the university community participated, and were paid $7 for the session. All were monolingual native speakers of English, with either good uncorrected vision, or soft contact lenses, and none had participated in earlier versions of the experiment, or in the rating study.

4.2. Materials and design

The displays used in this experiment were identical to those in Experiment 2. The critical trials contained modified referential expressions that were also identical to those in Experiment 2, with the following expectations: The noun phrases were indefinite (e.g. *a tall glass*) as opposed to definite (e.g. *the tall glass*), and, rather than being embedded in an instruction to pick up an object and move it to a new location, were uttered as part of a verification question, such as '*Is there a tall glass?*' Thus, subjects were free to decide whether the modified expression was an appropriate description of any of the objects present in the display. In fact, such as decision was a central aspect of the task required of the subject.

As in Experiment 2, the factors of contrast, typicality and display time were

manipulated, with contrast and typicality as within-subjects manipulations, and display time as a between-subjects factor. Thus, half of the experimental items contained a contrasting object, while the other half did not, and half of the target objects were good tokens of the modified expression, while the remaining half were poor tokens. The manipulation of display time was executed in the same way as in Experiment 2, with half of the subjects being permitted to freely preview the displays as they were being set up, and the other half being required to keep their eyes closed until hearing a cue word immediately prior to the experimental question.

As in the previous experiment, the 20 critical questions always occurred as the first question of any trial. The critical questions were followed by a second question pertaining to the displays. In addition, there were 22 distractor trials accompanied by two questions each. Again, 10 of the distractor trials included a pair of contrasting objects (e.g. square and rectangular boxes), and were accompanied by questions that involved adjectival modification, but with reference to an object outside of this contrasting pair (e.g. '**Is there a square block?**'). This was done to ensure that subjects did not induce a contrastive interpretation of adjectives based on the probabilistic use of adjectives internal to the experiment. In addition, many but not all, of the filler trials used adjectival modification for displays that had no contrasting sets at all. The distractor questions were constructed such that the anticipated answer to the verification questions was 'yes' half the time, and 'no' the remaining half.

4.3. Procedure

As the subjects performed the task, their eye movements to the display were monitored by means of the same apparatus as that used in Experiments 1 and 2. In addition, subjects' responses to the verification questions were recorded onto the video-audio record that contained the eye movement record. Subjects were instructed to respond as quickly as possible, and to limit responses to a simple 'yes' or 'no'. In addition, the experimenter encouraged the subjects to begin each trial by fixating the central cross to maximize the likelihood that subjects were fixating a point that was equidistant to all of the objects in the display.

4.4. Results and discussion

Three general measures were of interest in this task: (1) the nature of the subjects' responses, that is, whether they accepted or rejected the modified expression as a description of any of the objects present in the display; (2) the latencies of the responses, particularly in cases where the response was 'yes'; and (3) the pattern of eye movements over the course of the question.

Table 8 shows the percentage of 'yes' responses to the critical questions for each of the four conditions. The results indicated that in the absence of a contrasting object, subjects were very sensitive to the typicality manipulations, as would be expected on the basis of the rating study. However, in the presence of a contrasting object, the typicality effect was reduced. Analysis of variance revealed a significant

Table 8
Percentage of 'yes' responses to the verification question in Experiment 3

Token	With contrast (%)	Without contrast (%)
Long display time		
Good token	0.0	2.08
Poor token	5.0	33.33
Short display time		
Good token	8.75	1.67
Poor token	11.67	36.67
Combined means		
Good token	4.38	1.88
Poor token	8.33	35.00

main effect of contrast ($F_1(1,21) = 14.35$, $P = 0.001$; $F_2(1,19) = 12.35$, $P < 0.01$), and a main effect of typicality ($F_1(1,21) = 25.72$, $P < 0.001$; $F_2(1,19) = 15.32$, $P < 0.001$). Significant differences due to display time were found for items only ($F_2(1,19) = 6.56$, $P < 0.01$). The interaction of contrast and typicality was significant ($F_1(1,21) = 17.13$, $P < 0.001$; $F_2(1,19) = 19.45$, $P < 0.001$), with stronger effects of typicality evident for the no contrast conditions than the conditions with contrast.

In addition, a $2 \times 2 \times 2$ ANOVA was carried out for latencies of 'yes' responses, as measured from the onset of the noun (mean latencies are displayed in Table 9). The results of the analysis showed a significant main effect of both contrast ($F_1(1,21) = 20.00$, $P < 0.001$; $F_2(1,13) = 7.9$, $P < 0.05$) and typicality ($F_1(1,21) = 23.87$, $P < 0.001$; $F_2(1,13) = 22.95$, $P < 0.001$). The effect of display time was significant by items only ($F_2(1,13) = 7.2$, $P < 0.05$). The interaction between contrast and typicality was significant ($F_1(1,21) = 8.77$, $P < 0.01$; $F_2(1,13) = 22.95$, $P < 0.001$), with typicality exerting a stronger effect in the absence of the contrasting object. The interaction between display time and contrast was significant by subjects only ($F_1(1,21) = 4.52$, $P < 0.05$).

The eye movement data provided further confirmation of the role of contrast in

Table 9
Latencies in ms of 'yes' responses to the verification question in Experiment 3, as measured from the onset of the head noun

Token	With contrast (ms)	Without contrast (ms)
Long display time		
Good token	672	747
Poor token	802	1006
Short display time		
Good token	823	801
Poor token	879	1003
Combined means		
Good token	751	775
Poor token	842	1004

Table 10
Mean latencies in ms of first eye movements to the target object in Experiment 3, as measured from the onset of the head noun

Token	With contrast (ms)	Without contrast (ms)
Good token	751	775
Poor token	842	1004

this task. Table 10 shows the latencies of the first looks to the target in the critical region of the question (i.e. after the onset of the adjective). As in Experiment 2, there was an effect of contrast on eye movement latencies ($F_1(1,18) = 6.76$, $P < 0.05$; $F_2(1,17) = 4.36$, $P = 0.05$), with longer latencies for displays without contrast than displays with contrast. There was also an effect of typicality ($F_1(1,18) = 4.49$, $P < 0.05$; $F_2(1,17) = 17.38$, $P = 0.001$), with longer displays for poor tokens than good tokens. There was no main effect of display time and none of the interactions were significant.

Table 11 shows the proportion of trials including a look to an object other than the target. $2 \times 2 \times 2$ ANOVAs were conducted with proportion of trials containing a look to the competitor object as a dependent measure. Results of these analyses yielded a significant main effect of contrast ($F_1(1,19) = 12.83$, $P < 0.01$; $F_2(1,15) = 11.73$, $P < 0.01$). Typicality yielded only a marginally significant effect by subjects ($F_1(1,19) = 3.04$, $P < 0.1$). There was no main effect of display time, and none of the interactions were significant.

Additional ANOVAs were carried out on the proportion of trials containing a look to the contrasting object (or distractor item in the same location for displays without contrast). The only significant effect was a main effect of contrast ($F_1(1,19) = 70.29$, $P < 0.001$; $F_2(1,15) = 35.96$, $P < 0.001$), with a greater proportion of looks occurring to the contrasting object in displays with contrast than the unrelated distractor item in the same location for displays without contrast.

Fig. 6 shows the time course of eye movements over the duration of the trial. The patterns of eye movements across the four conditions were strikingly similar to those observed in Experiment 2. First, displays containing contrasting objects showed an earlier rise in looks to the target than do displays without contrast. Second, displays with contrast showed no evidence that subjects were more likely to look at the competitor object than the control distractor object at any point in the trial. A competitor effect was seen for displays without contrast, particularly for displays where the target object was a poor token. Third, as in Experiment 2, there was a clear difference in the timing of eye movements to the competitor object as compared to the contrasting objects. Again, looks to the competitor object occurred early, dropping off sharply after disambiguating information from the head noun accrued, while looks to the contrasting objects occurred considerably later in the trial, and persisted over a longer stretch of time. In fact, the general shape of the curves, and the points at which looks to the various objects in the display begin to diverge from one another was virtually identical for the two experiments.

The results of Experiment 3 lend further support to the hypothesis that the inter-

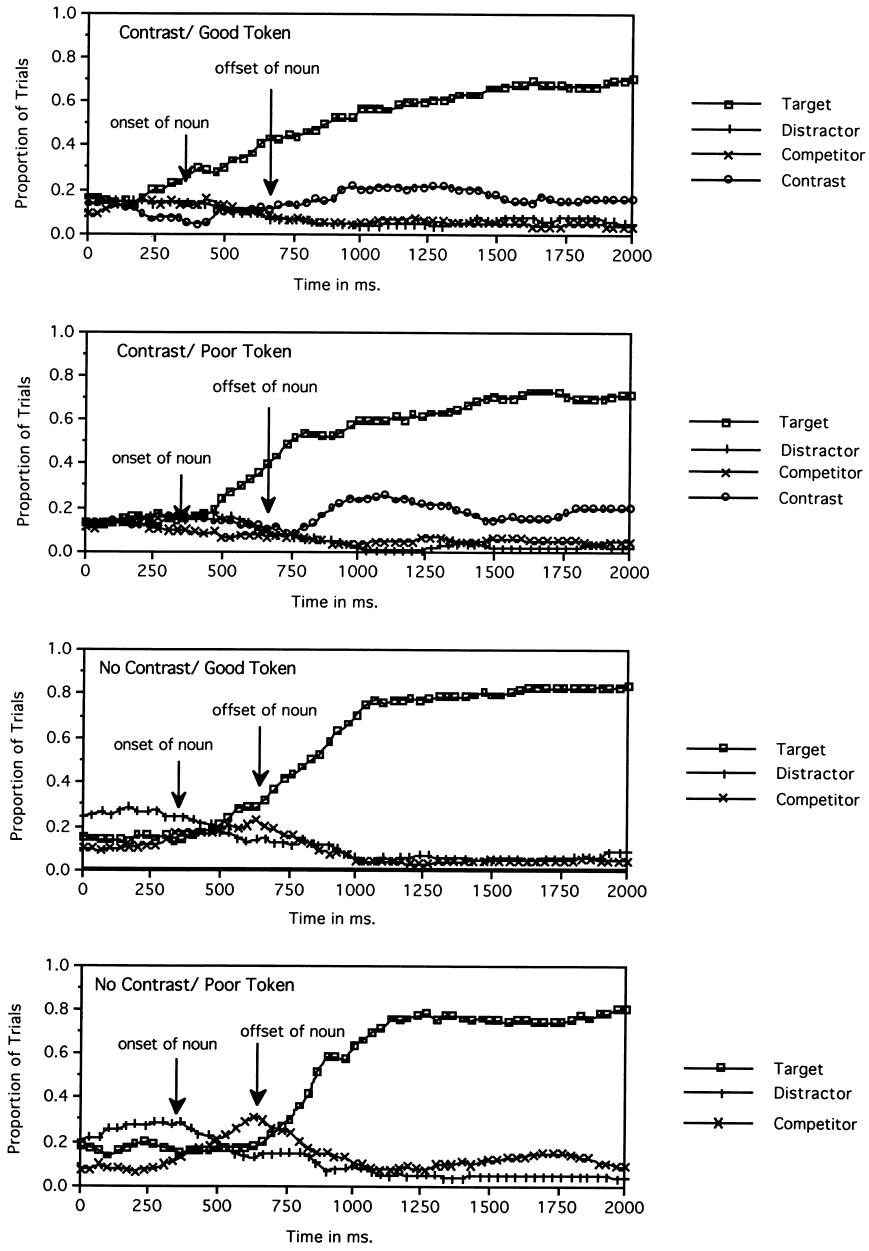

Fig. 6. Time course of eye movement data for Experiment 3 showing the proportion of trials that contain a look to each object in the display over time. Point 0 ms corresponds to the onset of the adjective in the instruction.

Table 11
Percentage of trials in Experiment 3 that included a look to an object in the display other than the target object

Look	Competitor (%)	Contrast/Distractor (%)	Distractor (%)
Contrast-good token	17.86	42.14	15.95
Contrast-poor token	24.44	48.65	13.10
No contrast-good token	31.90	8.10	12.38
No contrast-poor token	41.90	17.38	21.67

pretation of adjectival modifiers proceeds in an incremental fashion. As in Experiment 2, there is further evidence that the processing system was able to make use of contrastive information associated with scalar adjectives. Both contrast with respect to a stored representation of the class of objects denoted by the head noun, and contrast with respect to a contextually-available set of objects are relevant for incremental interpretation. The finding that effects of contextually-defined contrast are not dependent upon the presuppositions inherent in instructions such as those in Experiment 2 is particularly interesting and somewhat unexpected. Robust effects of contrast are found even with indefinite nouns, and in a situation where the experimental task did not carry any presuppositions of the presence of an object aptly described by the modified expression.

5. General discussion

The experiments in this paper converge upon the finding that interpretation of adjectives is incremental even when the adjective fails to have a stable core meaning. This incrementality is achieved by rapidly establishing contrast either between objects in the immediate visual context, or between an object and its corresponding typical representation in memory. It is worthwhile to consider how the rapid identification of contrast may be linked to the presence and properties of a modifying adjective.

One possibility is that the relationship between modification and contrast is based on Gricean principles of conversation, and reflects an expectation on the part of the hearer that the speaker is communicating in an optimally efficient manner, with neither more nor less information than necessary being linguistically expressed. Clifton and Ferreira (1989) assume that such an inferential mechanism underlies the contrastive nature of modifiers, and argue that Gricean inferences of this sort could not possibly be computed sufficiently quickly to have an impact upon online sentence processing:

> Faced with a post-nominal modifier, a listener/reader might reason, 'following Grice (1975) principle of quantity, the speaker/writer would not be giving me more information than necessary, therefore the modifier is probably needed to pick out the relevant item, so there are probably other such items or s/he may think I will have some other source of difficulty in identifying the intended referent.' Perhaps Altmann and Steedman would claim that conversational

> implicatures do play a role in initial parsing decisions. We consider this unlikely. Conversational implicatures are...not tied to the form of what is said, but rather, to its semantic content. To make a conversational implicature, a listener must have already parsed the sentence, assigned it its literal interpretation, realized that additional inferences must be added to make it conform to the Gricean maxim, and determined what these inferences are. Such activity could not reasonably affect the initial steps of parsing (Clifton and Ferreira, 1989, p. 84).

It remains unclear whether or not Gricean inferences are actually delayed as Clifton and Ferreira argue. Further investigation is needed before such a claim could be made.

However, regardless of the issue of the speed of processing conversational implicatures, there exist proposals of the contrastive function of modifiers that do not hinge on Gricean mechanisms. For instance, Steedman and Altmann (1989) argue that the contrastive presupposition associated with modifiers is, contrary to Clifton and Ferreira's claims, closely linked to linguistic form, and focuses on the particular contribution of modifiers and their relation to the discourse model. Specifically, the presupposition of contrast is triggered by the combination of definiteness of the noun phrase, and the presence of a modifier, and presumably hinges on the pragmatic presuppositions generally associated with definite noun phrases (Heim, 1982; Kadmon, 1990). A detailed discussion arguing for a form-based account of the contrastive presupposition can be found in Steedman and Altmann (1989).

In his analysis of scalar adjectives, Bierwisch (1987) proposes a form-based account of the contrastive presupposition that is somewhat different from Steedman and Altmann's account. Where Referential Theory has identified properties of definite (or presupposed) noun phrases as the locus of the contrastive effect, Bierwisch's analysis places the locus of the effect in the lexical entry for the adjective itself. For Bierwisch, contrast originates in the requirement that the adjective itself be interpreted with respect to a comparison class. Although the issue is not addressed by Bierwisch, the requirement for a comparison class that is inherent in scalar adjectives might plausibly interact with more general presuppositions associated with definite modified expressions to set the comparison class to a subset of the entities denoted by the head noun, namely those entities falling within that set that have been instantiated in the discourse. The phrase *tall snowman* (from previous example) then, would be understood with respect to a presupposed set of snowmen in a particular discourse context:

(9) My 2-year-old son built the tall snowman.

Under a theoretical view such as Bierwisch's, the determination of the semantic value of adjectives cannot be accomplished unless variables specified by the lexical entry for a particular adjective are fixed on the basis of contextually-supplied information. This approach suggests a dramatically different mechanism for accessing contrastive information pertaining to modifiers than the one suggested by Clifton and Ferreira (1989). The lexical representation of a scalar adjective contains a

variable whose value must be determined by a salient contrast set, thereby providing an inherent contrastive interpretation that does not rely on conversational principles.

There are close conceptual and formal similarities between Bierwisch's approach to the contrastiveness inherent in scalar adjectives and recent approaches to the semantics of focus (for explicit formal accounts, see Rooth, 1985, 1992; Krifka, 1991). Sentences involving focus marking have also been argued to exhibit presuppositions that are closely related to some aspect of linguistic form, rather than inferred indirectly from the semantic content of the sentence. Consider, for instance, sentences such as (10a) below, in which focus is indicated by the preposed phrase *It was Bill that* or (10b), where focus marking is indicated by intonational prominence on *Bill*. Both of these sentences involve the presupposition that the object of Mary's love is under discussion (Jackendoff, 1972) and suggest a contrast between Mary's feelings towards Bill, and Mary's feelings towards some contextually salient set of individuals. Sentence (13), however, which is semantically identical to both the sentences in (10) (i.e. it is true under the same conditions as (10a) and (10b)), does not carry such a presupposition, presumably because the syntactic or intonational correlates of focus are absent.

(10a) It was Bill that Mary loved.

(10b) Mary loved BILL.

(11) Mary loved Bill.

Rooth (1992) describes a theory of focus which involves formal mechanisms similar to Bierwisch's analysis of adjectives, namely, the use of free variables that are simultaneously constrained by the linguistic form of the sentence, and fixed by salient contextual factors, allowing for an intimate interleaving of semantic and pragmatic information.

Such formal analyses argue for a tighter linking between linguistic form and pragmatic presupposition than has frequently been conceived. They also suggest an intriguing framework in which to view experimental evidence for the very rapid use of pragmatic information that is associated with certain linguistic forms.

The results of the experiments reported here present clear evidence for the interaction of contextual and linguistic information at the earliest possible moments. Thus, not only is there evidence that semantic interpretation begins immediately upon hearing the adjective, but it is also clear that information about the contextual import of the adjective is accessed and integrated with information present in the visual context at this point in processing as well. What is not conclusive from the present study, is whether contextually-based effects of contrast with nominal modifiers are most effectively accounted for via inferred conversational implicatures, by means of presuppositional properties of modifiers, as argued by proponents of the Referential Theory, or by the specific contrastive requirements of scalar adjectives. Certainly, contrastive effects have been found with modifiers other

than adjectives, such as relative clauses, and prepositional phrases, primarily in studies focusing on syntactic ambiguity resolution, supporting the claims of Referential Theory. In addition, there is evidence that the definiteness of the noun phrase is implicated, with stronger presuppositional effects occurring with definite noun phrases than indefinite noun phrases (Spivey-Knowlton and Sedivy, 1995; Schelstraete, 1996).

However, Bierwisch's insight that vague scalar adjectives rely heavily on the identification of a comparison class does provide some explanation for the robustness of the effects of contextual contrast found in this study. Particularly striking is the finding that strong contrast effects are not in fact limited to definite noun phrases, but occur with indefinite noun phrases as well, even in an experimental task that carries no presuppositions that the modified noun is an appropriate description of any of the objects in the visual model. One explanation for pervasive effects of contextual contrast may lie in the vagueness of the adjectives used in Experiments 2 and 3. That is, adjectives such as *tall* which have no central value or stable norm independent of the noun they modify, rely more heavily on a comparison class than adjectives with more stable meanings, such as color adjectives. One might expect, then, that effects of contextually defined contrast would be more limited with color adjectives, particularly in their interaction with typicality norms associated with the central values for the adjectives themselves. Investigations focusing on different classes of adjectives are likely to be informative in determining the interaction between lexical information pertaining to specific adjectives and more general pragmatic properties.

More generally, the results presented in this study serve to give some shape to the broad problem of characterizing the nature of meanings that are interpreted online as part of human language processing. As discussed in Section 2, the lack of semantic constancy of adjectives has led to analyses in which adjectives do not directly refer to sets of entities, but are dependent upon the head noun for reference. Such analyses lead one to believe that semantic processing must proceed in a less-than-fully-incremental fashion.

The present study describes both a paradox and a potential solution for the problem of incremental semantic interpretation. The paradox lies in the evidence that referential interpretation is, in fact, not delayed, even for adjectives that appear to be heavily dependent upon the head noun for establishing reference. The solution lies in the evidence that referential processing is achieved incrementally through the interplay of the semantic content of the adjective and its relationship to the context of the utterance. Presumably, incremental semantic processing is not limited to utterances accompanied by a context sufficiently rich to resolve potential indeterminacy by allowing for the identification of a single referent (or set of referents). However, because of the contextual parameters involved in semantic interpretation, it may only be possible to fully identify the referent of a noun phrase prior to the head when there is sufficient contextual information. For instance, we suggest that the interpretation of adjectives generally involves some notion of a contrast set which serves as the comparison for the referent of the modified noun phrase. However, the degree to which this contrast set is specified is heavily

dependent upon the context. In the absence of such contextual specification, the precise identity of the referent of the phrase may not be possible until sufficient information has accrued to result in the specification of the contrast set, either via the head noun or perhaps by means of inferences triggered by linguistic material occurring after the head noun. In such cases, we suggest that the interpretation of adjectives is incremental, but vague or incompletely specified, insofar as the set of entities in the model that are denoted by the adjective has not been precisely identified. Thus, we would predict that although reference may not be fully resolved in favor of a particular entity, it should still be possible to observe effects of the elimination of some entities as possible candidates for reference. In fact, this is what we observe in the no-contrast conditions in Experiments 2 and 3, where there is relatively weak contextual information in support of the target entity. In these conditions, eye movements are more evenly distributed over both the target object and the competitor object; however, eye movements to objects in the display for which no relevant comparison class could plausibly be constructed are very sparse. The results of these studies suggest a system in which no a priori temporal constraints on semantic interpretation apply; rather, the processing system is characterized by a continuous uptake of information, including information pertaining to the context of the utterance, in an attempt to establish reference as quickly as possible.

In these studies, two ingredients were necessary to produce empirical evidence of the incremental interpretation of adjectives: First, it was important to explore and systematically manipulate the contextual parameters that were relevant for interpretation. Second, such an investigation needed to be conducted within an experimental paradigm that was amenable to observing interpretive processes in context, with the study of eye movements to visual contexts proving to be ideally suited for this purpose. The results obtained here argue not only for the general use of contextual parameters in incremental semantic interpretation, but also for the need to address questions of incrementality within an arena which pays attention to the function of linguistic forms, and how such forms make contact with contextual representations. Given recent advances in linguistic analyses that incorporate mechanisms for the interaction of linguistic form, meaning and context, as well as the development of experimental paradigms sensitive to the role of context in interpretation, there is reason to believe that significant progress will be made in investigating these questions.

Acknowledgements

The research presented in this paper was supported by NIH grant #HD 27206 and NSF grant #SBR 95-11340. We thank Eman Moustafa for assistance in data collection and analysis, and Paul Allopenna, Hana Filip, Joy Hanna, and Jim Magnuson for valuable comments on this work at various stages. We are also indebted to two anonymous reviewers for their insightful and thorough comments.

References

Altmann, G., Garnham, A., Dennis, Y., 1992. Avoiding the garden-path: eye movements in context. Journal of Memory and Language 31, 685–712.

Altmann, G., Garnham, A., Henstra, J., 1994. Effects of syntax in human sentence parsing: evidence against a structure-based proposal mechanism. Journal of Experimental Psychology: Learning. Memory and Cognition 20, 209–216.

Altmann, G., Steedman, M., 1988. Interaction with context during human sentence processing. Cognition 30, 191–238.

Bierwisch, M., 1987. The semantics of gradation. In: Bierwisch, M., Lang, E. (Eds.), Dimensional Adjectives. Springer, Berlin.

Britt, A., 1994. The interaction of referential ambiguity and arguments structure in the parsing of prepositional phrases. Journal of Memory and Language 33, 251–283.

Carroll, J., Bever, T., 1978. The perception of relations: a case study in the relation of knowledge and perception. In: Carterette, E., Friedmasn, M. (Eds.), Handbook of perception, Vol. 7. Academic Press, New York.

Clifton, C., Ferreira, F., 1989. Ambiguity in context. Language and Cognitive Processes 4, 77–103.

Crain, S., Steedman, M., 1985. On not being led up the garden path: the use of context by the psychological parser. In: Dowty, D., Karttunnen, L., Zwicky, A. (Eds.), Natural Language Parsing. Cambridge University Press, Cambridge, MA.

Eberhard, K., Spivey-Knowlton, S., Sedivy, J., Tanenhaus, M., 1995. Eye movements as a window into real-time spoken language processing in natural contexts. Journal of Psycholinguistic Research 24, 409–436.

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

Fox, B., Thompson, S., 1990. A discourse explanation of the grammar of relative clauses in English conversation. Language 66, 297–316.

Frazier, L., 1987. Theories of syntactic processing. In: Garfield, J. (Ed.), Modularity in Knowledge Representation and Natural Language Processing. MIT Press, Cambridge, MA.

Gelman, S., Markman, E., 1985. Implicit contrast in adjectives vs. nouns: implications for word-learning in preschoolers. Journal of Child Language 12, 125–143.

Grice, H., 1975. Logic and conversation. In: Cole, P. Morgan, J. (Eds.), Syntax and Semantics, Vol. 3, Speech Acts. Academic Press, New York.

Halff, H., Ortony, A., Anderson, R.C., 1976. A context-sensitive representation of word meanings. Memory and Cognition 4, 378–383.

Heim, I., 1982. The semantics of definite and indefinite NPs. Ph.D. Dissertation, University of Massachusetts at Amherst. GLSA, Amherst, MA.

Jackendoff, R., 1972. Semantic Interpretation in Generative Grammar. MIT Press, Cambridge, MA.

Kadmon, N., 1990. Uniqueness. Linguistics and Philosophy 13, 273–323.

Kamp, H., Partee, B., 1995. Prototype theory and compositionality. Cognition 57, 129–191.

Krifka, M., 1991. A compositional semantic for multiple focus constructions. Proceedings of Semantics and Linguistic Theory. Cornell University Press, Ithaca, NY.

Maloney, L., Gelman, S., 1987. Measuring the influence of context: the interpretation of dimensional adjectives. Language and Cognitive Processes 2, 205–215.

Marslen-Wilson, W., 1975. Sentence perception as an interactive parallel process. Science 189, 226–228.

Matin, E., Shao, K., Boff, K., 1993. Saccadic overhead: information processing time with and without saccades. Perceptual Psychophysics 53, 372–380.

Mitchell, D., Corley, M., Garnham, A., 1992. Effects of context in human sentence parsing: evidence against a discourse-based proposal mechanism. Journal of Experimental Psychology: Learning. Memory and Cognition 18, 69–88.

Murray, W., Liversedge, S., 1994. Referential context effects on syntactic processing. In: Clifton, C., Frazier, L., Rayner, K. (Eds.), Perspectives on Sentence Processing. Erlbaum, Hillsdale, NJ.

Ni, W., Crain, S., Shankweiler, D., 1996. Sidestepping garden paths: assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. Language and Cognitive Processes 11, 283–334.

Pierrehumbert, J., 1990. The phonology and phonetics of English intonation, Ph.D Dissertation. MIT Press, Cambridge, MA.

Pollard, C. Sag, I., 1994. Head-Driven Phrase Structure Grammar. Chicago University Press, Chicago, IL.

Rayner, K., Carlson, M., Frazier, L., 1983. The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. Journal of Verbal Learning and Verbal Behavior 22, 358–374.

Rooth, M., 1992. A theory of focus interpretation. Natural Language Semantics 1, 75–116.

Rooth, M., 1985 Association with focus, Ph.D Dissertation. University of Massachusetts at Amherst. GLSA, Amherst, MA.

Schelstraete, M.-A., 1996. Definiteness and prepositional phrase attachment in sentence processing. Current Psychology of Cognition 15, 463–486.

Sedivy, J., Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., Carlson, G., 1995. Using intonationally-marked presuppositional information in on-line language processing: evidence from eye movements to a visual model. In: Proceedings of the 17th Annual Conference of the Cognitive Science Society. Erlbaum, Hillsdale, NJ.

Spivey-Knowlton, M., Sedivy, J., 1995. Resolving attachment ambiguities with multiple constraints. Cognition 55, 227–267.

Spivey-Knowlton, M., Tanenhaus, M., 1994. Referential context and syntactic ambiguity resolution. In: Clifton, C., Frazier, L., Rayner, K. (Eds.), Perspectives on Sentence Processing. Erlbaum, Hillsdale, NJ.

Steedman, M., 1989. Grammar, interpretation, and processing from the lexicon. In: Marslen-Wilson, W. (Ed.), Lexical Representation and Process. MIT Press, Cambridge, MA.

Steedman, M., Altmann, G., 1989. Ambiguity in context: a reply. Language and Cognitive Processes 4, 211–234.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., Sedivy, J., 1995. Integration of visual and linguistic information during spoken language comprehension. Science 268, 1632–1634.