

Finding Referents in Time: Eye-Tracking Evidence for the Role of Contrastive Accents

**Andrea Weber¹, Bettina Braun²,
Matthew W. Crocker¹**

¹*Saarland University, Saarbrücken, Germany*

²*Max Planck Institute for Psycholinguistics, Nijmegen,
The Netherlands*

Key words

*contrastive
accent*

eye-tracking

focus

prosody

*reference
determination*

Abstract

In two eye-tracking experiments the role of contrastive pitch accents during the on-line determination of referents was examined. In both experiments, German listeners looked earlier at the picture of a referent belonging to a contrast pair (*red scissors*, given *purple scissors*) when instructions to click on it carried a contrastive accent on the color adjective (L + H*) than when the adjective was not accented. In addition to this prosodic facilitation, a general preference to interpret adjectives contrastively was found in Experiment 1: Along with the contrast pair, a noncontrastive referent was displayed (*red vase*) and listeners looked more often at the contrastive referent than at the noncontrastive referent even when the adjective was not focused. Experiment 2 differed from Experiment 1 in that the first member of the contrast pair (*purple scissors*) was introduced with a contrastive accent, thereby strengthening the salience of the contrast. In Experiment 2, listeners no longer preferred a contrastive interpretation of adjectives when the accent in a subsequent instruction was not contrastive. In sum, the results support both an early role for prosody in reference determination and an interpretation of contrastive focus that is dependent on preceding prosodic context.

1 Introduction

Suppose John is helping Mary to decorate a Christmas tree and she asks to be handed a red bulb. Since their ornament collection includes both red bulbs and red candy canes, John cannot identify the requested ornament on the basis of color alone but has to wait until he hears the specification of the ornament in the noun. Mary, however, might

Acknowledgments: This research was funded by SFB 378 'ALPHA' to the first and third author, and by a Ph.D. scholarship to the second author, both awarded by the German Research Foundation DFG. We would like to thank Amy Schafer and Emiel Krahmer for valuable comments on an earlier version of this paper.

Address for correspondence: Dr. Andrea Weber, FR 4.7 Psycholinguistik, Universität des Saarlandes, 66041 Saarbrücken, Germany; Tel. (+49) (681) 302 6553; Fax (+49) (681) 302 6561; e-mail: <aweber@coli.uni-sb.de>.

When, on the other hand, the accent in the sentence *Please hand me a red BULB* is associated with the noun rather than the adjective, a variety of focus interpretations are possible. Common is a noncontrastive reading of the utterance such as in an answer to the question *Can I help you?* Ladd (1980) introduced the term broad focus for such an interpretation (an idealized F0-track is shown in Fig. 1b). Broad focus is said to project to larger phrases, maximally to the whole utterance. Typically, the noun is marked with a high pitch accent, potentially preceded by a low tonal target (see Ladd & Schepman, 2003). It is still under dispute, however, whether or not accenting the noun always results in a broad focus reading and hence a noncontrastive interpretation. Some researchers claim that a contrastive interpretation of the noun (i.e., a red BULB, not a red candy cane) can be forced by a L+H* accent but this is not generally accepted.²

The main question this paper addresses, is whether listeners are able to use information from contrastively accented adjectives for an early determination of reference. We used the eye-tracking paradigm and German listeners for our investigation. In eye-tracking studies, looks to displayed objects are monitored as an utterance unfolds. This gives us a precise indication of which object in the display is being understood as the intended referent and how long it takes listeners to launch an eye movement to the understood referent (see Tanenhaus, Magnuson, Dahan, & Chambers, 2000, for a review of the paradigm). In 2002, Dahan, Tanenhaus, and Chambers studied the effect of accent on lexical competition using eye-tracking. In a first instruction, participants were asked to move an object in a display (*Put the candle above the triangle*); a second instruction used either an accented or unaccented noun to refer to the same object *candle* (*Now put the CANDLE above the square* vs. *Now put the candle ABOVE THE SQUARE*) or to introduce a new lexical competitor *candy*.³ They found that participants looked more often to the competitor *candy* when the noun in the second instruction was accented (nonanaphoric interpretation) and less often when it was deaccented (anaphoric interpretation). In their study, listeners immediately exploited the relation between pitch accents and discourse for the interpretation of referent nouns.

Not only has it repeatedly been shown with eye-tracking that referents in a scene are identified as soon as they are referred to in an utterance, there are several studies revealing that they can be identified prior to their mention. For example, Altmann and Kamide (1999) have shown that upon hearing the main verb in the sentence *The boy will eat the cake*, listeners prefer to look at edible objects on the display even

² Pierrehumbert and Hirschberg (1990) claimed, for instance, that the accent L+H* is a contrastive version of the more neutral H* (following the ToBI annotation system for intonation patterns by Beckman & Pierrehumbert, 1986). It is not generally accepted, however, that L+H* and H* do indeed represent two categorically distinct accent types. Pitrelli, Beckman, and Hirschberg (1994) describe the L+H* accent as a variant of a H* accent. Ladd and Schepman (2003) argue that a low tonal target is not restricted to L+H* accents but that H* accents can also be preceded by a low target; for accents involving a local F0 rise they argue for a single accent category (L+H)*. Furthermore, empirical evidence from Welby (2003) suggests that listeners' interpretation of focus is affected only minimally, if at all, by the two types of pitch accent.

³ Accented nouns carried either a L+H* or H* pitch accent, followed by a H-L% boundary tone; deaccented nouns carried a H+!H*, followed by a H-H% or H-L% boundary tone.

before they encounter the noun. Weber, Grice, and Crocker (2006), found that in the absence of clear case marking, German listeners anticipate an Object when sentence beginnings have a typical intonation for Subject-first sentences, and anticipate a Subject when sentence beginnings have a typical Object-first intonation (*Die KATZE jagt womöglich der Hund*, 'the cat[ACC] chases possibly the dog[NOM]').

German listeners in our study were asked in two consecutive instructions to click on objects on a computer display while their eye movements were monitored. Whereas the first instruction always introduced one member of a visually displayed contrast pair (*purple scissors*), the second instruction referred to either the other member of the contrast pair (*red scissors*) or to an object differing in form but not color from the other member of the contrast pair (*red vase*). Based on the lexical content of the second instruction alone, the point at which a unique referent could be selected was during the noun, thus after the adjective. In half of our trials, the adjective was unaccented and the nuclear accent was on the noun (*red SCISSORS*), which is the default position for broad focus sentences in German (Ladd, 1996). In the other half, the adjective carried a contrastive L+H* accent (*RED scissors*). Since only the red scissors contrasted in color with another displayed object (i.e., the purple scissors), the accent on the adjective was an appropriate cue for the upcoming referent. If listeners rapidly interpret prosodic information we would expect to find earlier looks to the red scissors when the adjective is accented.

In 1999, Sedivy, Tanenhaus, Chambers, and Carlson (Experiment 1b) tested in an English eye-tracking study whether variation in prosodic focus had an influence on reference resolution of contrastive and noncontrastive referents. They failed to find an effect. The setup of their study was very similar to the setup we used for our experiments. As in our study, their listeners heard two instructions to touch displayed objects. The first instruction referred to one member of a contrast pair (*red bowl*), the second instruction referred to either the other member of the contrast pair (*yellow bowl*) or to an object that shared color or material but not form with the other member of the contrast pair (*yellow comb*).⁴ Sedivy et al. (1999) analyzed latencies of eye movements to the target object of the second instruction after the onset of the noun, and found that contrastive referents were fixated earlier than noncontrastive referents which suggests a contrastive interpretation of adjectival modifiers per se; prosodic marking on the adjective did, however, not influence eye movement latencies.⁵

⁴ Nominal referents in our study were only modified by color adjectives since different materials (e.g., wooden spoon) were hard to depict in line drawings as we used them. Sedivy et al. (1999) displayed actual objects on a table.

⁵ In a previous eye-tracking study Sedivy, Tanenhaus, Spivey-Knowlton, Eberhard, and Carlson (1995) had found preliminary evidence that listeners can use contrastive accents to facilitate disambiguation of referents even without prior introduction of one member of a contrast pair. Contrastive referents in a single instruction had been identified faster when the adjectival modifier carried a contrastive accent. In 1999, Sedivy and colleagues speculated that their failure to replicate this prosodic effect was due to a ceiling effect. Firstly, in 1999 not all filler instructions included adjectival modification thereby potentially heightening the sensitivity to adjective presence. Secondly, including an adjective in the first instruction may have drawn attention to the contrasting object in the second instruction.

We want to argue that contrastively accented adjectives not only allow listeners to identify referents faster, but even help to anticipate upcoming referents. In contrast to Sedivy et al. (1999), we will analyze fixation patterns starting before the onset of the referent noun.

Experiment 1 hence sets out to revisit the role of contrastive accents in the determination of reference in German. Prosodic patterns employed for the instructions were closely matched with the Sedivy et al. (1999) study. The first instruction that introduced one member of a contrast pair (*click on the purple scissors*) always carried a high H* pitch accent on the noun. In the second instruction, a high L+H* pitch accent was placed either on the adjective or the subsequent noun.⁶ Following the findings of Sedivy et al., we predicted that the presence of an adjectival modifier in the second instruction would lead to more fixations of the contrastive object (*red scissors*) than of the noncontrastive object (*red vase*) even when the color adjective was unaccented. Further, we expected that this preference for the contrastive object would significantly be strengthened by a contrastive accent on the adjective. An additional effect of prosodic focus over and above the use of adjectival modifiers would be in opposition to Sedivy et al. who failed to find such an effect. Rather it would suggest that in the case of local indeterminacy, prosodic focus is exploited rapidly by listeners to derive information about the sentence semantics.

Experiment 2 aims at investigating the effect of prosodic marking for reference determination across sentences. More specifically, we want to test whether the contrastive interpretation of adjectival modifiers is influenced by prosodic focus in the preceding linguistic context. As mentioned above, contrasting accents are justified when the context provides a contrast set. This not only includes linguistic but also visual context (see exophoric and endophoric context in Pechmann, 1984); not only mentioning a contrast set but simply having a contrast set present in the environment can be sufficient to license the use of contrastive accents. Our setup fulfills this requirement as the contrast (*purple and red scissors*) is shown on the screen. A contrastive accent for the first mention of a contrast member is therefore appropriate. Prosodically marking the first member might strengthen the salience of the contrast for subsequent instructions. In particular, it might boost listeners' expectations for a contrastively accented second member; as a consequence, second referents that lack contrastive accents might no longer be interpreted contrastively. This would be evidence that adjectives are not automatically interpreted contrastively and that the interpretation of contrastive accents is context sensitive.

The design of Experiment 2 was identical to Experiment 1, except for the first instruction in which we used a contrastive L+H* accent on the adjective. We predicted more fixations on the contrastive *red scissors* than on the noncontrastive *red vase*

⁶ Sedivy and colleagues (1999) did not give a specific description of the prosodic pattern of their first instruction, but from the notation it is evident that there was no contrastive accent on the adjective in their first instruction. Unlike the present study, nouns in second instructions were accented with a H*. Assuming that the two accent patterns on the noun (L+H* and H*) constitute a real difference (see fn. 2 though), this still should not compromise a comparison between our study and Sedivy et al.: in both studies the prosodic manipulation of the adjective is crucial.

only for accented adjectives in the second instruction. The preference for contrastive objects was expected to disappear for unaccented second referents. Such a result would suggest that listeners not only process prosodic information incrementally, but that for the integration of prosodic and semantic structure they take preceding prosodic cues into consideration.

2 Experiment 1

2.1

Method

2.1.1

Participants

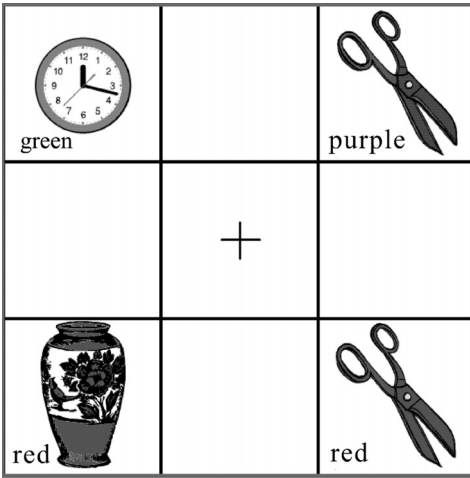
Twenty-four students from Saarland University were paid to take part in the experiment. They were native speakers of German with normal or corrected-to-normal vision and normal hearing.

2.1.2

Materials

Sixteen German nouns referring to illustratable objects were chosen as stimuli. Nouns were modified with a color adjective in the first instruction (e.g., *lila Schere*, ‘purple scissors’, denoted as first referent). For each first referent, two second referents were chosen: one referred to the same object but was modified with a different color adjective (e.g., *rote Schere*, ‘red scissors’, denoted as contrastive referent), the other referred to a different object that matched in color with the contrastive referent (e.g., *rote Vase*, ‘red vase’, denoted as noncontrastive referent). A fourth noun that referred to an object that neither shared form nor color with the other objects (e.g., *Uhr*, ‘clock’, denoted as distractor) was added. The pictures of a first referent, its contrastive and noncontrastive second referents, and the distractor were displayed together on a computer screen (see Fig. 2). To control for potential preferences for displayed objects, the role of first and second referent in a trial was switched for half of the participants and the displays and instructions adjusted accordingly. For example, rather than *lila Schere* (‘purple scissors’) being first referent and *rote Schere* (‘red scissors’) and *rote Vase* (‘red vase’) second referents, *lila Vase* (‘purple vase’) was then first referent and *rote Vase* (‘red vase’) and *rote Schere* (‘red scissors’) second referents.

Each trial consisted of two consecutive instructions to click on an object in the display. The first instruction named the first referent (e.g., *Klicke die lila Schere an*, ‘click on the purple scissors’); the second instruction either referred to the contrastive referent (*contrRef*, e.g., *Klicke jetzt die rote Schere an*, ‘click now on the red scissors’) or to the noncontrastive referent (*noncontrRef*, e.g., *Klicke jetzt die rote Vase an*, ‘click now on the red vase’). The distractor was never named during the experiment. Referents of a trial matched in gender (e.g., *Schere*, *Vase*, and *Uhr* are feminine in German). Since in the instructions referent nouns were preceded by their gender-marked definite article, sharing gender ensured that articles could not disambiguate between potential referents. Furthermore, the sequence of weak and strong syllables

**Figure 2**

Example of visual display presented to participants

in a trial were matched for color adjectives and for referent nouns. Referent nouns and their adjectival modifiers are listed in Appendix A.

To prevent participants from developing expectations that pictures with matching color or form were likely targets, 22 additional filler trials were constructed. Similar to target trials, filler trials also consisted of four displayed objects accompanied by two consecutive instructions to click on an object. Filler trials could differ in various ways: All four objects differed in color and form from one another, object pairs matched in color or form but were never referred to, referent nouns referred to objects with inherent color (e.g., *Banane*, ‘banana’) and were unmodified in the instructions, or both instructions referred to the same object (*Klicke jetzt nochmal darauf*, ‘click now on it again’).

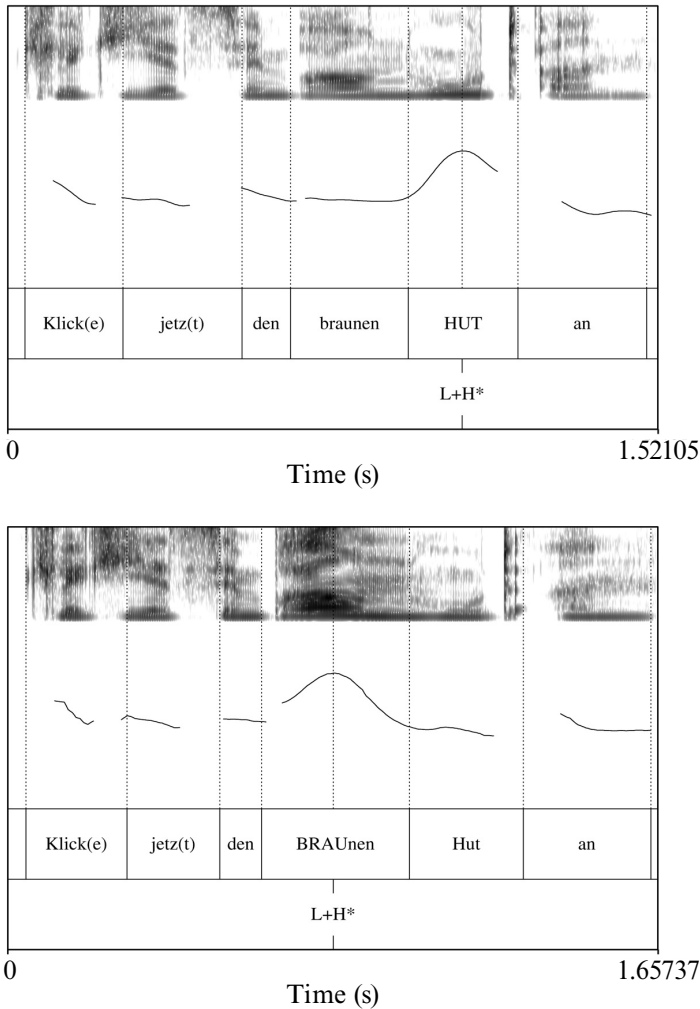
Pictures were selected from a commercially available collection of colored line drawings (IMSI MasterClips, 1990) and further processed using Adobe Illustrator. Spoken instructions were recorded onto DAT in a sound-attenuated room by a phonetically trained female native speaker of German, sampling at 48kHz. The recordings were then down-sampled to 20.48kHz and stored on disc. Instructions were recorded multiple times and the best pronunciation was selected by the first two authors.

The first instruction of experimental trials (e.g., *Klicke die lila Schere an*, ‘click on the purple scissors’) was always recorded with a noncontrastive H* pitch accent on the noun, followed by a low boundary tone. The maximum F0 of the H* was reached in the middle of the stressed vowel of the noun (mean F0-maximum 213.8Hz, *SD* 8.1 Hz). There was no local low tonal target preceding the maximum, and the pitch rose steadily over the utterance.

The second instruction, referring to either the contrastive or the noncontrastive referent was recorded twice, once with a noncontrastive L+H* pitch accent on the noun (*noncontrAccent*) and once with a contrastive L+H* pitch accent on the adjective (*contrAccent*). Instructions always ended with a low boundary tone, and the temporal adverb *jetzt* (‘now’) was never marked with a pitch accent (see Fig. 3).

Figure 3

Representative intonation contours for the second instruction with an accented noun (upper panel) and an accented adjective (lower panel). The example sentence *Klicke jetzt den braunen Hut an*, ‘Click now on the brown hat’, was chosen because of its high proportion of sonorant material



The two factors referent (with the 2 levels *contrRef* and *noncontrRef*) and prosodic accent (with the 2 levels *contrAccent* and *noncontrAccent*) constituted our four experimental conditions (see Table 1). In all four conditions, the L+H* pitch accent was implemented by a low tonal target aligned closely to the onset of the stressed syllable (on average 5ms after syllable onset, *SD* 10.2ms; 8 cases were excluded because F0 tracking proved impossible in the critical region). The F0 maximum was situated in the last quarter of the stressed syllable (on average at 77% of the syllable, *SD* 11%). Durational analyses showed for adjectives a significant main effect of accent,

with accented adjectives being on average 85 ms longer than unaccented adjectives, $F(1, 31) = 578.31, p < .001$. No effect of referent and no interaction between accent and referent was found for adjectives. Accented nouns were on average 15ms longer than unaccented nouns, $F(1, 31) = 9.72, p = .004$. Again, no effect of referent and no interaction between accent and referent was observed for the duration of nouns. Acoustic measurements, separate for each condition, are listed in Appendix B.

Table 1

Example of instructions accompanying the display in Figure 2

	First instruction	Second instruction
contrRef/non-contrAccent	<i>Klicke die lila SCHERE an.</i> 'Click on the purple SCISSORS.'	<i>Klicke jetzt die rote SCHERE an.</i> 'Click now on the red SCISSORS.'
contrRef/contrAccent	<i>Klicke die lila SCHERE an.</i> 'Click on the purple SCISSORS.'	<i>Klicke jetzt die ROTE Schere an.</i> 'Click now on the RED scissors.'
non-contrRef/non-contrAccent	<i>Klicke die lila SCHERE an.</i> 'Click on the purple SCISSORS.'	<i>Klicke jetzt die rote VASE an.</i> 'Click now on the red VASE.'
non-contrRef/contrAccent	<i>Klicke die lila SCHERE an.</i> 'Click on the purple SCISSORS.'	<i>Klicke jetzt die ROTE Vase an.</i> 'Click now on the RED vase.'

To vary prosodic patterns across the experiment, first instructions of filler trials were sometimes recorded with contrastive accent. Similar to experimental trials, some filler trials had contextually less appropriate prosodic patterns (i.e., contrastive accent for a noncontrastive referent). Using Praat (Boersma & Weenik, 1996), 1900ms silence was added between first and second instructions. During this interval, participants had to click on the first-referent object.

Four lists were constructed. Each list contained 16 experimental and 22 filler trials in pseudorandom order, such that before each experimental trial there was at least one filler trial. Filler trials appeared in the same sequential position in all four lists. Each experimental trial also appeared in the same sequential position, but in only one of its four conditions. Each list contained an equal number of trials of all four conditions. For half of the participants, the role of first and second referent of an experimental trial was switched (see above). Three representative practice trials were added at the beginning of each list.

2.1.3

Procedure

At the beginning of a session, participants received written instructions that included an example of a trial display and an explanation of the task. They were then seated in front of a computer monitor. After the eye tracker was calibrated, each participant was presented with one of the trial lists. All pictures were presented in the corner cells of a 3×3 grid with a cross in the middle cell (see Fig. 2). Each cell measured 7.5×7.5 cm. The positions of first and second referent objects were randomized across trials. Spoken instructions started 800ms after the appearance of the pictures on the screen. For each display, participants heard two instructions to click on an object using a computer mouse. Participants were told to look at the center cross after

carrying out the first instruction. Between trials, a dot appeared in the middle of the screen, and participants were instructed to fixate it. The experimenter then initiated an automatic drift correction.

Participants' eye movements were monitored using a SMI EyeLink Hispeed 2D eye-tracking system. A camera on a headband provided the input to the tracker. The center of the pupil was tracked to determine the position of the eye relative to the head. Throughout the experiment, the computer recorded onset and offset times and spatial coordinates of participants' fixations. The sampling rate of the eye tracker was 250Hz. Only the dominant eye of the participant was monitored. Along with the eye movements, time and location of the mouse click were stored. Auditory stimuli were presented binaurally over headphones at a comfortable volume level.

For the analysis, custom-made graphical software was used to display the locations of the participants' fixations as dots superimposed on the four pictures for each trial and each participant. Fixations were coded as pertaining to the cell of the first referent, contrastive referent, noncontrastive referent, the distractor, or the background. Saccade times in which the eye moves from one location to the next and blinks during which the eye tracker receives no signal of the pupil were not added to fixation times.

2.2

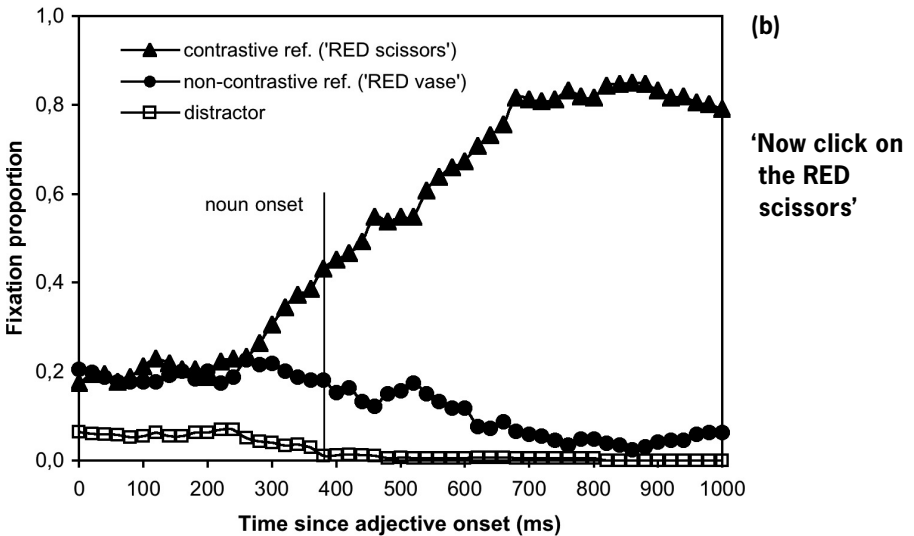
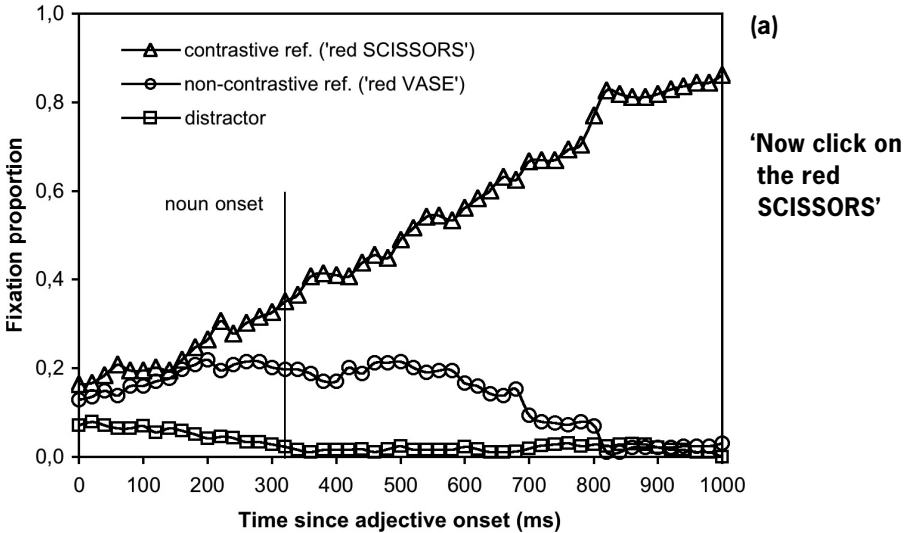
Results and Discussion

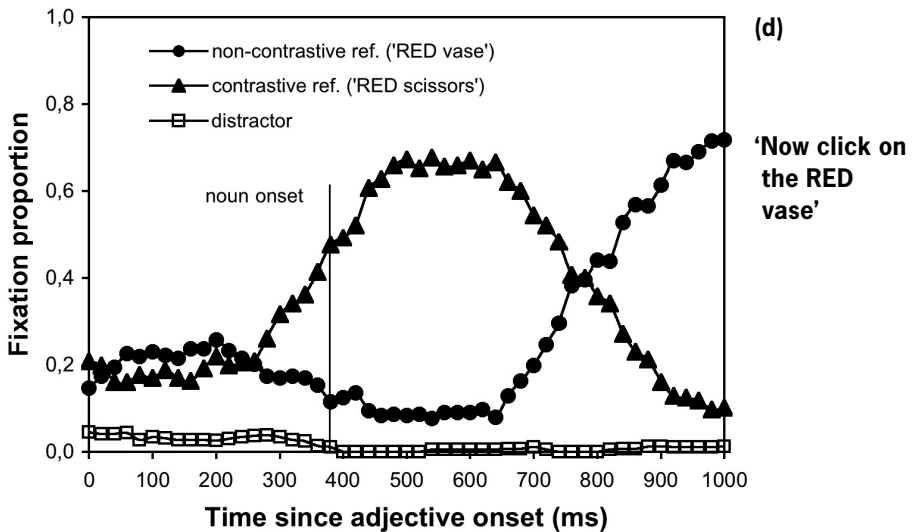
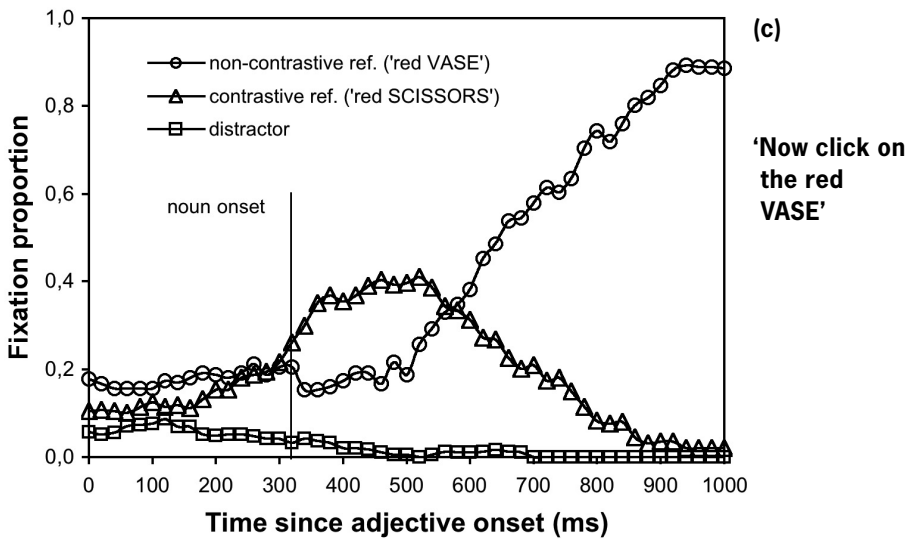
Only fixations occurring during the second instruction of a trial were analyzed. Nine trials were removed from the analysis because participants had clicked on an object other than the target referent (2% of all trials). Fixation proportions at successive 10ms time frames were calculated by adding the number of trials for each participant and each condition in which a picture type (first referent, contrastive referent, noncontrastive referent, distractor, background) was fixated during this interval. The sum for a picture type was then divided by the total sum of fixations during an interval. Fixation proportions were averaged over participants and items for separate analyses. ANOVAs were conducted with the two factors referent and prosodic accent as within-participants, within-items factors. It takes typically about 150 to 200ms before a programed eye movement is launched (e.g., Matin, Shao, & Boff, 1993). Observed fixations are therefore triggered by acoustic information that has been presented about 200ms earlier. For example, fixation proportions at 300ms after adjective onset were driven by approximately the first 100ms of the adjective. Fixation proportions for first referents and distractors were averaged, since none of the critical differences involve a comparison with these two picture types.

Prior to the point that fixations could be driven by acoustic input from the adjective, we found some variation between fixation proportions to the averaged distractor and to the contrastive and noncontrastive referents. Between 0 and 300ms distractors were significantly disfavored in all trials. Overlap in color or shape between objects might have diverted listeners' attention from the distractor. Importantly for this study, however, between 0 and 300ms no variation was found between fixation proportions to contrastive and noncontrastive referents in any condition (all F_1 and

Figure 4

Average fixation proportions over time for contrastive referents, noncontrastive referents, and averaged distractors in Experiment 1: (a) for *contrRef/noncontrAccent* trials, (b) for *contrRef/contrAccent* trials, (c) for *noncontrRef/noncontrAccent* trials, and (d) for *noncontrRef/contrAccent* trials





$F_2 < 1$). Thus, prior to adjective onset, contrastive referents were not fixated more often than noncontrastive referents.

When evaluating trials with instructions to click on the contrastive referent (*click on the red scissors*; Figs. 4a and b), we found that fixation proportions to the *red scissors* started to increase around 300ms after adjective onset, both when the noun

and when the adjective carried the accent. Thus, already before acoustic information of the referent noun was available, participants started to fixate the contrastive referent *red scissors*. No such increase in fixation proportions was found for the simultaneously displayed noncontrastive referent *red vase*, even though it matched in color with the *red scissors*.

A similar increase in early looks to contrastive referents was found for trials that referred to noncontrastive referents (*click on the red vase*; Figs. 4c and d). Evidently, upon encountering the adjective, participants expected the upcoming referent to belong to the contrastive set. For trials with noncontrastive target referents, however, this anticipation of a contrastive referent needed to be revised. Revision seemed to take place once acoustic information of the referent noun became available. Considering the 200ms delay for the execution of an eye movement, fixation proportions for the *red vase* started to increase, and for the *red scissors* to decrease, after the onset of the referent noun *vase*. Anticipation of a contrastive target referent resulted in significantly more looks to the *red scissors* than to the *red vase*. When the adjective was not accented, this difference was fully significant by participants between 300 and 550ms: 36% to contrastive referent, 19% to noncontrastive referent, $F_1(1, 23) = 4.89, p < .05$. By items, the difference was tending towards significance, $F_2(1, 31) = 3.40, p = .07$.⁷ When the adjective was accented the difference between contrastive and noncontrastive referents was robustly significant between 300 and 800ms: 55% to contrastive referent, 16% to noncontrastive referent, $F_1(1, 23) = 69.17, p < .001$; $F_2(1, 31) = 31.56, p < .001$.

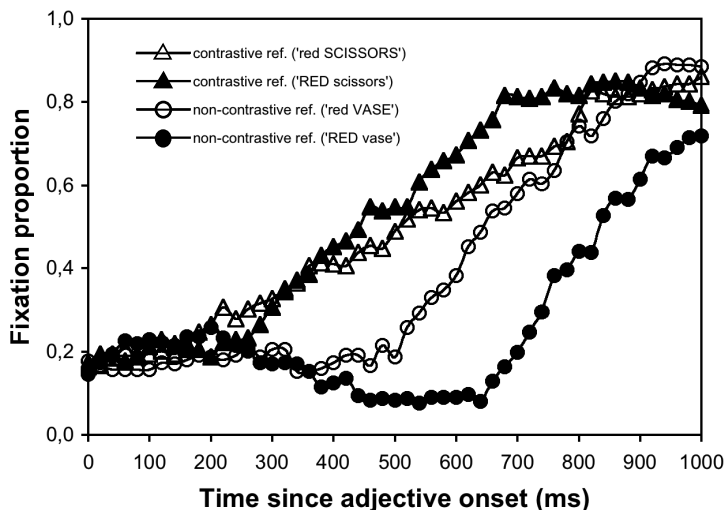
The extent of the contrastiveness effect can be illustrated by comparing the increase in target fixations across all four conditions. Figure 5 plots fixation proportions for target referents when the referent was contrastive (*red scissors*; Figs. 4a and b) and when the referent was not contrastive (*red vase*; Figs. 4c and d). It can be seen that although listeners do eventually look at the target referent to an extent that is similar across conditions, the rate at which they achieve this differs between conditions. Between 300 and 1000ms, fixations proportions were 60% to target referents in *contrRef/noncontrAccent* trials (*red SCISSORS*), 67% in *contrRef/contrAccent* trials (*RED scissors*), 47% in *noncontrRef/noncontrAccent* trials (*red VASE*), and 26% in *noncontrRef/contrAccent* trials (*RED vase*). In a two-factor ANOVA there was a robust main effect of referent, $F_1(1, 23) = 38.34, p < .001$; $F_2(1, 31) = 62.64, p < .001$, such that fixation proportions for contrastive referents increased faster than for noncontrastive referents. We also found a main effect of accent, $F_1(1, 23) = 6.81, p < .05$; $F_2(1, 31) = 5.53, p < .05$, and, crucially, a significant interaction between referent and accent, $F_1(1, 23) = 44.71, p < .001$; $F_2(1, 31) = 11.83, p < .01$. This suggests that accented adjectives speeded recognition of contrastive target referents but hindered recognition of noncontrastive target referents. Note, that durational differences between accented and unaccented adjectives could not be responsible for the pattern of results; looks to the target *red scissors* increased faster when the adjective was accented, even though accented adjectives were on average 73ms longer, therefore making acoustic information of the target noun available later than for trials with noncontrastive

⁷ The difference reached full significance after the exclusion of one particularly extreme item that exhibited a reversal of the effect, $F_2(1, 29) = 5.69, p < .05$.

accent. Fixations for the target *red vase* began to increase around 650ms with contrastive accent and around 500ms with noncontrastive accent. Durational differences between adjectives (68ms on average) alone could not account for this 150ms delay in fixation increase. Indeed, ANCOVAs with adjective durations as covariates, still showed a significant influence of referent, $F_2(1, 30) = 61.29, p < .001$, and an interaction between referent and accent, $F_2(1, 30) = 12.39, p = .001$.

Figure 5

Average fixation proportions over time for target referents when the referent was contrastive (Figs. 4a and b) and when the referent was not contrastive (Figs. 4c and d)



Our findings are in contrast with Sedivy et al.'s (1999) results who, in a very similar study, found an effect of referent, but no interaction between referent and accent. However, both their procedure and their analysis were different. For example, real objects were used and participants were permitted to watch the display as it was being changed between trials, which took approximately 20s. Furthermore, rather than looking at the time course of fixation proportions after adjective onset, Sedivy and colleagues analyzed latencies for first fixations on the target object after noun onset, including only trials in which the initial fixation was to the correct target object. When we analyzed our data accordingly, we too found an effect of referent, $F_1(1, 14) = 5.48, p < .05$; $F_2(1, 20) = 25.72, p < .001$, and no interaction with accent. This result has to be taken with great caution, however, since in only 54% of our trials the first fixation after noun onset was on the target object (as opposed to more than 80% in Sedivy et al.). We suggest that listeners in our study fixated objects other than the target object much more often, because the preceding exposure to the display was much shorter. More importantly, in our study, fixation probabilities over time revealed that listeners start fixating referent objects prior to noun onset. Anticipatory eye movement latencies were not considered in the Sedivy et al. study which might be the reason why they failed to uncover the interaction between referent and accent.

In sum, the results of Experiment 1 suggest that following the introduction of one member of a contrast pair, listeners interpreted subsequent adjectival modifiers as referring to the other member of that contrast pair. Prosodic marking of the adjective heightened the contrastive interpretation even further. As described above, not only linguistic context but also visual context can justify the use of contrastive accents. Thus using contrastive accents in a first instruction in our setup can be considered appropriate since the contrast pair is present in the visual context. Identifying one member of a contrast pair with contrastive marking might strengthen the salience of the contrast pair for subsequent instructions. In particular, introducing the first member of a contrast pair with contrastive accent (e.g., *LILA Schere*, 'PURPLE scissors') might boost listeners' expectations for a contrastively accented second member (e.g., *ROTE Schere*, 'RED scissors'). When second referents lack contrastive accents they might therefore no longer be interpreted contrastively. We tested this hypothesis in Experiment 2 by presenting the materials of Experiment 1 again, this time using a first instruction that carried a contrastive accent rather than a noncontrastive accent.

3 Experiment 2

3.1

Method

3.1.1

Participants

Twenty-four native speakers of German, students at Saarland University, were paid to take part in the experiment. They had not participated in Experiment 1. As before, all had normal or corrected-to-normal vision and normal hearing.

3.1.2

Materials

The materials were as in Experiment 1, except that in the first instruction the adjective rather than the noun carried the accent. For example, the first instruction *Klicke die lila SCHERE an*, 'click on the purple SCISSORS', with a H* accent on the noun, became in Experiment 2 *Klicke die LILA Schere an*, 'click on the PURPLE scissors', with a L+H* accent on the adjective. The bitonal pitch accent had a low tonal target aligned to the start of the stressed syllable of the adjective. On average, the F0 minimum was at 171.8 Hz (*SD* 11.2 Hz). The high target was positioned in the last quarter of the stressed syllable and was realized with 241.8 Hz (*SD* 18.0 Hz). There was a sudden drop in F0 after the maximum, and the noun was deaccented. The second instructions had the same four phonological contours as in Experiment 1. The recordings for Experiments 1 and 2 were conducted on different days. To avoid differences in voice quality, pitch range, and speaking rate between first and second instructions in Experiment 2, the second instructions were rerecorded. Prosodically, the second instructions of the two recording sessions were very similar with only minor differences in the phonetic implementation of the contours. As in Experiment 1, the low tonal target of the L+H* pitch accent was in all four conditions positioned closely to the

onset of the stressed syllable (on average 6 ms after syllable onset, SD 9 ms; nine cases were excluded because F_0 tracking proved impossible in the critical region). The F_0 maximum was reached in the last quarter of the stressed syllable (on average at 70% of the syllable, SD 16%). As in Experiment 1, adjectives in the second instruction were significantly longer when they were accented than when they were unaccented: 78 ms, $F(1, 31) = 219.92$, $p < .001$. Neither the effect of referent nor the interaction between referent and accent were significant. Likewise, accented nouns were longer than unaccented ones: 6 ms, $F(1, 31) = 6.32$, $p < .02$, with no effect of referent or interaction. For acoustic measurements per condition see Appendix B.

3.1.3

Procedure

The procedure was the same as in Experiment 1.

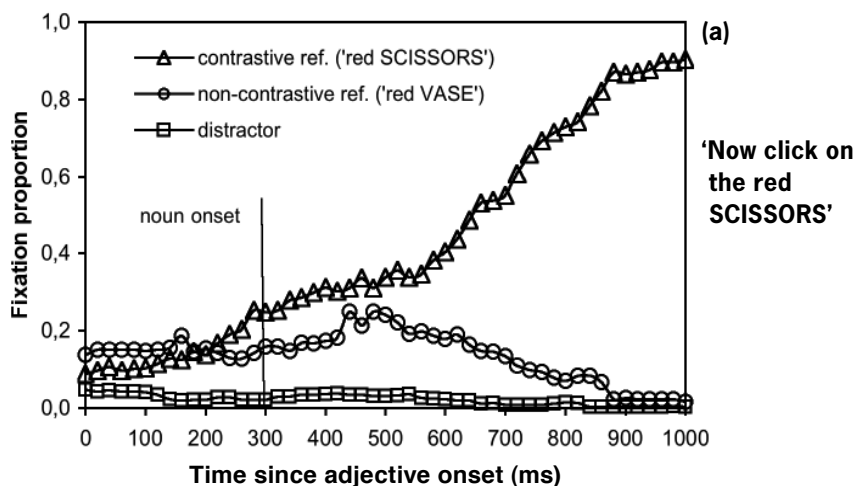
3.2

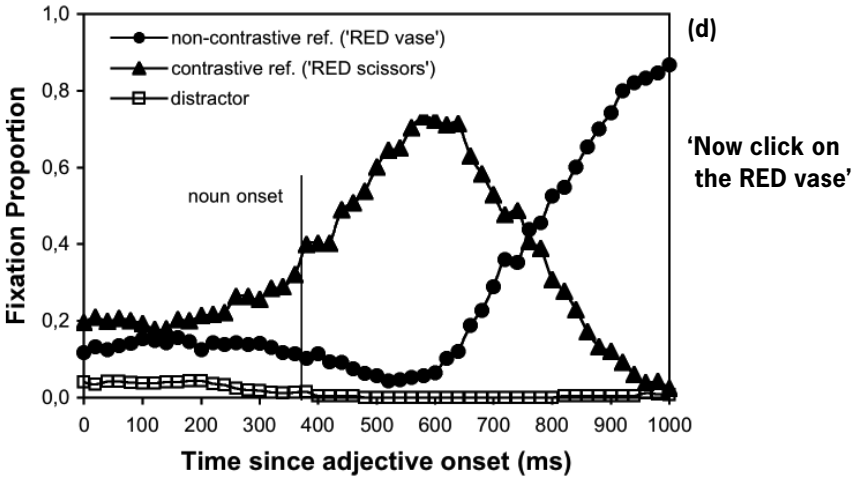
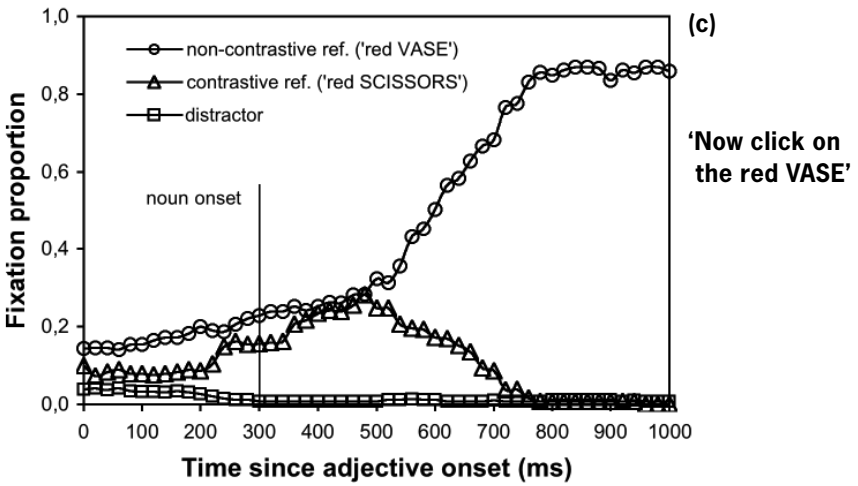
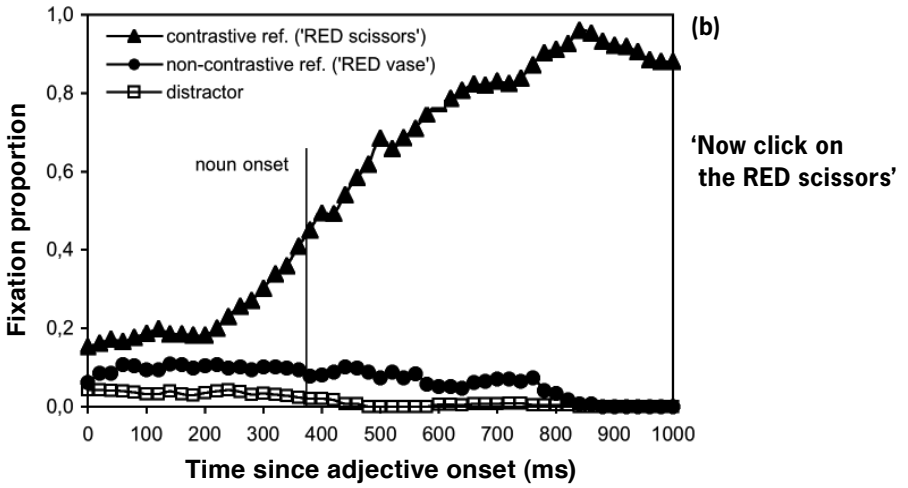
Results and Discussion

As in Experiment 1, only fixations occurring during the second instruction were analyzed. Twenty-five trials were removed from the analysis because participants had clicked on an object other than the target referent or no fixation on the target object was found (6.5% of all trials). Figure 6 presents the averaged proportions of fixations for trials with contrastive referents (*red scissors*; Figs. 6a and b) and noncontrastive referents (*red vase*; Figs. 6c and d). Fixation proportions for first referents and distractors were again averaged.

Figure 6

Average fixation proportions over time for contrastive referents, noncontrastive referents, and averaged distractors in Experiment 2: (a) for *contrRef/noncontrAccent* trials, (b) for *contrRef/contrAccent* trials, (c) for *noncontrRef/noncontrAccent* trials, and (d) for *noncontrRef/contrAccent* trials





As before, some early variation between fixation proportions for different picture types was found. Between 0 and 300ms after adjective onset, the averaged distractor was disfavored in all but *contrRef/contrAccent* trials. The crucial comparison for this experiment was, however, between contrastive and noncontrastive referents, for which no significant difference in fixation proportions was found between 0 and 300ms in any of the four conditions: highest variation for *contrRef/contrAccent* trials, $F_1(1, 23) = 3.38, p > .07$; $F_2(1, 31) = 2.05, p > .1$. A difference between fixations to the contrastive and noncontrastive referent in a later time window can therefore not be attributed to a general bias toward one of the referents.

Figures 6a and b show that regardless of accent placement, fixation proportions to the contrastive target *red scissors* started to increase prior to noun onset, just as they had done in Experiment 1. Already 200 to 300ms after adjective onset, increasingly more looks went to the anticipated *red scissors*. At variance with Experiment 1, however, the probability of fixating the *red scissors* seemed to diverge from that of the *red vase* at different times for different accent placements. For trials with accented adjectives (Fig. 6b), the divergence started around 250ms after adjective onset. For trials with unaccented adjectives (Fig. 6a), both referent types seemed to have received a similar amount of looks until 500ms after adjective onset. ANOVAs, however, revealed a marginal significant difference from 300 to 550ms between fixation proportions for contrastive and noncontrastive referents, $F_1(1, 23) = 4.28, p = .05$; $F_2(1, 31) = 5.34, p < .05$. Nevertheless, 500ms after adjective onset, the difference between contrastive and noncontrastive referents was much smaller when the adjective was unaccented (10%; Fig. 6a) than when the adjective was accented (61%; Fig. 6b), suggesting that noncontrastive accents reduced the expectation of an upcoming contrastive referent.

This effect is even clearer when looking at instructions with noncontrastive target referents (*click on the red vase*; Figs. 6c and d). When instructions carried a noncontrastive accent (Fig. 6c), looks to the *red vase* increased along with looks to the *red scissors* until 500ms after adjective onset when fixation proportions for the *red scissors* started to drop. Between 300 and 500ms, the mean proportion of fixations was 23% for noncontrastive referents and 28% for contrastive referents (F_1 and $F_2 < 1$).⁸ The preference to interpret adjectives contrastively was apparently neutralized when the adjective was unaccented. This finding is in obvious contrast with Experiment 1 (Fig. 4c), where even unaccented adjectives were interpreted contrastively. We suggest that contrastive accents in the first instruction in Experiment 2 highlighted the contrast set and that therefore listeners expected contrastive accents for the contrast set in the second instruction too. When second instructions did not carry a contrastive accent, the *red scissors* were no longer preferred. When second instructions did carry a contrastive accent (Fig. 6d), on the other hand, listeners started fixating the *red scissors* as soon as they encountered the adjective. Thus, as in Experiment 1, listeners initially interpreted an accented adjective as referring to the contrastive referent; this initial interpretation was only corrected after some acoustic information of the noun of the noncontrastive target referent was available. The probability of fixating the *red scissors* was greater than that of the *red vase* until approximately 800ms after adjective

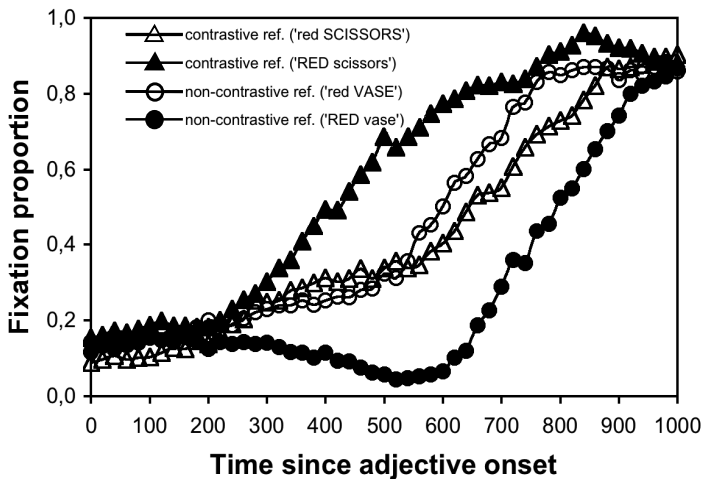
⁸ Also between 300 and 550ms, the time window used in Experiment 1, no significant difference was found.

onset. Between 300 and 800ms, the proportion of fixations was 48% for the contrastive referent and 17% for the noncontrastive referent. A two-factor ANOVA showed that the difference was significant, $F_1(1, 23) = 22.39, p < .001$; $F_2(1, 31) = 39.95, p < .001$.

Again, the extent of the contrastiveness effect can be demonstrated best by comparing the increase in target fixations across all four conditions. Figure 7 shows fixation proportions for contrastive target referents (taken from Figs. 6a and b) and for noncontrastive target referents (taken from Figs. 6c and d), both when the adjective was accented and when it was not. Between 300 and 1000ms, 53% of the fixations went to the target in *contrRef/noncontrAccent* trials (*red SCISSORS*), 71% in *contrRef/contrAccent* trials (*RED scissors*), 57% in *noncontrRef/noncontrAccent* trials (*red VASE*), and 31% in *noncontrRef/contrAccent* trials (*RED vase*). Two-factor ANOVAs, comparable to the ones in Experiment 1, showed a main effect of referent between 300 and 1000ms, $F_1(1, 23) = 42.36, p < .001$; $F_2(1, 31) = 30.12, p < .001$, such that fixation proportions for the contrastive referent *red scissors* increased faster than for the noncontrastive referent *red vase*, and a significant interaction between referent and accent, $F_1(1, 23) = 105.64, p < .001$; $F_2(1, 31) = 35.76, p < .001$. The cross-over interaction was balanced such that the factor accent was not significant, $F_1(1, 23) = 1.53, p > .1$; $F_2(1, 31) = 2.56, p > .1$. Post hoc pair-wise comparisons, however, showed a significant effect of accent for both contrastive referents, $F_1(1, 23) = 18.72, p < .001$; $F_2(1, 31) = 17.96, p < .001$, and noncontrastive referents, $F_1(1, 23) = 40.34, p < .001$; $F_2(1, 31) = 30.26, p < .001$. As in Experiment 1, accented adjectives speeded the recognition of the contrastive *red scissors*. When, however, the spoken instruction referred to the noncontrastive referent *red vase* and the adjective was accented, recognition of the *red vase* was slowed down. As in Experiment 1, durational differences between accented and unaccented adjectives could not be responsible for the pattern of results: ANCOVAs with adjective durations as covariates still showed a significant influence of referent, $F_2(1, 30) = 38.02, p < .001$, and an interaction between referent and accent, $F_2(1, 30) = 36.16, p < .001$.

Figure 7

Average fixation proportions over time for target referents when the referent was contrastive (Figs. 6a and b) and when the referent was not contrastive (Figs. 6c and d)



By replicating the interaction between referent and accent from Experiment 1, we also replicated the difference in results with Sedivy et al. (1999) who had found no such interaction when they looked at eye movement latencies to target objects after noun onset. We thereby provide further support that fixation proportions after adjective onset are better suited to investigate the role of contrastive accent for referent identification.

In sum, listeners in Experiment 2 interpreted accented adjectives contrastively. Upon hearing an accented adjective, contrastive referents received more looks than noncontrastive referents. This is in line with Experiment 1, where contrastive referents were also preferred when the adjective carried an accent. Crucially different from Experiment 1, however, contrastive referents in Experiment 2 were no longer preferred when the adjective was unaccented; rather, contrastive and noncontrastive referents in Experiment 2 then received an equal amount of looks. In Experiment 1, the first instruction carried a noncontrastive accent, in Experiment 2, a contrastive accent. Thus introducing the first member of a contrast pair with contrastive accent in Experiment 2, appeared to have raised listeners' expectations for the second member to be referred to with contrastive accent too. In consequence, noncontrastive accents neutralized the contrastive interpretation of adjectives.

4 General discussion

Our experiments support the view that listeners rapidly exploit prosodic information for the interpretation of referential expressions: German listeners identified referents in a contrast set faster when instructions to click on the referent picture carried a contrastive accent. Referential expressions in our experiments were modified with a color adjective (*red scissors*). Experiment 1 showed that contrastive accents on the adjective (L+H*) speeded referent determination over and above the contrastive interpretation found for adjectives as such. That is, during the adjective we found more looks to referents that contrasted in color with another displayed object (*red scissors* when simultaneously *purple scissors* were shown) than to referents that did not contrast in color (*red vase*). The early preference for the contrastive referent was not only observed when the adjective carried a contrastive accent but also when it did not. This suggests that prosodic focus was not necessary for a contrastive interpretation of the color adjective. However, prosodic focus further strengthened the contrastive interpretation: contrastively accented adjectives induced even earlier looks to contrastive referents than unaccented adjectives did. Thus, by manipulating prosodic focus, it was possible to alter the point in the instructions which allowed for the identification of a unique referent. Listeners exploited contrastive accents on preceding adjectives rapidly enough to anticipate target referents even before the referent noun was mentioned.

In Experiment 2, we investigated whether the contrastive function of adjectival modifiers was sensitive to prosodic context influences. In both Experiments 1 and 2, listeners heard two consecutive instructions to click on objects in a display. The first instruction introduced one member of a contrast pair, the second instruction either referred to the other member of the contrast pair or to a noncontrasting referent. In Experiment 2, but not in Experiment 1, the first instruction carried a contrastive

accent on the adjective. Counter to Experiment 1, we found in Experiment 2 that the contrastive function of adjectives depended on prosodic focus. German listeners only looked more often at the contrastive referent than at the noncontrastive referent when the adjective in the second instruction was accented. When it did not carry a contrastive accent, the contrastive referent was no longer preferred. While both experiments showed that contrastive accents are exploited rapidly to derive information about sentence semantics, Experiment 2 in addition demonstrated that listeners take preceding prosodic information into consideration for the determination of referents. An unaccented color adjective was no longer sufficient for a contrastive interpretation when the other member of the contrast set had been introduced with an accented color adjective.

Our finding that prosodic focus facilitates the identification of referents in German is in contrast with Sedivy et al.'s (1999) study which failed to find such an effect of prosody for English listeners. Since the accentual realizations are very similar in English and German (see Grabe, 1998), the difference in findings cannot easily be attributed to language differences. In the study by Sedivy and colleagues, trials of the contrastive accent study were presented together with trials of a study that did not manipulate prosody, and rather than using prerecorded stimuli, stimuli were read aloud by the experimenter from a script. Furthermore, no specific requirements for the realization and the placement of pitch accents in the first instruction were given. It is possible that the combination of these factors caused too much variation in pitch accents to allow an influence on listeners' behavior. Sedivy and colleagues themselves explained the lack of a prosodic effect as a ceiling effect caused by a prominent position of adjectives in the experimental setup (see fn. 5). Since the experimental setup was identical with our study this seems unlikely. We rather believe that the time window Sedivy et al. used for their analysis may have been responsible for the difference in findings. Looking at the time course of fixation proportions in our experiments we found that participants started fixating potential referents during the adjective, prior to the referent noun. That is, listeners anticipated target referents. Sedivy et al. analyzed latencies of eye movements after the onset of the noun, thereby excluding trials in which participants started looking at the target referent during the adjective. Excluding trials with early looks to the target referent might have been responsible for the failure to find an effect of prosodic focus for reference processing.

The placement of pitch accent has been shown to influence speech comprehension at various levels. It has been shown, for example, that phonemes in accented syllables are detected more quickly (Cutler & Foss, 1977; Shields, McHugh, & Martin, 1974), that multiple meanings of homophones are activated if the words are in focus (Blutner & Sommer, 1988), and that word forms are remembered better when the word was accented (Birch & Garnsey, 1995). But also the interpretation of structural ambiguities (Carlson, 2001; Schafer, Carter, Clifton, & Frazier, 1996; Snedeker & Trueswell, 2003; Weber et al., in press) as well as the processing of discourse structure have been shown to be influenced by accent placement (e.g., Baumann & Hadelich, 2003; Hruska & Alter, 2004; Read, Kraak, & Boves, 1980; Terken & Nooteboom, 1987; Toepel & Alter, 2004). In the choice of accent placement, speakers encode important information for listeners, be it information about syntactic structuring or sentence semantics. The listeners' task is then to identify the accented words and interpret

their meaning. A long-standing question has been how quickly listeners can make use of such prosodic cues for comprehension. Recent studies using the eye tracking paradigm support a very rapid and early role for prosody in comprehension. For example, Snedeker and Trueswell (2003) found that when utterances are syntactically ambiguous, prosodic information can guide listeners' interpretation even prior to the onset of the ambiguous phrase. In their study, prosodic information had an early influence on whether English listeners preferred a modifier or instrument interpretation for ambiguous prepositional phrase attachments (e.g., "with the feather"). But also semantic processing has been shown to be influenced by prosody early on. For example, Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1996) found evidence that the determination of reference can be speeded by pitch accents. Accented adjectives in their study (e.g., "click on the LARGE blue triangle") helped English listeners to delimit the set of potential referents (see also Sedivy, Tanenhaus, Spivey-Knowlton, Eberhard, & Carlson 1995). Our results are in line with the assumption that semantic processing is incremental and that listeners compute contrast sets immediately when they encounter an utterance with contrastive accent, at least when the visual context provides potential referents. Furthermore, they suggest that prosody can allow reference determination before the disambiguating referent name is mentioned.

The interpretation of adjectives in our study was influenced by the preceding use of contrastive accents (see Experiment 2). Observed context effects spread over more than one utterance and unaccented adjectives were only interpreted contrastively when a previous adjective with a contrastive function was unaccented as well. Thus, rather than a simple one to one mapping between prosodic information and semantic interpretation, the results suggest that broader discourse context influenced the appraisal of prosodic information. There is evidence from research that concentrates on the relation between accents and information structure that the interpretation of prosodic cues is context dependent (see Cutler, Dahan, & Doneslaar, 1997, for a review). Birch and Clifton (1995) found, for instance, that listeners' judgments of prosodic appropriateness in question-answer pairs were better when new information in the answer was accented and given information de-accented. Dahan et al. (2002) found in an eye-tracking study that listeners interpreted only deaccented nouns anaphorically. Even though evidence from studies in information structure support the view that the interpretation of prosody is context dependent, they usually differ from the present study in that they manipulate prior discourse structure and not just previously encountered intonation patterns.

Taken together, we conclude from our results that (a) prosodic focus can facilitate on-line referential processes and that (b) the interpretation of prosodic information is dependent upon expectations set up in preceding sentences. The rapid influence of prosody in triggering anticipatory eye movements suggests not only that prosodic information is used in the earliest stages of processing, but also that its interpretation is guided by context. Thus, incoming information needs to be processed incrementally and, at least for prosodic information, the interpretation not be fixed a priori.

manuscript received: 05. 15. 2005

manuscript accepted: 11. 15. 2005

References

- ALTMANN, G., & KAMIDE, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, **72**, 247–264.
- BAUMANN, S., & HADELICH, K. (2003). Accent type and givenness: An experiment with auditory and visual priming. In *Proceedings of the 15th International Conference of Phonetic Sciences 1811–1814*. Barcelona, Spain.
- BECKMAN, M. E., & PIERREHUMBERT, J. Z. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, **3**, 255–309.
- BIRCH, S., & CLIFTON, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, **38**, 365–391.
- BIRCH, S., & GARNSEY, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language*, **34**, 232–267.
- BLUTNER, R., & SOMMER, R. (1988). Sentence processing and lexical access: The influence of the focus-identifying task. *Journal of Memory and Language*, **27**, 359–367.
- BOERSMA, P., & WEENIK, D. (1996). Praat, a system for doing phonetics by computer, version 3.4. Institute of Phonetic Sciences of the University of Amsterdam, Report 132. (updated version on <<http://www.praat.org>>.)
- BOLINGER, D. L. (1978). Intonation across languages. In J. H. Greenberg (Ed.), *Universals of human language*, Vol. 2. *Phonology* (pp.471–524). Stanford: Stanford University Press.
- BUERING, D. (1997). *The meaning of topic and focus — The 59th Street Bridge accent*. London: Routledge.
- CARLSON, K. (2001). The effects of parallelism and prosody in the processing of gapping structures. *Language and Speech*, **44**, 1–26.
- CUTLER, A., DAHAN, D., & van DONSELAAR, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, **40**, 141–201.
- CUTLER, A., & FOSS, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, **20**, 1–10.
- DAHAN, D., TANENHAUS, M., & CHAMBERS, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, **47**, 292–314.
- GRABE, E. (1998). Comparative intonational phonology: English and German. Ph.D. Dissertation, University of Nijmegen (Max Planck Institute series in Psycholinguistics, 7), The Netherlands. Wageningen: Ponsen and Looijen bv.
- GRICE, M., & BAUMANN, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte*, **191**, 267–298.
- GRICE, M., BAUMANN, S., & BENZMUELLER, R. (2005). German intonation in autosegmental-phonology. In S.-A. Jun (Ed.), *Prosodic typology*. Oxford: Oxford University Press.
- GUSSENHOVEN, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris Publications.
- HRUSKA, C., & ALTER, K. (2004). Prosody in dialogues and single sentences: How prosody can influence speech perception. In A. Steube (Ed.), *Information structure: Theoretical and empirical aspects* (pp.221–226). Berlin: Walter de Gruyter.
- IMSI MASTERCLIPS, (1990). Premium image collection 303,000 [Image database]. (<<http://www.imsisoft.com>>)
- JACKENDOFF, R. (1972). *Semantic interpretation in generative grammar*. Cambridge: MIT Press.
- KRAHMER, E., & SWERTS, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, **34**, 391–405.
- LADD, D. (1980). *The structure of intonational meaning: Evidence from English*. Bloomington: Indiana University Press.
- LADD, D. R. (1996). *Intonational phonology*. Cambridge University Press, Cambridge, U.K.
- LADD, D., & SCHEPMAN, A. (2003). Sagging transitions between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, **31**, 81–112.

- MATIN, E., SHAO, K., & BOFF, K. (1993). Saccadic overhead: Information processing time with and without saccades. *Perceptual Psychophysics*, **53**, 372–380.
- PECHMANN, T. (1984). *Überspezifizierung und Betonung in referentieller Kommunikation*. Ph.D. Dissertation. University of Mannheim.
- PIERREHUMBERT, J. (1990). *The phonology and phonetics of English intonation*. Ph.D. Dissertation. Cambridge: MIT Press.
- PIERREHUMBERT, J., & HIRSCHBERG, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Coehn et al. (Eds.), *Intentions in communication*. Cambridge: MIT Press.
- PITRELLI, J., BECKMAN, M., & HIRSCHBERG, J. (1994). Evaluation of prosodic transcription labelling reliability in the ToBI framework. In *Proceedings of the Third International Conference on Spoken Language Processing* (pp. 123–126). Yokohama, Japan.
- READ, C., KRAAK, A., & BOVES, L. (1980). The interpretation of ambiguous who-questions in Dutch: The effect of intonation. In W. Zonneveld & F. Weerman (Eds.), *Linguistics in the Netherlands, 1977–1979* (pp. 389–410). Dordrecht: Foris.
- ROOTH, M. (1985). *Association with focus*. Ph.D. dissertation. University of Massachusetts at Amherst. Amherst: GLSA.
- ROOTH, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, **1**, 75–116.
- SCHAFER, A., CARTER, C., CLIFTON, C., & FRAZIER, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes*, **11**, 135–163.
- SEDIVY, J., TANENHAUS, M., CHAMBERS, C., & CARLSON, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, **71**, 109–147.
- 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* (pp. 375–380). Hillsdale, NJ: Erlbaum.
- SELKIRK, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge: MIT Press.
- SHIELDS, J. L., MCHUGH, A., & MARTIN, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, **102**, 250–255.
- SNEDEKER, J., & TRUESWELL, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, **48**, 103–130.
- STEEDMAN, M. (1989). Grammar, interpretation, and processing from the lexicon. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 463–504). Cambridge: MIT Press.
- TANENHAUS, M., MAGNUSON, J., DAHAN, D., & CHAMBERS, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, **29**, 557–580.
- TANENHAUS, M., SPIVEY-KNOWLTON, M., EBERHARD, K., & SEDIVY, J. (1996). Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. In T. Inui & J. McClelland (Eds.), *Attention and performance XVI: Integration in perception and communication* (pp. 457–478). Cambridge, MA: MIT Press.
- TERKEN, J., & NOOTEBOOM, S. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*, **2**, 145–163.
- TOEPEL, U., & ALTER, K. (2004). On the independence of information structural processing. In A. Steube (Ed.), *Information structure: Theoretical and empirical aspects* (pp. 227–240). Berlin: Walter de Gruyter.
- WEBER, A., GRICE, M., & CROCKER, M. C. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, **99**, B63–B72.
- WELBY, P. (2003). Effects of pitch accent position, type, and status on focus projection. *Language and Speech*, **46**, 53–81.

Appendix A

Experimental stimuli used in Experiments, 1 & 2, with English translations

<i>First referent</i>	<i>Contrastive second referent</i>	<i>Noncontrastive second referent</i>	<i>Unrelated distractor</i>
roter Kamm 'red comb'	gelber Kamm 'yellow comb'	gelber Ball 'yellow ball'	Hammer 'hammer'
rosa Schal 'pink shawl'	blauer Schal 'blue shawl'	blauer Fön 'blue hairdryer'	Löffel 'spoon'
grüner Sonnenschirm 'green sunshade'	roter Sonnenschirm 'red sunshade'	roter Teddybär 'red teddybear'	Wecker 'alarm'
rote Blume 'red flower'	gelbe Blume 'yellow flower'	gelbe Schale 'yellow bowl'	Flöte 'flute'
lila Schere 'purple scissors'	rote Schere 'red scissors'	rote Vase 'red vase'	Uhr 'clock'
grüne Lampe 'green lamp'	rosa Lampe 'pink lamp'	rosa Bluse 'pink blouse'	Gabel 'fork'
brauner Koffer 'brown suitcase'	grüner Koffer 'green suitcase'	grüner Eimer 'green bucket'	Schlitten 'sledge'
blaues Flugzeug 'blue airplane'	rotes Flugzeug 'red airplane'	rotes Sparschwein 'red piggybank'	Glas 'glas'
grünes Telefon 'green telephone'	blaues Telefon 'blue telephone'	blaues Lineal 'blue ruler'	Messer 'knife'
gelbes Hemd 'yellow shirt'	lila Hemd 'purple shirt'	lila Buch 'purple book'	Fass 'barrel'
grüne Flasche 'green bottle'	braune Flasche 'brown bottle'	braune Jacke 'brown coat'	Leiter 'ladder'
orange Mütze 'orange cap'	türkise Mütze 'turquoise cap'	türkise Kanne 'turquoise jug'	Bürste 'brush'
türkises T-Shirt 'turquoise t-shirt'	oranges T-Shirt 'orange t-shirt'	oranges Kissen 'orange pillow'	Paddel 'paddle'
lila Hut 'purple hat'	brauner Hut 'brown hat'	brauner Stuhl 'brown chair'	Pinsel 'paint brush'
gelber Becher 'yellow mug'	grüner Becher 'green mug'	grüner Stiefel 'green boot'	Nagel 'nail'
blaue Maske 'blue mask'	rote Maske 'red mask'	rote Kerze 'red candle'	Geige 'violine'

Appendix B***Acoustical measurements for the second instruction in Experiments 1 and 2****Mean duration (and SD) in ms*

	<i>preceding context "click on the"</i>	<i>adjective</i>	<i>noun</i>	
<i>Experiment 1</i>				
a) contrRef/noncontrAccent	569.7 (45.5)	301.1 (61.7)	424.9 (93.7)	
b) contrRef/contrAccent	580.4 (47.6)	388.6 (56.8)	413.3 (94.3)	
c) noncontrRef/noncontrAccent	565.1 (40.4)	303.1 (60.1)	431.4 (97.4)	
d) noncontrRef/contrAccent	580.3 (42.6)	386.1 (57.7)	413.7 (96.4)	
<i>Experiment 2</i>				
a) contrRef/noncontrAccent	555.8 (34.7)	301.1 (50.7)	415.9 (93.2)	
b) contrRef/contrAccent	582.8 (44.9)	383.4 (52.8)	410.3 (94.5)	
c) noncontrRef/noncontrAccent	558.6 (42.3)	302.2 (53.3)	420.3 (91.3)	
d) noncontrRef/contrAccent	583.0 (39.6)	375.5 (59.5)	413.3 (90.0)	
	<i>Position of L relative to onset of stressed syllable in ms</i>	<i>Height of L in Hz</i>	<i>Proportion H with respect to duration of stressed syllable</i>	<i>Height of H in Hz</i>
<i>Experiment 1</i>				
a) contrRef/noncontrAccent	4.6 (26.3)	166.8 (6.1)	0.76 (0.14)	247.1 (10.5)
b) contrRef/contrAccent	7.0 (6.9)	170.2 (11.5)	0.78 (0.08)	252.4 (11.7)
c) noncontrRef/noncontrAccent	1.4 (10.1)	167.3 (6.1)	0.75 (0.14)	247.3 (11.3)
d) noncontrRef/contrAccent	7.2 (7.6)	170.3 (6.4)	0.79 (0.08)	252.1 (11.3)
<i>Experiment 2</i>				
a) contrRef/noncontrAccent	3.3 (10.7)	182.7 (10.3)	0.67 (0.18)	225.7 (11.5)
b) contrRef/contrAccent	9.9 (7.4)	182.4 (12.0)	0.72 (0.09)	234.9 (10.7)
c) noncontrRef/noncontrAccent	3.1 (9.1)	184.9 (10.9)	0.70 (0.22)	227.1 (10.9)
d) noncontrRef/contrAccent	7.3 (7.7)	181.5 (12.4)	0.71 (0.10)	233.8 (12.5)