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*

1. Introduction

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

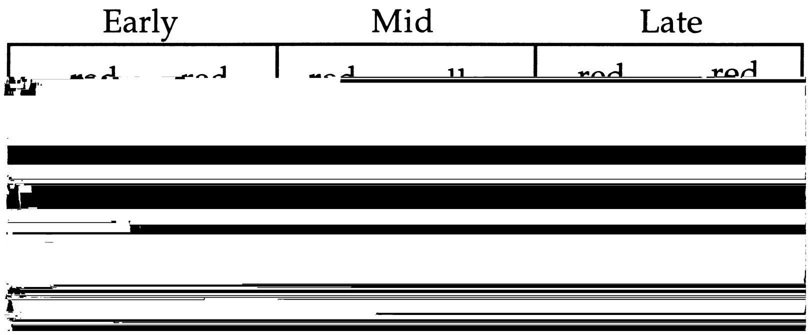


Fig. 1. Example displays from conditions manipulating the point at which a spoken instruction becomes unambiguous with respect to its referent. The accompanying instruction to this example was ‘Touch the plain red square’.

(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.e.g.Tae nsean’ic wependsncey-474 wependsng si fare moe sobjlt

The adjective *good*

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,

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 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 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. Beca.401ofssible.yad wa T* toeldthe ed9iennged

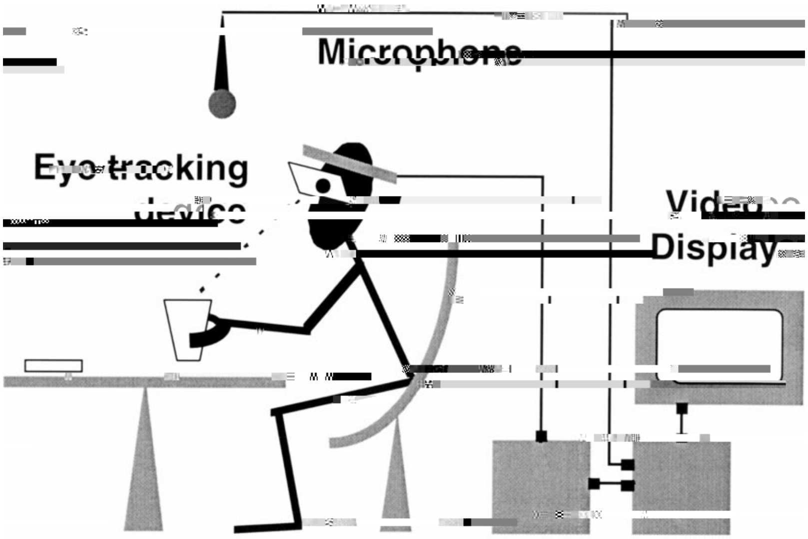


Fig. 2. An illustration of the configuration of the eyetracking equipment. Both eye image and scene image were taken in by camera mounted onto the headband. The CPU computed and superimposed the eye fixation over the scene image, with the resulting video data recorded by the VCR and displayed on the 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

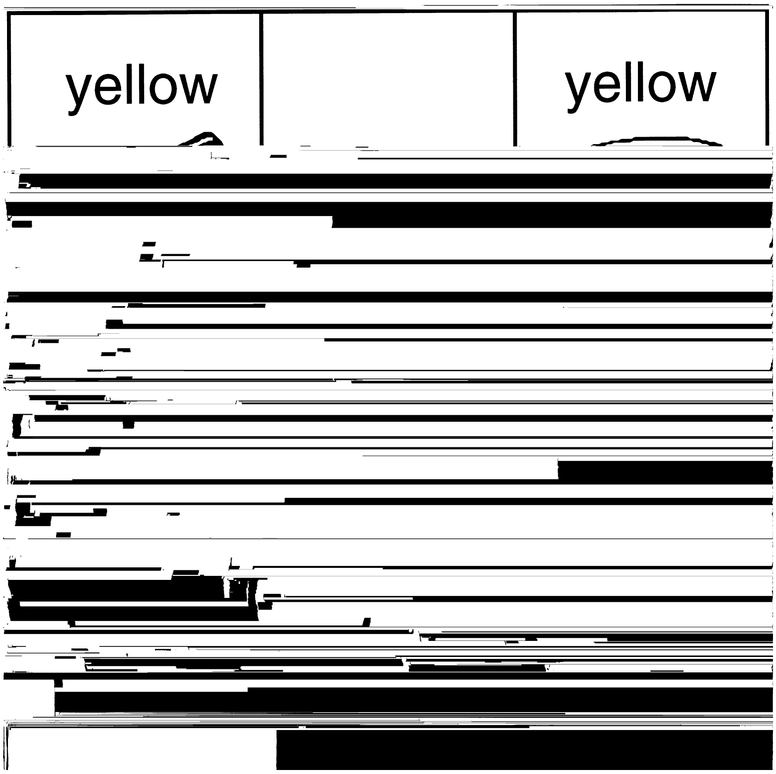
any object other than the target was not reliably different for the two conditions, there were significantly more looks to the competitor object in the late disambiguation condition than there were to the object in the same location for the early disambiguation condition. This difference was reliable 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.



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 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 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 ($F_2(1,19) = 78e25f9$

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

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 hor-

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

ANOVA20355cntro(degr-27.217([(placed)-0 Tcy)-1-stduc42-251)-,2639(ithe)-2493(7(ity)-263(rs)-29493(isplat97)-st)

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

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

manipulated, with contrast and typicality as within-subjects manipulations, and dismain 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 ($F_2(1,13) = 7.2$, $P <$

this task. Table 10 shows the latencies of the first looks to the target in the critical

Fig. 6. Time course of eye movement data for Experiment 3 showing the proportion of trials that contain a

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

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

References

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

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.