Now for something completely different: Anticipatory effects of intonation

Bettina Braun[a,b] and Aoju Chen[a]
[a]Max Planck Institute for Psycholinguistics, the Netherlands
[b]University of Konstanz, Germany

1 INTRODUCTION

It is nowadays well established that spoken sentence processing is achieved in an incremental manner. As a sentence unfolds over time, listeners rapidly process incoming information to eliminate local ambiguity and make predictions on the most plausible interpretation of the sentence. Previous research has shown that these predictions are based on all kinds of linguistic information, explicitly or implicitly in combination with world knowledge.[1] A substantial amount of evidence comes from studies on online referential processing conducted in the visual-world paradigm (Cooper 1974; Eberhard, Spivey-Knowlton, Sedivy, and Tanenhaus 1995; Tanenhaus, Sedivy-Knowlton, Eberhard, and Sedivy 1995; Sedivy, Tanenhaus, Chambers, Carlson 1999).[2] In this paradigm, listeners are shown a visual scene containing a number of objects and listen to a short sentence about the scene. They are asked to either inspect the visual scene while listening or to carry out the action depicted in the sentence (e.g. 'Touch the blue square'). Participants' eye movements directed to each object in the scene are monitored and time-locked to pre-defined time points in the auditory stimulus. Anticipatory effects and their triggers in the auditory signal can be examined by analyzing fixations to a given referent *before* acoustic information on the referent is available.[3]

Various studies have demonstrated that within an elaborate referring expression (e.g. adjective(s) + noun), listeners use information from each word preceding the noun to reduce the set of possible referents to the intended one (Eberhard et al 1995; Sedivy et al. 1999). For example, Eberhard et al. (1995,

[1] In the literature, some authors talk about anticipatory effects or prediction while others claim incremental effects. In our view, anticipation or prediction can only be achieved by incrementally evaluating all sources of information available before the respective linguistic expression is produced and processed.
[2] Anticipatory effects have also been observed in phoneme processing (see Fowler and Brown 2000 for details) and word recognition (see Salverda, Dahan, and McQueen 2003 for a brief literature review).
[3] Anticipatory effects can be reflected in different measures, e.g. first saccade, first fixation or mean proportion of fixations to a certain object. A fixation is an interval in which the eye rests at a region of interest; a saccade is a fast movement of an eye between two fixations (Salvucci and Goldberg 2000).

Experiment 1) presented listeners with short instructions like 'Touch the starred yellow square' with disambiguating information at three different positions in the referring expression (i.e., adj1, adj2, noun). They found that the first saccade to the target referent was launched earliest when adj1 contained the disambiguating information but latest when only the noun contained the disambiguating information. Furthermore, it has been shown that listeners can predict an upcoming direct object before the onset of the referring expression itself by using verb-based information (i.e., the semantic constraints that the verb imposes on permissible objects). For example, listeners directed the first saccade earlier and launched more first saccades in general towards an edible object upon hearing 'They boy will eat ...' compared to 'The boy will move ...' (Altmann and Kamide 1999). In subsequent studies, Kamide, Altmann and colleagues (Kamide, Scheepers and Altmann 2003; Altman and Kamide 2007) established that listeners also exploit morphosyntactic and syntactic information available before and/or during the verb region (e.g. case-marking, grammatical voice, tense, and aspect) for reference resolution.

More relevant to the present study is the role of *intonation* in predicting the future course of a sentence. Work in the 1970's has shown that listeners process intonational information incrementally as a sentence unfolds itself. Cutler (1976), for instance, found that listeners were significantly faster in detecting a given word-initial target phoneme (e.g., /d/) in a sentence when the intonation of the part preceding the target-bearing word predicted an accent on that word than when this was not the case (e.g., 'She managed to remove the *dirt* from the rug, but not the berry stains' vs. 'She managed to remove the *dirt* from the rug, but not from their clothes'). Since the target-bearing word (e.g., dirt) in both conditions was spliced from a neutral rendition of a third sentence (e.g., 'She managed to remove the dirt from the rug'), listeners must have exploited prosodic information prior to the target word.[4]

More recently, researchers have begun to examine whether and how listeners exploit intonational information to predict upcoming referents by means of the visual-world paradigm and its adapted versions. Unlike adjectival modifiers, verb-related information, morphosyntactic and syntactic cues, intonation can only affect reference resolution indirectly, i.e. via the interface between intonation and information status.[5] For instance, referents new to the discourse are usually accented, while already mentioned referents typically receive no accent. Since speakers tend to also accent already mentioned

[4] Listeners also make predictions based on repetitions of the rhythmic organization of speech (see Dilley and McAuley 2008 and references therein).

[5] Therefore, in studies on anticipatory effects of intonation, participants are usually presented with two utterances so that the referent mentioned in the second utterance can be defined as given or new relative to the preceding utterance. In such a set-up, the new referent is also simultaneously contrastive.

referents (Terken and Hirschberg 1994; Terken and Nooteboom 1987), there is at most a strong association between accent placement and information status, but not an absolute one-to-one form-function mapping (but see Niebuhr 2006 for arguments for a one-to-one relationship between accent placement and information status). Most intonation patterns that are claimed to convey a certain meaning only represent the most frequent pattern that speakers choose to use in that context (e.g., Caspers 2003; Braun 2006; Braun and Chen 2010). Therefore, intonation may not have as strong a predictive role as verb-related semantic constraints or (morpho)syntactic information. However, despite the lack of a one-to-one mapping from intonational form to function, listeners have been shown to use intonational information within a referring expression efficiently as soon as it becomes available to identify upcoming referents.

Dahan, Tanenhaus, and Chambers (2002), for instance, asked listeners to follow two consecutive instructions to move an object in a grid on the computer screen (e.g., 'Put the candle above the square; now put the candle/candy below the circle'). The grid contained pictures of two referents with an overlapping first syllable (e.g., 'candle' and 'candy'), two phonemically unrelated distractor objects, and four geometrical shapes. The direct object in the second instruction (target word) was either accented (signalling new or contrastive information) or unaccented (signalling given or non-contrastive information). Following an initial bias towards the referent unmentioned in the first instruction (i.e. contrastive referent – 'candy'), listeners launched even more fixations to the contrastive referent when it was accented but shifted their fixations to the first referent when it was unaccented. Since the first syllable of the target word (e.g., /kæn/) did not segmentally disambiguate between the two referents in question, this difference could only be caused by the intonational realization of the first syllable. Thus, by exploiting intonational information - in particular the presence or absence of accentuation in the segmentally ambiguous first syllable of the target word - listeners got a head start in resolving the referential ambiguity. Using the same eye tracking paradigm, Chen, den Os, and de Ruiter (2007) found that listeners were also sensitive to the shape of the accent (i.e. accent type). Listeners fixated the contrastive referent more when the target word was produced with a fall (H*L) or delayed fall (L*HL) than when it was spoken with a rise (L*H) or no accent.

Intonational information available before the referent noun but still within the referential expression has also been shown to be effective in guiding listeners' expectations. Using the same method as Dahan et al. (2002), Weber, Braun, and Crocker (2006) presented German listeners with instructions such as 'Klicke die lila Vase an. Klicke jetzt die rote Vase an' ('Click on the purple scissors. Now click on the red vase'). The colour adjective in the referring expression of the second instruction ('rote') was either accented with L+H* (an accent with a high tonal target preceded by a steep rise from a rather low pitch value) or left unaccented (e.g., 'ROTE Vase' vs. 'rote VASE'). Listeners

launched more fixations towards the same *type* of object as in the first instruction (e.g., vase) but with the different colour when they heard an accented colour adjective than when they heard an unaccented one. This anticipatory effect was even stronger when the colour adjective of the first instruction was already accented, suggesting that listeners already anticipated a contrastive accent on the adjective in the second instruction. Ito and Speer (2008, Experiments 1 and 2) conducted a similar study in English with real world objects. Participants' task was to pick up objects from different cells of a grid to decorate a real holiday tree, following instructions such as 'Hang the blue angel. And next, hang the GREEN ball/angel'. The adjective in the second instruction was either spoken with a L+H* accent (which signals contrast, cf. Pierrehumbert and Hirschberg 1990) or with an H* accent (high tone, which has no contrastive connotation). Results were similar to those of Weber et al.: listeners fixated the cell with the same type of objects as in the first instruction (e.g., angels) more often and earlier when the adjective was realized with an L+H* accent compared to when it was realized with an H* accent. Eberhard et al.'s (1995) Experiment 3 showed a similar effect for the first adjective in referential expressions with more than one adjectival modifier.

The remaining question is then whether listeners also make use of intonational information *prior to the entire referential expression* to identify the upcoming referent, in analogy to verb-based information. A test case for this is the intonational realization of phrase-initial discourse markers, which are generally used to increase discourse coherence and to mark relations between utterances and events. Therefore, they are also frequently used in the experimental materials of the above-mentioned eye tracking studies, e.g. 'now' in Dahan et al. (2002) and Chen et al. (2007), 'jetzt' in Weber et al. (2006), 'and next', 'and then', 'after that', and 'finally' in Ito and Speer (2008). In a recent production experiment on British English and Dutch, Braun and Chen (2010) elicited similar bipartite instructions with a movie-clip description task, and found that the intonational realization of the discourse markers 'now' in English and 'nu' in Dutch were adapted to the information structure of the upcoming sentence. They were mostly unaccented when the following referent was contrastive ('Put the candle in cell 1. Now put the candy in cell 1') but mostly accented with a steep rise (L*H) when the location was contrasted ('Put the candle in cell 1. Now put the candle in cell 9'). The interesting question is then whether listeners can make use of the intonational realization of the phrase-initial discourse markers to predict the upcoming referent.

A case in point is Ito and Speer's Experiment 3 (2008) in which the authors examined the effect of a discourse marker's intonational realization on the interpretation of subsequent referents. They used the same tree decoration task as described above with instructions such as 'And next, pick the blue ball'. The adverbs of the discourse markers 'and next', 'and then', and 'after that' were either realized with a L+H* L-H% or with an H* L-H% patterns (a steep rise

followed by a rising boundary tone vs. a weak rise or a high level tone followed by a rising boundary tone). They examined whether the 'contrastive' accent L+H* evoked more anticipatory looks to a referent differing only in colour from the preceding referent. The intonation patterns of the discourse markers were either matched or mismatched with the information status and intonational realization of the target ornament. Strikingly, in both intonation conditions, participants rarely fixated the target cell before noun onset. Thus, an L+H* on a discourse marker did not trigger the anticipation of an upcoming colour contrast, unlike the L+H* on a colour adjective itself. The authors speculated that L+H* on the discourse markers may have at most provided an attention-orientating cue or signalled upcoming contrasts in general. While these might be possible functions of an L+H*-accented discourse marker, it is perplexing that listeners in Ito and Speer's study hardly fixated the target cell at all before noun onset. This total absence of anticipatory eye movements is especially noteworthy in the light of the existing evidence for anticipatory effects arising from intonation as well as other linguistic information.

We see a number of potential reasons why Ito and Speer (2008) did not observe anticipatory effects based on the intonational realization of the discourse markers. First, the hypothesized effect of L+H* as a marker for contrastive information may not be applicable to the discourse markers examined in their study in the first place, and there is no reference to relevant production data to back up this claim – as there is for the intonation patterns of adjectives and nouns (Ito and Speer 2006). Second, the presence of a phrase boundary after the discourse marker sets the discourse marker intonationally apart from the rest of the sentence. Consequently, listeners might have regarded the discourse markers as pure attention getters and did not interpret the intonational realization of the discourse markers in connection with the information in the following intonational phrase.

In the present study, we focused on anticipatory effects based on the intonational realization of the discourse marker *nu* ('now') in Dutch. In two eye tracking experiments listeners moved objects in a grid following bipartite instructions such as *Verplaats het boek naar vak 1. Verplaats nu de film naar vak 1* ('Put the book in cell 1. Now put the film in cell 1'). We examined listeners' eye movements towards the first referent (i.e., the referent mentioned in the first instruction) and the contrastive referent (i.e., the referent not mentioned in the first instruction) after they processed the intonational information in *nu* but before they could process the target word. The intonation of our auditory materials was determined on the basis of the most frequent realizations in comparable conditions in Braun and Chen's (2010) study.. The discourse marker *nu* was realised with L*+H when the referent was maintained but the location was contrastive, and with no accent when the referent was contrastive but the location was maintained. Further, to ensure that listeners

interpreted the discourse marker as an integral part of the instruction, there was no phrase break following the discourse marker.

On the assumption that listeners use the language production system to make predictions on what is coming next in others' speech during comprehension, as for instance argued by Pickering and Garrod (2007) and other authors cited therein, we predicted that listeners should make use of the form-function mappings between intonation in *nu* and the information status of the upcoming referent present in production to predict the upcoming referent during comprehension. More specifically, following a likely initial bias towards the contrastive referent as found in earlier studies (Dahan et al. 2002; Weber et al. 2006), listeners should launch even more fixations towards the contrastive referent on hearing an unaccented *nu* than on hearing an L*+H-accented *nu*. Conversely, they should launch more fixations towards the first referent on hearing an L*+H-accented *nu* than on hearing an unaccented *nu*.

2 EXPERIMENT 1

The experiments in this paper made use of the visual world eye tracking paradigm with printed words (McQueen and Viebahn 2007; Reinisch, Jesse and McQueen 2010; Salverda and Tanenhaus 2010). One of the referents was referred to in the first instruction (hereafter first referent), while the other one was not mentioned in the first instruction (hereafter contrastive referent).

In the first instruction, participants had to move one of the two referents to a new cell in the display (e.g., *Verplaats nu het woord ball naar vak 1*). The padding *het woord* ('the word') was added since it sounded more appropriate with the use of printed words. The auditory materials intonationally mirrored the patterns most frequently produced to express an object or a location contrast in Braun and Chen (2010). The discourse marker *nu* was produced with a steep rise (L*+H) to signal a contrast in the location, and produced with no accent to signal a contrast in the referent. There was no intonational mismatch condition, as a mismatch between information structure and intonational realization is likely to draw listeners' attention to intonation, which may create experimental artefacts. If the intonation of the discourse marker affects whether listeners expect a contrastive or a given referent, we should find anticipatory effects also when listeners are not explicitly aware of the prosodic manipulation.

2.1 PARTICIPANTS

Twenty-four native speakers of Dutch participated in the experiment for a small fee. Participants were all unaware of the purpose of the experiment and

had not taken part in the experiments reported in Braun and Chen (2010). They were all students of Radboud University Nijmegen, reported to have normal hearing and normal or corrected-to-normal vision.

2.2 MATERIALS

Twenty-four disyllabic Dutch word pairs with lexical stress on the first syllable were selected. The two words in every pair had an identical initial consonant-vowel sequence (e.g., *zegel-zetel,* 'stamp-seat'*, panda-panter,* 'panda-lepard', see Table 1 in the Appendix for the full list) and did not differ in lexical frequency according to the CELEX word form dictionary (Baayen, Piepenbrock and Gulikers 1995): $t(23) = 0.01$, $p > 0.9$. One word of each word pair served as first referent, the other as contrastive referent. The role of first and contrastive referent was counterbalanced across participants and conditions to reduce effects due to particular words.

An additional set of 10 cohort pairs and a set of 24 non-cohort pairs were selected for filler sentences. From the non-cohort pairs, 18 instruction sequences similar to the experimental ones were created, half with a location contrast and half with an object contrast. To keep participants attentive, the remaining 16 word pairs were used in trials with only a single instruction. In these trials, participants were asked to click on a word or to move a word above or below a square or a triangle.

A female native speaker of Dutch who had been known for being able to produce intonation patterns very consistently read the instructions from a recording list. This list contained four sentences for a given experimental word pair and all filler sentences. Recordings were made in a sound-attenuated cabin at the MPI. Utterances were directly recorded onto a PC using Adobe Audition (44.1 kHz, 16 bit). For the trials with two instructions, the first instruction was recorded with a low initial boundary tone (%L) and a high final boundary tone (H%) (Figure 1).

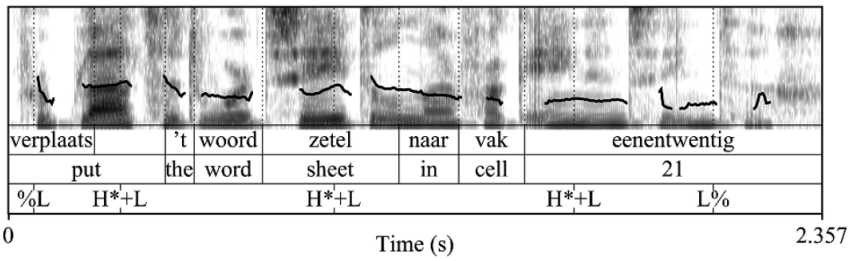

Figure 1: Example pitch track of a first instruction with a high boundary tone in Experiment 1.

H% signals continuation or non-finality (Cruttenden 1997; Pierrehumbert 1981; Pierrehumbert and Hirschberg 1990; 't Hart, Collier and Cohen, 1990), and was hence assumed to better connect the two instructions in each trial than a low boundary tone.

The second instructions always started and ended with a low boundary tone. Those with a contrast in the location (hereafter location contrast) were always recorded with a rising accent on the discourse marker *nu* and a falling accent on the location (Figure 2, upper panel); those with a contrast in the referent (hereafter object contrast) were always recorded with a single falling accent on the object noun (Figure 2, lower panel). The filler trials with only a single instruction were recorded with a low boundary tone both at the start and at the end, and a falling accent on the noun. All intonation contours proved to be very natural; our speaker did not have to be specifically instructed. The recorded utterances were used in the experiment without further manipulation.

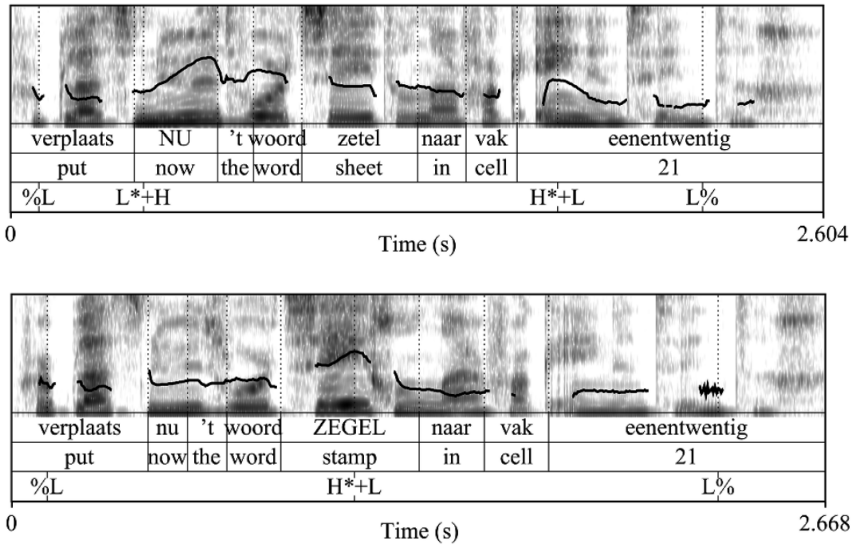

Figure 2: Example pitch tracks of an utterance with an accented *nu* to signal a contrast in the location (upper panel) and an utterance with an unaccented *nu* to signal a contrast in the object (lower panel)

To quantify the prosodic structure up to the target word, we measured the duration and f0-excursion of the constituents preceding the target word (the verb, *nu*, and the padding *het woord*). Table 1 gives an overview of the mean values of these acoustic measures. A series of paired-samples two-tailed t-tests were conducted to assess the significance of the differences in duration and f0-excursion between the object-contrast condition and the location contrast

condition for each constituent. Regarding the verb, neither the average duration ($t(47) = 1.8$, $p > 0.05$) nor the average f0-excursion ($t(47) = 0.92$, $p > 0.3$) differed significantly between the two conditions. Regarding *nu*, both the average duration ($t(47) = 17.2$, $p < 0.0001$) and the average f0-excursion ($t(47) = 27.9$, $p < 0.0001$) differed significantly between the two conditions. Its duration was 63.1 ms longer and its average f0-excursion was 60.8 Hz larger when accented (in the location contrast condition) than when unaccented (in the object contrast condition). Finally, the average duration of the padding *het woord* did not differ between conditions ($t(47) = 0.7$, $p > 0.4$), but the average f0-excursion of the padding was 37 Hz larger in the accented-*nu* condition than in the unaccented-*nu* condition ($t(47) = 11.4$, $p < 0.0001$).

	Duration in ms		F0-excursion in Hz	
	Object contrast	Location contrast	Object contrast	Location contrast
verb	433.6	421.5	36.1	36.1
nu	132.5	195.6	12.1	72.0
padding	371.8	371.8	31.9	68.9

Table 1: Mean duration and f0-excursion of the constituents in Experiment 1.

2.3 PROCEDURE

The first referent, the contrastive referent, a square and a triangle were displayed on a 5×5 grid on a computer screen (Figure 3). The size of the individual cells was 96 x 96 pixels (which corresponds to a size of 2.54 x 2.54 cm). The words were displayed in boldface black Arial 24 against a white background, centred in each cell. The cell number was shown in light grey Arial 12 pt in the top left corner of each cell.

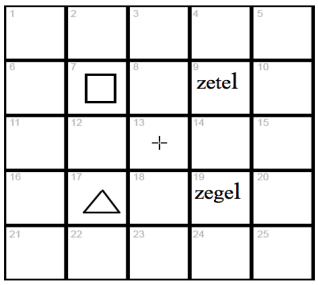

Figure 3: Example display of a trial.

Four basic lists of experimental stimuli were constructed. The role of first and second referent as well as contrast condition was counterbalanced across lists, following a Latin-Square design. More specifically, the word pairs were split into two groups, with a matched average frequency for first and contrastive referents in each group. In one list, the first half of the word pairs was assigned to the object-contrast condition (produced with an unaccented *nu*), the other half to the location-contrast condition (produced with an accented *nu*). In the second list, the order of first and contrastive referent was swapped to minimize a potential bias for one of the two words in each pair (e.g. *zegel* was the first referent in list 1, and *zetel* was the first referent in list 2). In lists 3 and 4, every pair that was assigned to the object-contrast condition in list 1 and 2 was assigned to the location-contrast condition, and vice versa (e.g. the pair *zegel-zetel* appeared in an object-contrast condition in list 1 and in a location-contrast condition in list 3). Twelve of the 34 filler items were used as familiarization trials. The remaining filler trials were interleaved with the experimental trials. There were three randomizations for each of the four lists, resulting in 12 experimental lists.

Participants were randomly assigned to the experimental lists and were tested individually in a sound-attenuated cabin. They were first given written instructions on the task, and were then seated in front of a computer screen at a comfortable distance. An SMI Eyelink II eye tracking system was fitted and calibrated. At the start of each trial, the two words and the two geometric shapes were displayed in cells 7, 9, 17, and 19 of the grid (see Figure 3). Their positions were counterbalanced across conditions so that each of the words and shapes occurred equally often in each of these four cells for each condition. Auditory stimuli were presented binaurally over headphones. The first instructions started simultaneously with the display of the grid. The second instructions started after participants had dropped the word mentioned in the first instruction into its new cell (but not before the end of the first instruction). An automatic drift correction was initiated after each block of six trials.

Participants' eye movements and mouse actions were monitored during the second instructions. The centre of the pupil was tracked to determine the position of the eye relative to the head. Onset and offset as well as the coordinates of the fixations were recorded with a sampling rate of 250 Hz.

2.4 CODING PROCEDURE

The data from each participant's left eye were coded in terms of fixations, saccades, and blinks, using the algorithm provided in the Eyelink software. Blinks and saccades were discarded. For every 4ms-frame recorded by Eyelink, fixations were coded as pertaining to the cell of first referent,

contrastive referent, square, or triangle. This information was subsequently used to calculate the proportion of fixations to each referent.

2.5 RESULTS

For visualization of changes in participants' eye movements over time, the proportion of fixations to the first and contrastive referent (and the two shapes) was calculated in 20ms-intervals by dividing the total number of fixations to a respective referent or shape by the total number of fixations in a given time interval (excluding blinks or saccades). Fixation proportions to first and contrastive referents for each condition are displayed in Figure 4, time-locked to the acoustic onset of the discourse marker *nu,* starting ca. 70 ms after the start of the second instruction (350 ms before *nu* onset) and ending 500-600 ms after the onset of the target noun (1100 ms after *nu* onset).

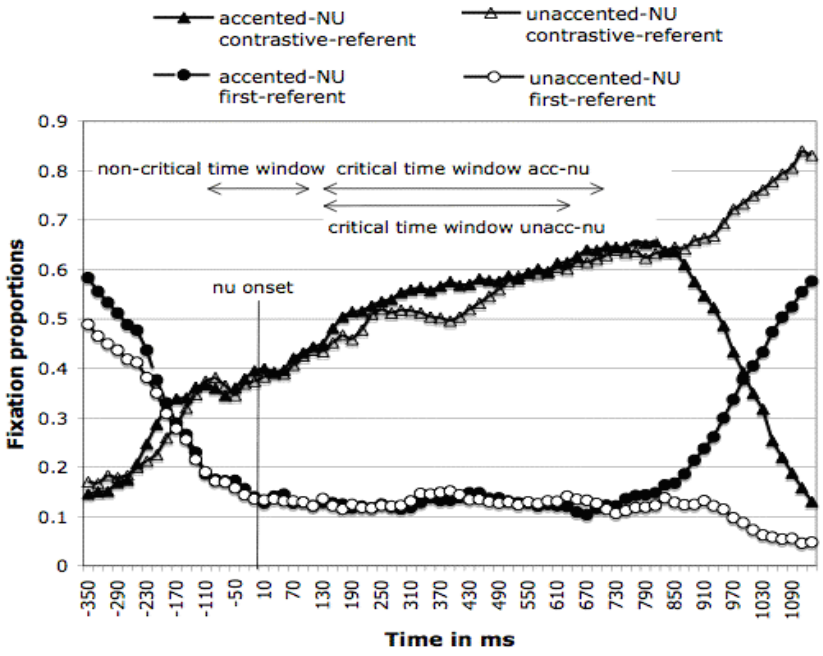

Figure 4: Averaged fixation proportions in Experiment 1 to first (circles) and contrastive referents (triangles) starting from 350 ms before *nu* onset till 1100 ms after *nu* onset. Acoustic onset of *nu* as well as the ranges of the non-critical and critical time windows are marked.

The critical time window to observe anticipatory eye movements started when prosodic information of the discourse marker began to be reflected in participants' eye movements and ended when information of the target referent is being processed. A usual estimate of the time taken to launch an eye movement in such a visual searching task is 150-200ms (Fischer, 1992; Hallet, 1986; Matin, Shao, & Boff, 1993; Rayner, Slowiaczek, Clifton, & Bertera, 1983; Saslow, 1967). Since our displays contained only two printed words, and since participants had sufficient time to preview it (during the first instruction), we used the lower bound of this range (150 ms). The critical time window thus started 150 ms after the onset of the discourse marker *nu* and ended 150ms after the onset of the target word, shown by the long horizontal arrows in Figure 4. The absolute begin and end of the critical time window were determined for each trial individually. The average duration of the critical time window across all trials was 535.8 ms.

Not surprisingly, participants' gaze initially rested on the object just moved (first referent) before it shifted to the contrastive referent. We performed three kinds of analyses. The first two tested for participants' initial preference for any of the referents prior to the point at which fixation patterns could be influenced by the intonational realization of the discourse marker *nu*. These two analyses included fixations in the time window from 150ms before *nu* onset till 150 ms after *nu* onset (hereafter 'non-critical time window', shown by the short horizontal arrow in Figure 4). The third and main analysis tested for anticipatory effects of intonation in the critical time window, in which information on the intonation of the discourse marker was processed but information on the realization of the target was not yet available. This only included fixations that started 150ms after *nu* onset (or later) but excluded all fixations that started 150ms after target onset.

Following the analysis protocol described above, we first tested for an initial bias towards the contrastive referent during the non-critical time window, as previous experiments have shown such a bias (e.g., Dahan, Tanenhaus and Chambers 2002) and the fixation proportions during the non-critical time window in Figure 4 suggest exactly such a bias. More specifically, we compared the ratios of fixation proportions to the contrastive and first referents to 0.5 using a one-sample t-test in separate by-subject and by-item analyses (cf., Dahan and Tanenhaus 2005; Huettig and McQueen 2007). The ratios of fixation proportions were calculated by dividing the fixation proportions to the contrastive referent by the sum of fixation proportions to the contrastive and first referents. Statistical analyses confirmed the visual impression of a bias towards the contrastive referent in the non-critical time window. The average ratio of fixation proportions to the contrastive referent

was 67%, which was significantly higher than chance ($t1(23) = 3.57$, $p = 0.005$; $t2(23) = 6.06$, $p < 0.001$). [6]

We then tested whether there was already an effect of intonation of *nu* on the fixations to the contrastive and first referents during the non-critical time window. Effects of intonation might come from the boundary tone of the first instruction or the intonation of the verb preceding *nu* (e.g., Cutler 1976; Xu and Xu 2005). The fixation proportions to the respective referents in each intonation condition were averaged (by subjects and by items) and subjected to a paired-samples t-test with intonation of *nu* as the independent variable. There was no effect of intonation on the fixation proportions to the first or contrastive referent in the non-critical time window (average fixation proportions to the contrastive referent in accented-*nu* condition: 39.3% compared to 38.8% in unaccented-*nu* condition, $t1 = t2 < 1$; average fixation proportion to the first referent in accented-*nu* condition: 14.6% compared to 14.2% in unaccented-*nu* condition, $t1 = t2 < 1$).

More important is the question as to whether the intonation of *nu* influenced the fixation proportions to the contrastive and first referent in the critical time window (from 150ms after *nu* onset till 150 ms after noun onset). Results of repeated measures ANOVAs with intonation of *nu* as the independent variable showed no effect of intonation on the fixation proportions to the contrastive or first referent in the critical time window (averaged fixation proportions to contrastive referent in accented-*nu* condition: 57.3% compared to 53.3% in the unaccented-*nu* condition, $F1(1,23) = 3.18$, $p = 0.09$, $MSE = 0.56$; $F2(1,23) = 2.62$, $p > 0.1$, $MSE = 0.68$; averaged fixation proportion to the first referent in the accented-*nu* condition: 12.6% compared to 13.1% in the unaccented-*nu* condition, $F1 = F2 < 1$).

2.6 DISCUSSION

Like in earlier eye tracking studies using bipartite instructions (Dahan et al., 2002 for English; Weber, Braun and Crocker, 2006 for German), participants quickly shifted their gaze away from the object just moved and developed a bias towards the novel, yet unmentioned referent. This bias was present already before the onset of the discourse marker *nu*. The initial bias towards the contrastive referent sustained throughout the padding. Listeners only started to fixate the intended referent when information about the referent was processed. One possible source for this bias towards the contrastive referent is the high final boundary tone used in the first instruction. As mentioned before, a high boundary tone signals non-finality and continuation in general. It was chosen here for the end of the first instruction as it enhanced the connection

[6] Additional analyses showed that this initial bias towards contrastive referents held for both intonation conditions.

between the two instructions. For listeners, however, the high boundary tone in the first instruction seemed to be an unambiguous signal for a second instruction involving a *different* referent. Apparently, they expected to be instructed to move *both* of the objects in a given display. Since they already moved one of the two objects after the first instruction, the high-boundary tone might have triggered the guess that the second action would be about the other object. This guess sustained even though in half of the trials with two instructions, their task was to move the same referent again.[7]

Since participants seemed to be convinced that this second action was going to involve the yet unmentioned referent, the discourse marker *nu* (and its intonational realization) may have lost its informativeness. Turning this argument around, a low boundary tone in the first instruction might actually increase the informativeness of the discourse marker *nu*. A low pitch at the end of an utterance signals that the speaker is finished with his turn (or topic) and that there is nothing more to come. The use of a low boundary tone at the end of the first instruction might hence suppress the guess that a new object was to be moved in the subsequent instruction. In other words, participants might judge both referents as equally likely candidates, which would in turn make the intonational realization of *nu* informationally more relevant for the listeners to anticipate the object to be moved. We therefore conduced a second experiment with a low boundary tone at the end of the first instruction to investigate the effect of the intonational realisation of *nu* on reference resolution.

3 EXPERIMENT 2

3.1 PARTICIPANTS

Another group of twenty-four native Dutch participants from the MPI subject pool took part and were paid a small fee. They had neither participated in the experiments in Braun and Chen (2010) nor in Experiment 1.

3.2 MATERIALS

The second instructions were identical to the ones used in Experiment 1. The first instructions were produced with a low final boundary tone (see Figure 5).

[7] The results were identical when analyzing the second half of the experiment separately, suggesting that participants did not change this strategy as they encountered more trials with a contrast in the location only.

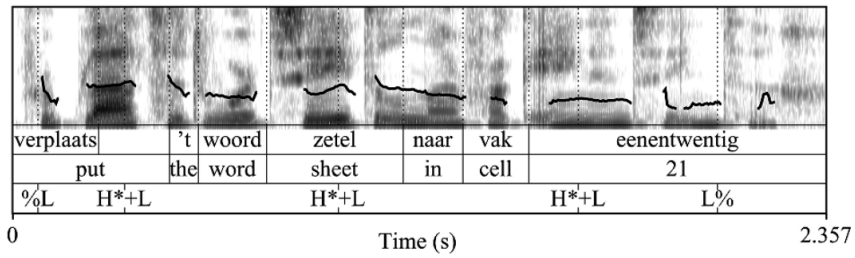

Figure 5: Example pitch track of a first instruction with a low final boundary tone

3.3 PROCEDURE

The testing and coding procedure was identical to that of Experiment 1.

3.4 RESULTS

The time course of fixation proportions towards the first and contrastive referents is plotted in Figure 6. We found no bias towards the contrastive referent in the non-critical time window (mean ratio: 56%, $t1(23) = 1.0$, $p > 0.3$; $t2(23) = 0.7$, $p > 0.4$).

Second, the paired-samples t-test with intonation of *nu* as independent variable and fixation proportions to the first referent and the contrastive referent in the non-critical time window showed that the intonation of *nu* did not affect fixation proportions to the first or contrastive referent in the non-critical time window (average fixation proportions to first referent 29%, $t1 = t2 < 1$; average fixation proportions to contrastive referent was 32%, $t1(32) = 1.6$, $p > 1$, $t2 = 1.4$, $p > 1$).

Third, the paired-samples t-tests with intonation of *nu* as the independent variable and fixation proportions to the first referent and the contrastive referent in the critical time window revealed a strong effect of the intonation of *nu* on the fixation proportions to the contrastive referent (average fixation proportion to contrastive referent in the accented-*nu* condition: 54.9% compared to 40.5% in unaccented-*nu* condition ($t1(23) = 3.5$, $p < 0.005$, $t2(23) = 2.9$, $p < 0.01$). No effect of intonation on fixation proportions to the first referent was found (on average 19.3%, $t1 = t2 < 1$).

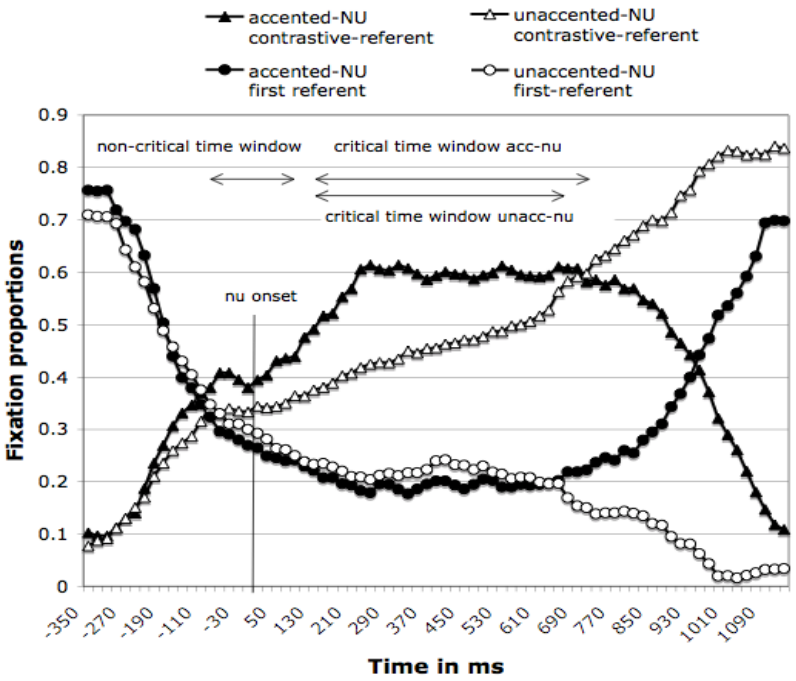

Figure 6: Averaged fixation proportions in Experiment 2 to first (circles) and contrastive referents (triangles) in the two intonation conditions starting from 350 ms before *nu* onset till 1100 ms after *nu* onset. Acoustic onset of *nu* as well as the ranges of the non-critical and critical time windows are marked.

Additionally, Figure 6 suggests that the fixation proportions to the contrastive referent kept increasing till the noun onset when *nu* was unaccented, while there was no such a sustained increase in fixations when *nu* was accented. To statistically verify this observation, we split the critical time window into two equally long time windows (calculated for each item separately) and conducted a repeated measures 2x2 ANOVA, with fixation proportions as the dependent variable, and time window (first half vs. second half of critical time window) and intonation (accented vs. unaccented *nu*) as the independent variables. Results showed a significant main effect of intonation ($F1(1,23) = 5.55$, $p < 0.05$, $MSE = 0.25$, $F2(1,23) = 5.34$, $p < 0.05$, $MSE = 0.98$) and a significant interaction between time window and intonation ($F1(1,23) = 4.22$, $p < 0.05$, $MSE = 1.68$; $F2(1,23) = 5.3$, $p < 0.05$, $MSE = 2.07$). The average fixation proportions to the contrastive referent in the accented-*nu* condition remained relatively stable (59.8% in the first half of the time window compared to 57.2% in the second half of the time window), while

fixation proportions to the contrastive referent increased in the unaccented-*nu* condition (from 41.2% in the first half of the critical time to 55.4% in the second half). There were no main effects and no interactions on the fixation proportions to the first referent.

3.5 DISCUSSION

With a low boundary tone in the first instruction, the initial bias towards the contrastive referent disappeared. Importantly, when the fixation proportions to both referents were approximately equal, the fixation patterns were strongly influenced by the intonational realization of the discourse marker *nu*. Fixation proportions to the contrastive referent increased immediately after an accented *nu* was processed, showing an enhanced anticipation of the contrastive referent. Intriguingly, this pattern of anticipatory eye movements is not what one would expect given the pattern observed in the production experiment reported in Braun and Chen (2010). There, *nu* was mostly produced with a rising accent when the referent remained unchanged but with no accent when the referent changed in the second instruction. If listeners use production patterns to make predictions in comprehension, as suggested by Pickering and Garrod (2007), we should observe more fixations to the contrastive referent when *nu* was unaccented and more fixations to the first referent when *nu* was accented. It would thus seem that the intonational realization of the discourse marker *nu* was processed paralinguistically, at least initially. Following the Effort Code (Gussenhoven 2002, 2004), more articulatory effort of the speaker leads to a wider pitch range; a wider pitch range is in turn interpreted in the light of motivations for using more articulatory effort, such as the need to convey new information. The wide pitch range in an accented *nu* may have caused the listeners to allocate their attention to something new, i.e. the contrastive referent, which was reflected in the immediate increase in fixations to the contrastive referent. The role of accentuation in attention allocation has also been noted in recent ERP (even-related-potentials) studies. For example, Li, Hagoort, and Yang (2008) have found that an accented noun triggered a larger N400 response than an unaccented noun in a sentence in an early time window (about 120-130 ms after the onset of the target noun), independent of whether the noun conveyed new or given information.

Importantly, our data also show that fixations to the contrastive referent gradually increased after an unaccented *nu* was processed while they remained largely constant upon processing an accented *nu*. This change in fixation proportions over time in the unaccented-*nu* condition suggests a gradual increase in the anticipation of the contrastive referent, which renders the pattern more similar to what one would expect given the patterns found in

production. It thus appears that the linguistic interpretation of the intonation of *nu* lags behind a paralinguistic interpretation. That is, the paralinguistic function resulted in a rapid change in fixation, whereas fixations driven by the linguistic interpretation only slowly increased over time. This is comparable to Li et al.'s (2008) finding that the linguistic meaning of accentuation (i.e. newness, contrast) was processed later in the brain (at about 300-310 ms after the onset of the target noun).

4 GENERAL DISCUSSION

We examined whether listeners could make use of the observed association between the intonation of *nu* (unaccented vs. accented with a rise) and the information status of the upcoming referent (contrastive vs. given) in reference resolution *before* sensory information on the referent is available. Based on previous production data (Braun and Chen 2010), we expected higher fixation proportions to contrastive referents upon hearing an unaccented *nu* than upon hearing an accented *nu*.

Interestingly, we observed clear anticipatory effects based on the boundary tone of the first instruction. In Experiment 1, the first instruction ended with a high boundary tone; listeners expected a change in the referent, i.e. they had a strong bias towards the referent not mentioned in the first instruction. A similar bias was also reported in Dahan et al. (2002). We have argued that participants may have understood the task as to drag and drop *both* objects in the display to a new location As a result, they anticipated moving a different printed word in the second instruction instead of moving the same printed word again.

The bias towards the contrastive referent was eliminated when the first instruction ended in a low boundary tone (Experiment 2). As discussed earlier, a low boundary tone is usually associated with finality and hence separates the two instructions more from each other than a high boundary tone, which signals continuation. Intonationally connected instructions may strengthen the interpretation that both objects have to be moved in order to proceed to the next trial/display. Intonationally disconnected instructions, on the other hand, may be interpreted as separate units. Our results show that following an instruction with a low boundary tone, both referents were interpreted as equally likely to be mentioned in the next instruction.

Importantly, when there was no initial bias towards either of the referents (Experiment 2), we saw a clear, immediate anticipatory effect of the intonation of *nu*. An accented version of the discourse marker *nu* rendered the contrastive referent initially the more likely candidate, a pattern that is opposite to the results of the production data. Possibly, listeners did not interpret the accentuation of *nu* in purely information-structural terms but rather in paralinguistic terms. An accented *nu* may be interpreted as more engaging,

more emphatic and thereby signal a change in what is coming up. As in the linear order of the sentence, the referent appears before the location, it makes sense to be prepared for a change in the referent. This is in line with Ito and Speer's (2008: 564) speculation that the rising accent on a discourse marker may be an 'attention-orientating cue' or signal upcoming contrasts in general.

Although an accented *nu* initiated such an attentional shift towards the contrastive referent, the fixations to the contrastive referent did not increase any further during the remainder of the critical time window. On the other hand, when the discourse marker was *not* accented, the fixations to the contrastive referent increased over time, reflecting a gradually increasing anticipation of the contrastive referent. The presence of a gradual increase in the fixations to the contrastive referent across the critical time window in the unaccented-*nu* condition and the absence of such an increase in the accented-*nu* condition were in line with the linguistic functions of the intonation in *nu*.

To conclude, our results show that despite the probabilistic nature of intonation, there are robust anticipatory effects of intonation on reference resolution. Our experiments have provided evidence for two kinds of anticipatory effects of intonation. First, the choice of final boundary tone in the first instruction modulated listeners' initial guess as to which referent to be moved next. A high boundary tone triggered a bias towards the contrastive referent. Second, in the absence of such an initial bias as in the case of a low boundary tone in the first instruction, an accented *nu* initiated an immediate attentional shift towards the contrastive referent. But in the subsequent time frames, this initial attention to the contrastive referent was modified as we predicted on the basis of the linguistic functions of the intonation in *nu* found in Braun and Chen (2010). An unaccented *nu* led to a gradual increase in the fixations to the contrastive referent but an accented *nu* did not. Thus, the paralinguistic intonational meaning is processed before linguistic meaning (here information-structure related meanings) not only in nouns, which can have information status (Li et al. 2008), but also in the discourse maker *nu*, which does not have information status itself but whose intonation varies with the information status of the upcoming referent.

APPENDIX

Word 1	Log Lemma frequency	English translation	Word 2	Log Lemma frequency	English translation
zegel	2.18	stamp	zetel	2.61	seat (in parliment)
kever	0	beetle	ketel	2.46	boiler
toeter	1.72	horn	toekan	0.78	toucan
havik	2.08	hawk	hamer	2.54	hammer
puzzel	2.09	puzzle	pudding	1.98	pudding
poedel	1.69	poodle	poema	1.20	puma
panda	1.96	panda	panter	2.10	panter
hennep	1.67	hennep	hengel	2.12	fishing rod
beker	0	beaker	bever	1.11	beaver
tube	2.00	tube	tuba	0.85	tuba
duivel	3.20	devil	duiker	1.75	diver
sikkel	2.01	sickle	singel	2.03	zone
lepel	2.68	spoon	lever	2.74	liver
toga	1.82	gown	totem	1.41	totem
zwaluw	2.31	swallow	zwavel	2.12	sulphur
motor	3.20	motor	molen	2.55	mill
disco	1.65	disco	distel	1.08	thistle
navel	2.41	navel	nagel	2.14	nail
kabel	2.55	cable	kater	2.68	hangover
schommel	1.88	swing	schoffel	1.50	hoe
drummer	1.56	drummer	druppel	2.62	drop
sofa	2.56	sofa	soda	1.84	soda
visser	2.75	fisherman	vinger	3.30	finger
merel	2.33	blackbird	metro	2.09	metro
average	2.01			1.98	

Table 1: Word pairs with log lexical lemma frequency

REFERENCES

Altman, G.T.M. and Kamide, Y. (2007): The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. Journal of Memory and Language 57, 502–518.

Altman, G.T.M. and Kamide, Y. (1999): Incremental interpretation at verbs: restricting the domain of subsequent reference. Cognition 73, 247-264.

Baayen, H. R., Piepenbrock, R., and Gulikers, L. (1995): The CELEX lexical database [CD-ROM]. : Linguistic Data Consortium. Philadelphia, PA: University of Pennsylvania.

Braun, B. (2006): Phonetics and phonology of thematic contrast in German. Language and Speech 49(4), 451-493.

Braun, B. and Chen, A., (2010): Intonation of 'now' in resolving scope ambiguity in English and Dutch. Journal of Phonetics 38, 431-444.

Caspers, J. (2003): Local speech melody as a limiting factor in the turn-taking system in Dutch. Journal of Phonetics 31, 251-276.

Chen, A., den Os, E., and de Ruiter, J. P. (2007): Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. The Linguistic Review 24 (2-3), 317-344.

Cruttenden, A. (1997): Intonation (2nd edition). Cambridge: CUP.

Cooper, R. M. (1974): The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. Cognitive Psychology 6, 84–107.

Cutler, A. (1976): Phoneme-monitoring reaction time as a function of preceding intonation contour. Perception and Psychophysics 20(1), 55-60.

Dahan, D. and Tanenhaus, M. K. (2005): Looking at the rope when looking for the snake: conceptually mediated eye movements during spoken-word recognition. Psychonomic Bulletin and Review 12, 453–459.

Dahan, D., Tanenhaus, M.K., and Chambers, C.G., (2002): Accent and reference resolution in spoken-language comprehension. Journal of Memory and Language 47, 292-314.

Dilley, L. and McAuley, J. D. (2008): Distal prosodic context affects word segmentation and lexical processing. Journal of Memory and Language 59(3), 291-311.

Eberhard, K.M., Spivey-Knowlton, M.J., Sedivy, J.C. and Tanenhaus, M.K. (1995): Eye movements as a window into real-time spoken language comprehension in natural contexts. Journal of Psycholinguistic Research 24, 409-436.

Fischer, B. (1992): Saccadic reaction time: Implications for reading, dyslexia and visual cognition. In: K. Rayner (ed.): Eye movements and visual cognition: Scene perception and reading, 31–45. New York: Springer.

Fowler, C.A. and Brown, J.M. (2000): Perceptual parsing of acoustic consequences of velum lowering from information for vowels. Perception and Psychophysics 62, 21-32.

Gussenhoven, C. (2002): Intonation and interpretation: Phonetics and phonology. Proceedings of the First International Conference on Speech Prosody, Aix-en-Provence, 47-57.

Gussenhoven, C. (2004): The Phonology of Tone and Intonation, Cambridge: Cambridge University Press.

Hallett, P. E. (1986): Eye movements. In K. Boff, L. Kaufman, and J. Thomas (eds), Handbook of Perception and Human Performance (Vol 1, Chapter 10). New York: Wiley.

Huettig, F., and McQueen, J. M. (2007): The tug of war between phonological, semantic and shape information in language-mediated visual search. Journal of Memory and Language 57(4), 460-482.

Ito, K., and Speer, S. R. (2008): Anticipatory effect of intonation: Eye movements during instructed visual search. Journal of Memory and Language 58, 541-573.

Ito, K., and Speer, S. R. (2006): Immediate effects of intonational prominence in a visual search task. In Proceedings of the 3rd intonational conference on speech prosody, Dresden, Germany.

Kamide, Y., Scheepers, C., and Altmann, G.T.M. (2003): Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. Journal of Psycholinguistic Research 23, 37-55.

Li, X., Hagoort, P., and Yang, Y. (2008): Event-related potential evidence on the influence of accentuation in spoken discourse comprehension in Chinese. Journal of Cognitive Neuroscience 20(5), 906–915.

Matin, E., Shao, K., and Boff, K., (1993): Saccadic overhead: Information processing time with and without saccades. Perceptual Psychophysics, 53, 372-380.

McQueen, J.M. and Viebahn, M., (2007): Tracking recognition of spoken words by tracking looks to printed words. The Quarterly Journal of Experimental Psychology 60(5), 661-671.

Niebuhr, O. (2006): Perzeption und kognitive Verarbeitung der Sprechmelodie - Theoretische Grundlagen und empirische Untersuchungen. Unpublished PhD thesis. University of Kiel.

Pickering M.J. and Garrod S. (2007): Do people use language production to make predictions during comprehension? Trends in Cognitive Sciences 11, 105-110.

Pierrehumbert, J.B. (1981): Synthesizing intonation. Journal of the Acoustical Society of America 70, 985-995.

Pierrehumbert, J.B. and Hirschberg, J. (1990): The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, M.,

and E. Pollack (eds): Intentions in Communication. 271 – 311. MIT Press, Cambridge, MA.

Rayner, K., Slowiaczek, M. L., Clifton, C., Jr. and Bertera, J. H. (1983): Latency of sequential eye movements: Implications for reading. Journal of Exp. Psychology: Human Perception and Performance 9, 912–922.

Reinisch, E., Jesse, A., and McQueen, J.M. (2010): Early use of phonetic information in spoken word recognition: Lexical stress drives eye-movements immediately. The Quarterly Journal of Experimental Psychology 63, 772-783.

Salverda, A.P., and Tanenhaus, M.K. (2010): Tracking the time course of orthographic information in spoken-word recognition. Journal of Exp. Psychology: Learning, Memory, and Cognition 36, 1108-1117.

Salverda, A. P., Dahan, D., and McQueen, J. M. (2003): The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. Cognition 90, 51-89.

Salvucci, D. D., and Goldberg, J. H. (2000): Identifying fixations and saccades in eye-tracking protocols. Proceedings of the Eye Tracking Research and Applications Symposium.New York, 71-78.

Saslow, M. G. (1967): Latency for saccadic eye movement. Journal of the Optical Society of America 57, 1030–1033.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C.G., and Carlson, G. N. (1999): Achieving incremental semantic interpretation through contextual representation. Cognition 71, 109-147.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995): Integration of visual and linguistic information in spoken language comprehension. Science 268 (5217), 1632-1634.

Terken, J.M.B. and Nooteboom, S. (1987): Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. Language and Cognitive Processes 2, 145 – 163.

Terken, J. and Hirschberg, J. (1994): Deaccentuation of words representing 'given' information: effects of persistence of grammatical function and surface position, Language and Speech 37(2), 125–145.

't Hart, J., Collier, R., and Cohen, A. (1990): A Perceptual Study of Intonation. Cambridge: Cambridge University Press.

Weber, A., Braun, B., and Crocker, M., (2006): Finding referents in time: eye-tracking evidence for the role of contrastive accents. Language and Speech 49(3), 367-392.

Weber, A., Grice, M., and Crocker, M., (2006): The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. Cognition 99, B63-B72.

Xu, Y. and Xu, C. X. (2005): Phonetic realization of focus in English declarative intonation. Journal of Phonetics 33, 159-197.