Does Narrow Focus Activate Alternative Referents?

Bettina Braun‡*, Andrea Weber*∗*, Matthew Crocker*∗

‡Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands
bettina.braun@mpi.nl

∗Department of Psycholinguistics
Saarland University, Germany
{aweber,crocker}@coli.uni-sb.de

Abstract

Narrow focus refers to accent placement that forces one interpretation of a sentence, which is then often perceived contrastively. Narrow focus is formalised in terms of alternative sets, i.e. contextually or situationally salient alternatives. In this paper, we investigate whether this model is valid also in human utterance processing. We present an eye-tracking experiment to study listeners' expectations (i.e. eye-movements) with respect to upcoming referents. Some of the objects contrast in colour with objects that were previously referred to, others do not; the objects are referred to with either a narrow focus on the colour adjective or with broad focus on the noun. Results show that narrow focus on the adjective increases early fixations to contrastive referents. Narrow focus hence activates alternative referents in human utterance processing.

1. Introduction

One of the principal functions of accent is the focusing on salient information in utterances [1]. Whereas in the syntactic literature various rules have been put forward on how to determine the most natural position of the element the accent falls on [2, 3], in the semantic literature proposals have been made on how to formalise focus (e.g., [4, 5, 6, 7, 8]). Most of these proposals are based on the assumption that focus activates alternative referents of some sort. Accenting the adjective of a noun phrase is generally perceived as contrastive. In the sentence 'He was warned to look out for an ex-convict with a RED shirt' (example from [4]), the accent on the adjective contrasts the ex-convict with the red shirt from an ex-convict with, for example, a blue or green shirt. When the accent is associated with the noun rather than with the adjective (He was warned to look out for an ex-convict with a red SHIRT), a variety of focus interpretations are possible [9], none of which are contrastive[1].

We used the eye-tracking paradigm to investigate the effect of narrow focus on reference interpretation in German. In eye-tracking studies, fixations to displayed objects are monitored as the utterance unfolds. This reveals precisely which object in the display is being understood as the intended referent and how long it takes listeners to launch an eye movement [13]. It has been shown that referents in a scene are identified as soon as they are referred to in an utterance; also, referents can be identified even before they are mentioned (e.g., [14, 15]). The paradigm is therefore highly suited to investigate incremental referential processing.

Results of previous eye-tracking experiments concerned with reference resolution in different prosodic contexts are inconclusive, however. There is some evidence that when contextual presuppositions were met, contrastive accent on a modifier speeded referent identification [14]. In a very similar study to ours, however, Sedivy et al.,did not find an effect of focus on eye-movement latencies to target referents in English [16]. The experimental setup used in the present experiment is very similar to [16] but rather than analysing eye-movement data after noun onset, we analyse fixations from adjective onset onwards.

Listeners in our study were asked in two consecutive instructions to click on one object among four on a computer display while their eye movements were monitored. The first instruction always referred to one member of a visually displayed contrast pair (e.g., purple scissors in the presence of both purple and red scissors); the second instruction referred to either the other member of the contrast pair (e.g., red scissors) or to an object differing in form but not colour from the other member of the contrast pair (e.g., red vase). A contrast pair is defined here as containing two identical objects contrasting in colour. Based on the content of the second instruction, the point at which a unique referent could be selected was during the noun, thus after the colour adjective was heard. To investigate the processing of narrow focus, in half of our trials the second instruction carried a contrastive accent on the adjective. Since only the red scissors contrasted in colour with an other displayed object, the accent on the adjective was an appropriate cue for the red scissors as upcoming referent. Thus, if narrow focus on the adjective triggered alternatives, we should find earlier looks to the red scissors when the adjective is accented than when it is not.

2. Experiment

2.1. Participants

Twenty-four native speakers of German were paid to take part in the experiment. They had normal or corrected-to-normal vision and normal hearing.

2.2. Materials

Sixteen German nouns referring to illustratable objects were chosen as stimuli. Nouns were modified with a colour adjective in the first instruction (first referent, e.g., lila Schere, 'purple scissors'). For each first referent two second referents were chosen: one referred to the same object but was modified with a different colour adjective (contrRef, e.g., rote Schere, 'red scissors'), the other referred to a different object that matched in colour with the contrastive referent (non-contrRef, e.g., rote

[1] Some researchers claim that a contrastive interpretation of the noun (i.e., the ex-convict with a red shirt not with a red hat) can be forced by a particular accent type. Pierrehumbert and Hirschberg, for instance, claimed that the accent L+H* is a contrastive version of the more neutral H* [10]. It is under dispute, however, whether L+H* and H* do indeed represent different accent types. Ladd and Schepman (2003) showed, for instance, that H* can be preceded by a low target as well [11]. Watson et al. found that the interpretations of referents with H* and L+H* accent were strongly overlapping [12].

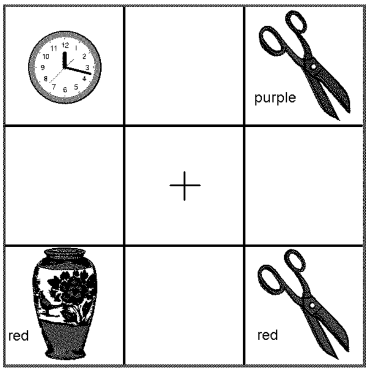

Figure 1: *Black and white example display presented to participants.*

Vase, 'red vase'). A fourth noun referring to an object that shared neither form nor colour with the other objects (distractor, e.g., Uhr, 'clock') was added. The pictures of a first referent, its contrastive and non-contrastive second referents, and distractor were displayed together on a computer screen, see Figure 1. Pictures were selected from a commercially available collection of coloured line drawings [17] and further processed using Adobe Illustrator.

Each trial consisted of two consecutive instructions to click on an object in the display. The first instruction named the first member of a contrast pair (e.g., Klicke die lila Schere an, 'click on the purple scissors'); the second instruction either referred to the contrastive referent (e.g., Klicke jetzt die rote Schere an, 'click now on the red scissors') or to the non-contrastive referent (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., scissors, vase, and clock are feminine in German); the metrical pattern in a trial was matched for colour adjectives and nouns.

Spoken instructions were recorded onto DAT in a sound-attenuated room by a phonetically trained female native speaker of German, sampling at 48 kHz. The recordings were then down-sampled to 20.48 kHz and stored on disc. The first instruction of experimental trials was always recorded with a high pitch accent on the adjective (i.e., contrastive accent)[2]. The second instruction, referring to either the contrastive or the non-contrastive referent was recorded twice, once with a high pitch accent on the noun (i.e., non-contrastive accent, broad focus) and once with a high pitch accent on the adjective (i.e., contrastive accent, narrow focus).

The two factors referent (with the two levels contrRef and non-contrRef) and prosodic accent (with the two levels contrAccent and non-contrAccent) constituted our four experimental conditions (see Table 1). Using Praat [18], 1900 ms silence was added between first and second instructions so that subjects had time to click on the first-referent object.

To prevent participants from developing expectations that pictures with matching colour or form were likely targets, 22 filler trials were added. Filler trials also consisted of four displayed objects accompanied by two consecutive instructions to click on an object. Prosodic patterns on filler items were varied. Some filler trials had contextually unexpected prosodic patterns (i.e., contrastive accent for a non-contrastive referent).

Four lists were constructed, each containing 16 experimental and 22 filler trials in pseudo-random 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.

[2]This is appropriate given that there are two identical referents in the display that only differ in colour.

2.3. Procedure

Participants' eye movements were monitored using a SMI Eye-Link Hispeed 2D eye-tracking system with a camera on a headband. The centre of the pupil was tracked to determine the position of the eye relative to the head. Onset and offset times and spatial coordinates of participants' fixations were stored. The sampling rate of the eye tracker was 250 Hz. Only the dominant eye of a participant was monitored. Along with the eye movements, time and location of the mouse click were stored.

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 four trial lists. All pictures were presented in the corner cells of a 3×3 grid with a cross in the middle cell, see Figure 1. Each cell measured 7.5×7.5 cm. The positions of first and second referent objects were randomised across trials. Spoken instructions started 800 ms after the appearance of the pictures on the screen. For each display, participants heard two instructions to click on objects using a computer mouse. Participants were told to look at the centre 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. Auditory stimuli were presented binaurally over headphones at a comfortable loudness 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, non-contrastive referent, distractor, or the background. Saccade times and blinks were not added to fixation times.

2.4. Results

Only fixations during the second instruction were analysed. Twenty-five trials were discarded 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). Fixation proportions were averaged over participants (F1) and items (F2) at successive 10 ms time frames for separate analyses. ANOVAs were conducted with the two factors referent and prosodic accent as within-participants, within-items factors. Figure 2 presents the averaged fixation proportions from adjective onset for trials with contrastive referents (e.g., rote Schere, 'red scissors'; Figures 2a and 2b) and non-contrastive referents (e.g., rote Vase, 'red vase'; Figures 2c and 2d). Fixation proportions for first referents and distractors were averaged. Since it takes typically about 150 to 200 ms before a programmed eye movement is launched [19], observed fixations are triggered by acoustic information presented about 200 ms earlier.

For contrastive referents, Figures 2a and b show that regardless of type of accent, fixation proportions to contrastive target referents (e.g., rote Schere, 'red scissors') started to increase prior to noun onset; already 200 to 300 ms after adjective onset, increasingly more looks went to the anticipated contrastive referent. The probability of fixating the contrastive

	Condition	First instruction	Second instruction
contrRef/non-contrAccent	a	Klicke die LILA Schere an. ‘Click on the PURPLE scissors.’	Klicke jetzt die rote SCHERE an. ‘Click now on the red SCISSORS.’
contrRef/contrAccent	b	Klicke die LILA Schere an. ‘Click on the PURPLE scissors.’	Klicke jetzt die ROTE Schere an. ‘Click now on the RED scissors.’
non-contrRef/non-contrAccent	c	Klicke die LILA Schere an. ‘Click on the PURPLE scissors.’	Klicke jetzt die rote VASE an. ‘Click now on the red VASE.’
non-contrRef/contrAccent	d	Klicke die LILA Schere an. ‘Click on the PURPLE scissors.’	Klicke jetzt die ROTE Vase an. ‘Click now on the RED vase.’

Table 1: *Example set of instructions.*

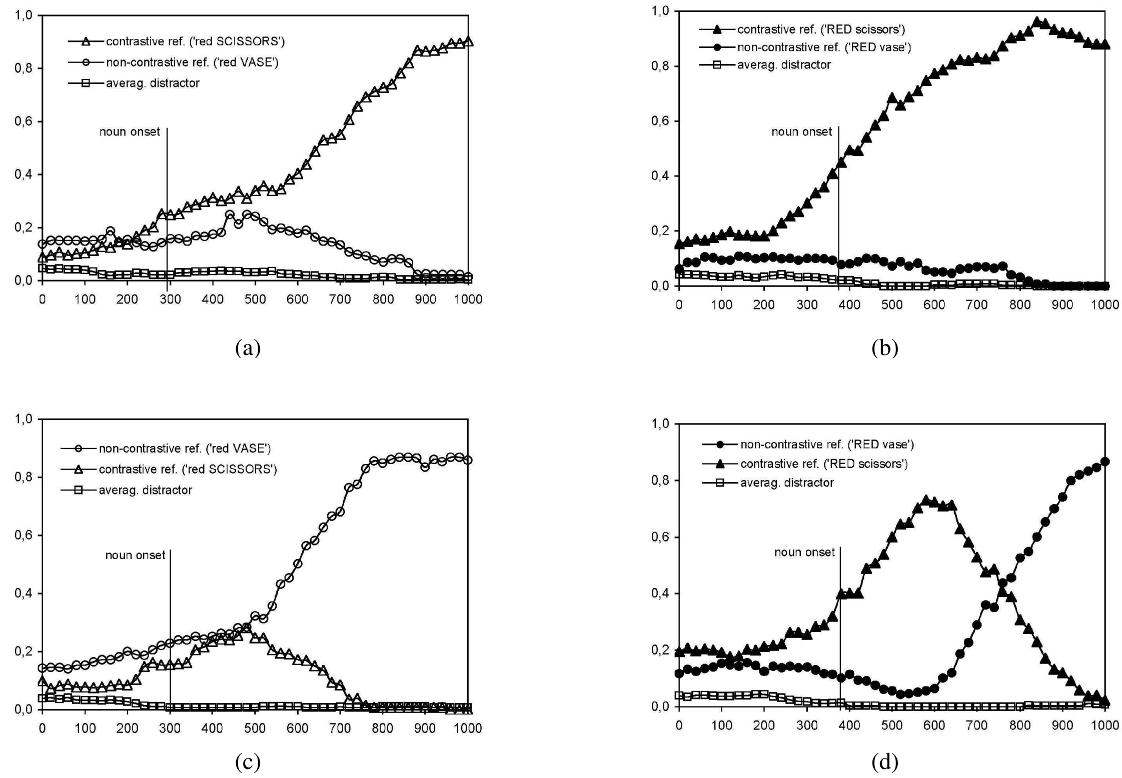

Figure 2: *Average fixation proportions over time starting from adjective onset for (a) contrRef/non-contrAccent trials, (b) contrRef/contrAccent trials, (c) non-contrRef/non-contrAccent trials, and (d) non-contrRef/contrAccent trials.*

referent diverged from that of the non-contrastive referent at different times for contrastive and non-contrastive accent. For contrastive accent trials (Fig. 2b), the divergence started around 250 ms after adjective onset. For trials with a non-contrastive accent (Fig. 2a), both referent types seem to receive a similar amount of looks until 500 ms after adjective onset. ANOVAs, however, revealed a marginal significant difference for 2a between 300 and 550 ms between fixation proportions for contrastive and non-contrastive referents ($F1[1, 23] = 4.28$, $p = .05$; $F2[1, 31] = 5.34$, $p < .05$). Nevertheless, 500 ms after adjective onset, the difference of fixation proportions to contrastive and non-contrastive referents was much smaller in non-contrastive accent trials (Fig. 2a) than in contrastive accent trials (Fig. 2b), suggesting that non-contrastive accents reduced the expectation of an upcoming contrastive referent.

Figures 2c and d show fixation proportions in trials with non-contrastive target referents. When instructions carried a non-contrastive accent (Figure 2c), looks to non-contrastive referents increased along with contrastive referents until 500 ms after adjective onset. Between 300 and 500 ms, there was no significant difference between fixations to contrastive and non-contrastive referents ($F1$ & $F2 < 1$). When second instructions carried a contrastive accent (Figure 2d), on the other hand, listeners started fixating the contrastive referent as soon as they encountered the adjective. Thus, 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 was available. The probability of fixating the contrastive referent was greater than that of the non-contrastive target referent until approximately 800 ms after adjective onset. Between 300 and 800 ms, the proportion of fixations was 48% for the contrastive referent and 17% for the non-contrastive referent. A two-factor ANOVA showed that this difference was significant ($F1[1, 23] = 22.39$, $p < .001$; $F2[1, 31] = 39.95$, $p < .001$).

Finally, we compare the increase in *target* fixations across the four conditions. There was a significant main effect of referent between 300 and 1000 ms ($F1[1, 23] = 42.36$, $p < .001$; $F2[1, 31] = 30.12$, $p < .001$), such that fixations proportions for contrastive referents increased faster than for non-contrastive referents. Also, we found a significant interaction between referent and accent ($F1[1, 23] = 105.64$, $p < .001$; $F2[1, 31] = 35.76$, $p < .001$). Post-hoc pairwise comparisons furthermore showed a significant effect of accent for both contrastive referents ($F1[1, 23] = 18.72$, $p < .001$; $F2[1, 31] = 17.96$, $p < .001$) and non-contrastive referents ($F1[1, 23] = 40.34$, $p < .001$; $F2[1, 31] = 30.26$, $p < .001$). Contrastive accent hence speeded up the recognition of contrastive target referents. When, however, the spoken instruction carried a contrastive accent and referred to the non-contrastive referent, recognition of that target referent was slowed down. Durational differences between accented and unaccented adjectives could not solely be responsible for the pattern of results: ANCOVAs with adjective durations as covariates, still showed a significant influence of referent ($F2[1, 30] = 38.54$, $p < .001$) and an interaction between referent and accent ($F2[1, 30] = 43.19$, $p < .001$). The interaction between referent and accent deviates from the results of [16] who had found no effect of accent and no interaction.

3. Discussion

The present experiment investigated the time course of referent interpretation using the eye-tracking paradigm. Participants were asked in two subsequent instructions to click on an object on the screen. The first instruction was presented with a narrow focus on the adjective, which was appropriate given that the display contained two identical items that only differed in colour. Analysing fixation proportions after adjective onset of the second instructions showed a significant effect of referent (contrastive and non-contrastive object) and an interaction between referent and accent. These results deviate from those of Sedivy et al. 1999 who conducted a similar experiment for English [16]. They had only found an effect of referent on eye-movement latencies and concluded that referents with adjectival modification are generally interpreted contrastively, independent of prosodic marking. The differences are probably due to the dependent variables used. Their analysis was based on eye-movement latencies after noun onset. We analysed the time course of fixation probabilities and found differences in eye-movements already before noun onset.

We found that the preference to interpret adjectives contrastively (as concluded by [16]) was neutralised when adjectives were unaccented. We suggest that the visual context and the first instruction in narrow focus set expectations for a contrastively accented contrastive referent in the second instruction. When adjectives were unaccented in the second instruction, contrastive referents were no longer preferred; contrastive and non-contrastive referents received an equal amount of looks.

When the colour adjectives were accented (narrow focus) listeners looked significantly more often to the contrastive referents than to the non-contrastive ones. Narrow focus hence clearly activates alternative referents. Listeners' increased fixations to contrasting objects in the narrow focus conditions is coherent with the way narrow focus is formalised in semantics.

4. References

[1] Bolinger, D., Intonation across languages. In. J.H. Greenberg (ed.), 1987. *Universals of Human Language, Vol. 2, Phonology*. Stanford University Press.

[2] Selkirk, E., *Phonology and Syntax: The Relation between Sound and Structure*. MIT Press, Cambridge, Massachusetts. 1984.

[3] Gussenhoven, C., *On the grammar and Semantics of Sentence Accents*. Foris Publications, Dordrecht. 1984.

[4] Jackendoff, R., *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge, Massachusetts. 1972.

[5] Rooth, M., *Association with Focus*. Ph.D. thesis, University of Massachusetts, Amherst. 1985.

[6] Büring, D., *The Meaning of Topic and Focus – the 59th Street Bridge Accent*. Routhledge, London. 1997.

[7] Steedman, M., Information structure and the syntax-phonology interface. *Linguistic Inquiry* 31(4):649–689. 2000.

[8] Krahmer, E. and Swerts, M., On the alleged existence of contrastive accents. *Speech Communication* 34:391–405. 2001.

[9] Ladd, D.R., *The Structure of Intonational Meaning*. Indiana University Press, Bloomington. 1980.

[10] Pierrehumbert, J. B. and Hirschberg, J., The meaning of intonation in the interpretation of discourse. In Cohen, P., Morgan, J., and Pollack, M. (eds.), 1990. *Intentions in Communication*. MIT Press, Cambridge, Massachusetts. chap. 14, pp 271–311.

[11] Ladd, D. R. and Schepman, A., "Sagging transitions" between high pitch accents in English; experimental evidence. *Journal of Phonetics* 31:81–112. 2003.

[12] Watson, D., Tanenhaus, M.K., and Gunlogson, C.A., Interpreting pitch accents in on-line comprehension. ms.

[13] Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M, and Sedivy, J.C., Integration of visual and linguistic information in spoken language comprehension. *Science* 268:632–634. 1995.

[14] Sedivy, J.C., Tanenhaus, M.K., Eberhard, K., Spivey-Knowlton, M., and Carlson, G.N. Using intonationally-marked presuppositional information in on-line language processing: evidence from eye movements to a visual model. Proc. of 17th Annual Conference of the Cognitive Science Society, pp 375–380. 1995.

[15] Kamide, Y., Scheepers, C., and Altman, G., Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research* 32:37–55. 2003.

[16] Sedivy, J.C, Tanenhaus, M.K, Chambers, C. and Carlson, G.N. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71:109–147. 1999.

[17] IMSI Masterclips. Premium image collection 303,000 [Image Database] (http://www.imsisoft.com). 1990.

[18] Boersma, P. and Weenink, D., Praat, a system for doing phonetics by computer, version 3.4. Institute of Phonetic Sciences of the University of Amsterdam, Report 132. 1996.

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