Comb or *coat:* The role of intonation in online reference resolution in a second language

Aoju Chen and Vicky Lai

1 Introduction

In spoken sentence processing, listeners do not wait till the end of a sentence to decipher what message is conveyed. Rather, they make predictions on the most plausible interpretation at every possible point in the auditory signal on the basis of all kinds of linguistic information (e.g.. Eberhard et al. 1995; Alman and Kamide 1999, 2007). Intonation is one such kind of linguistic information that is efficiently used in spoken sentence processing. The evidence comes primarily from recent work on online reference resolution conducted in the visual-world eyetracking paradigm (e.g., Tanenhaus et al. 1995). In this paradigm, listeners are shown a visual scene containing a number of objects and listen to one or two short sentences about the scene. They are asked to either inspect the visual scene while listening or to carry out the action depicted in the sentence(s) (e.g., 'Touch the blue square'). Listeners' eye movements directed to each object in the scene are monitored and time-locked to pre-defined time points in the auditory stimulus. Their predictions on the upcoming referent and sources for the predictions in the auditory signal are examined by analysing fixations to the relevant objects in the visual scene before the acoustic information on the referent is available.

Past work has shown that listeners make use of the links between intonation and information status (i.e. new vs. given in a discourse context) to predict the upcoming referent before the segmental information of that referent is fully available. Dahan et al. (2002), for instance, examined the role of accent placement (i.e. accented vs. unaccented) in online reference resolution in English. They monitored eye fixations to phonemically related referents (e.g., comb, coat) as participants followed pre-recorded bipartite instructions (e.g.. *Put the coat above the square; now put the comb below the circle)* to move objects displayed on a computer screen using a computer mouse. The object noun in the second part of the instruction

' A fixation is an interval in which the eye rests at a region of interest. It is different from a saccade, which is a fast movement of an eye between two fixations (Salvucci and Goldberg 2000).

(hereafter the critical word) was either accented or unaccented. The analysis on fixations concerned the distribution of fixations to the phonemically related referents after the ambiguous segmental sequence (i.e. /kau/) was processed but before the disambiguating segmental information was available (i.e. *Ill* and /m/). Dahan et al. found that accent placement mattered. Following an initial bias towards the referent unmentioned in the first part of the instruction (i.e. comb - the new referent), listeners launched even more fixations to the new referent when the critical word was accented than when it was unaccented. Using the same paradigm, Chen et al. (2007) found that listeners were also highly sensitive to the differences in accent type (i.e. shape of accent). They fixated a new referent more often when the critical word was spoken with a fall (H*L) or a delayed fall (L*HL) than when it was spoken with a rise (L*H) or no accent. Further, they fixated the referent previously mentioned (the 'wrong' referent) longer when it was spoken with L*H or no accent than when it was spoken with H*L or L*HL before they shifted their visual attention to the 'correct' referent."

The evidence reviewed above is confined to online reference resolution in one's native language. In the current study, we address the question as to whether listeners can process intonational information in such an anticipatory fashion when resolving referential ambiguity in a second language (L2). To this end, we conducted an adapted version of Chen et al.'s (2007) Experiment 1 with three intonation conditions (H*L, L*H, and 'deaccented') with intermediate and advanced Dutch learners of English. We focussed on these three conditions because they are highly frequent patterns in English.

Work on intonational processing in L2 is still rather sparse. The limited work suggests that L2 listeners do not process intonation as efficiently as native listeners. Akker and Cutler's (2003) study is most pertinent to our study. They examined the processing of intonational structure of a sentence (i.e. determining where the sentence accent goes) by proficient Dutch learners of English. They found that similar to native listeners, L2 learners could predict which word should be accented in a sentence and directed their attention anticipatorily to that word. The effect of predicted accent was reflected in faster phoneme detection when the phoneme-bearing word was predicted to be accented than when it was not. Further, the effect of predicted accent was significantly smaller when the predicted accent-bearing word was also focused (i.e. carrying new information) than otherwise in native listeners, because focus also led to faster phoneme detection and the focus effect shrank the effect of predicted accent for the sake of parsimonious processing. There was however little difference in the size of the effect of predicted accent between the 'focused' and 'unfocused' conditions in L2 learners. These results imply that L2 learners are able to process intonation in an intonationally similar L2

" Intonational information available before the referential noun (still within or prior to the entire referential expression) can also guide listeners' expectations (Weber et al. 2006; Ito and Speer 2008; Braun and Chen, in press). As these studies are not directly relevant to the current study, we will not discuss them in detail here.

(English vs. Dutch), because of L1 transfer and/or learning effect, as well as focus information. But they fail to efficiently integrate the two processes and adjust the 'depth' of intonational processing accordingly.

Different from Akker and Cutler (2003), we are not concerned with the balancing between two processes but with the processing of intonation as markers of information status. As there is no comparable study to Chen et al. (2007) in Dutch, we can only speculate on what Dutch learners of English may do on the basis of relevant production data. Braun and Chen (2010) examined the intonation of nominal referential expressions in the second part of bipartite instructions like the ones mentioned above in English and Dutch. They found that in both languages, speakers preferred to accent the noun, most frequently with H*L, if it referred to a new object. But if the location to which the new object was moved was also new, Dutch speakers frequently used L*H to realise the noun, creating a hat pattern together with the H*L in the noun depicting the location. Assuming that native speakers of Dutch also associate H*L and L*H with newness and deaccentuation with givenness in online reference resolution in Dutch, three predictions can be put forward on the use of intonation in online reference resolution by Dutch learners of English in English: (1) If L1 transfer is in effect, L2 learners would map H*L and L*H to new information and fixate the new referent more when the critical word is spoken with H*L and L*H than when it is deaccented; (2) If L2 learners have acquired the links between intonation and information status in English, they would map H*L to new information and fixate the new referent more when the critical word is spoken with H*L than when it is spoken with L*H or deaccented; (3) L2 learners may have learner-specific strategies that do not closely resemble the processing of native listeners in either Dutch or English. Akker and Cutler's results would seem to suggest that predictions 1 and 2 were more plausible than prediction 3.

2 Method

2.1 Participants

Thirty-six Dutch learners of English participated in this study with monetary compensation. Among the participants, twenty-four were 4th grade high school pupils with intermediate English proficiency, tested in Best. Twelve were 3rd year English-major university students with advanced English proficiency, tested in Nijmegen. All participants reported to have normal hearing and normal or corrected-to normal vision.

2.2 Materials

Eighteen pairs of phonemically similar nouns (i.e. the cohort pairs) served as the materials for the experimental trials. The two nouns in each pair shared either the same stressed syllable (e.g., candy vs. candle) or the same onset-peak sequence (e.g., comb vs. coat). They were similar in the mean lexical frequencies (33.6 vs. 27.5 per million) (Francis and Kucera 1982). One noun in each pair served as the target (e.g.. comb) and the other as the competitor (e.g., coat).

The cohort pairs were embedded in bipartite instructions. The second part of the instruction (or the second instruction) always mentioned the target (e.g., *now put the comb below the diamond).* The first part of the instruction (or the first instruction) mentioned either the target (e.g.. *Put the comb below the triangle),* marking the target at the onset of the second instruction as 'given' but the competitor as 'new', or the competitor (e.g., *Put the coat below the triangle),* marking the target at the onset of the second instruction as 'new' but the competitor as 'given'. The intonation of the first instruction was kept the same throughout the experiment; the intonation of the second instruction was varied such that the critical word (i.e. comb) was produced with H*L, L*H or no accent, followed by a low intonational phrase boundary tone. Combining the two types of information status of the target/competitor during the second instruction and the three intonation conditions gave us six experimental conditions. Three cohort pairs were assigned to each experimental condition by means of a Latin Square design, leading to six lists of experimental stimuli. In addition, 48 filler trials were constructed.

Each cohort pair was associated with two distractor nouns, resulting in four pictures on each display (Figure 1). In total, 272 Pictures (18 experimental trials 4 pictures + 48 filler trials x 4 pictures) were selected from Snodgrass and Vanderwart's (1980) picture database and the picture database of the Max Planck Institute for Psycholinguistics (MPI). All pictures were black and white line drawings.

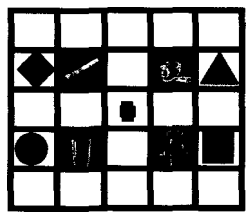

Figure 1. Example of a visual display. Geometric shapes were blue.

The bipartite instructions were recorded (48 kHz, 16 bits) by a prosodically trained male speaker of Southern Standard British English in the sound-proof studio at the MPI. Figure 2 shows example *fo* tracks for the critical word *comb* produced in the three intonation conditions.

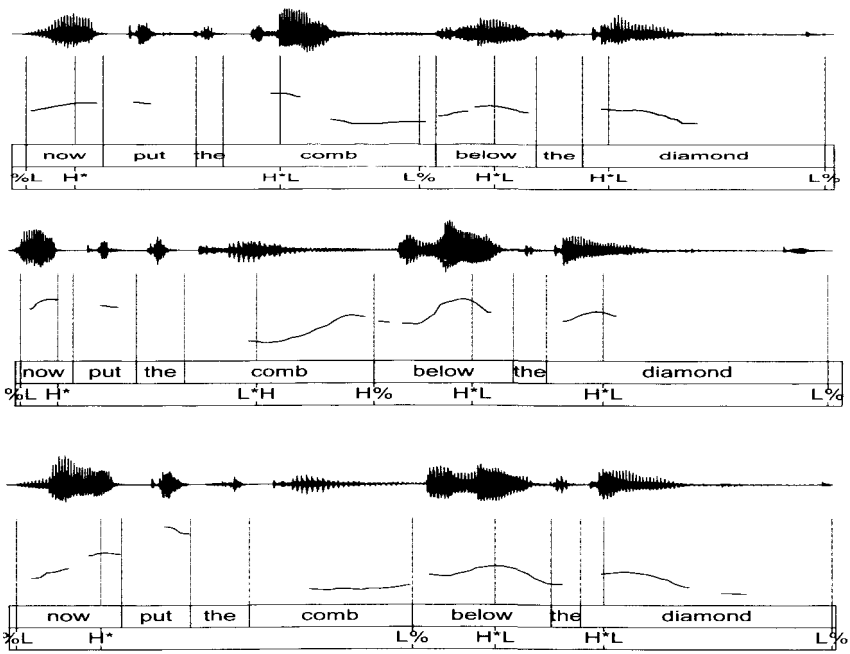

Figure 2. F_0 **tracks for** ***now*** *put* ***the comb below*** *the* ***diamond*** **in** 3 **intonation conditions.**

2.3 Procedures

Participants were tested individually, either in a quite room in Heerbeeck College or in an eyetracking lab at the MPI. They were seated at a comfortable distance from the computer screen. The eye tracker was mounted and calibrated. Eye movements were monitored with a portable SR EyeLink II system. The pre-recorded instructions were presented to participants through headphones. Each trial began with a central fixation point on the screen, which stayed for 500 ms. Then, the grid with four pictures and four geometric shapes appeared on the screen and the first instruction was initiated. The positions of the pictures were randomised across four fixed positions of the grid, while the geometric shapes appeared in fixed positions on every trial. As soon as a picture was moved after the first instruction ended, the second instruction was initiated. Once a picture was moved following the second instruction, the next trial began. The position of the mouse cursor on the computer screen was sampled and recorded, along with the eye-movement data. Automatic drift correction was done after every five trials.

Four advanced learners and two intermediate learners were assigned to each stimulus list. The participants who received the same stimulus list were presented with the stimuli in different orders. To make sure that the intermediate learners understood all object names in English, they were given the pictures and the corresponding nouns a week in advance to familiarise themselves with the words.

2.4 Coding procedure

The data from each participant's dominant eye obtained during the second instructions were coded in terms of fixations, saccades, and blinks by means of the algorithm provided in the Eyelink software. For every 4ms-frame recorded by Eyelink, the fixations from the onset of the critical word till the mouse click to drag and drop the picture were further coded as pertaining to the cell of the competitor referent, the cell of the target referent, the cells of the distractor objects, or elsewhere.

3 Results and discussion

The proportion of fixations to each location (i.e. competitor referent, target referent, distractor objects, and elsewhere) was calculated in l0ms-intervals for each participant per condition. It was done by dividing the total number of trials in which a location was fixated in a given time interval by the total number of trials in which a fixation was launched to any location in this time interval (Salverda et al. 2003). As the minimal latency to plan and launch a saccade is 200-300 ms in visual search tasks (Hallett 1986; Viviani 1990), fixations realised in the first 300 ms of the critical word were likely to be related to speech input preceding the critical word. Since the phonemically similar sequence was on average 232 ms long in the 'deaccented' condition and 274 ms long in the H*L/L*H conditions, effects of intonation conditions were expected in the time window of 300-500 ms after critical word onset (hereafter the critical time window).

We plotted changes in mean proportion of fixations to the competitor referent and the target referent in different conditions over time. Both groups of listeners launched a small number of fixations to the new referent and the new competitor during the critical time window in all three intonation conditions. They began to fixate the new target referent more and the new competitor even less only after the whole critical word was heard and processed. There was thus no effect of intonation in fixations to new referents. However, we did observe changes in fixations to the given target referent and the given competitor referent in the critical time window. We subsequently analysed the changes in fixations in the critical time window for the given target referent and the given competitor referent separately by means of repeated measures ANOVAs. The dependent variable in these analyses was the mean fixation proportion. The independent variables included INTONATION (2 levels) and TIME INTERVAL (20 levels: 300-500 ms in a 10-ms interval). Note that the variable INTONATION consisted of three levels in our experimental design, i.e. H*L, L*H and 'deaccented'. The pattern of fixations to the given target/competitor referent differed consistently between one and the other two conditions. Repeated measures ANOVAs with all three conditions did not yield a significant interaction of INTONATION x TIME INTERVAL, probably because of the similarity between two of the conditions and individual variation in the data. We thus conducted separate repeated measures ANOVA in which we compared the fixation pattern in one

condition to the fixation pattern in each of the other two conditions. Hence the variable INTONATION consisted of two levels in each analysis.[3]

3.1 The intermediate learners of English

3.1.1 Fixations to the given competitor referent

During the critical time window, there was a steady increase in fixations to the given competitor referent when the critical word was spoken with L*H or deaccented, but not when it was spoken with H*L. In the H*L condition, the proportion of fixations was overall low (< 30%) and exhibited a decreasing trend in the middle part of the critical time window. These patterns were in line with Chen et al.'s (2007) finding that native speakers of English associated L*H and deaccentuation with givenness but H*L with newness. However, the differences in fixation proportions between L*H and deaccentuation on the one hand and H*L on the other hand did not reach statistical significance. This might be due to individual variation in the data.

3.1.2 Fixations to the given target referent

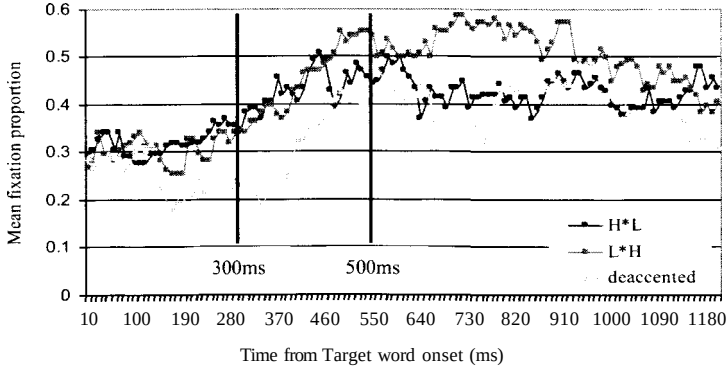

Figure 3. Mean proportion of fixations to the given target referent in the intermediate learners in three intonation conditions, starting from the critical word onset till 1200 ms after the critical word onset. The two vertical dark lines indicate the begin and end of the critical time window (300-500 ms after the critical word onset).

Figure 3 displays the changes in the mean proportion of fixations to the given target referent in the intermediate learners. As can be seen, there was a steady increase in fixations to the given target in all three intonation conditions throughout the critical

We are aware that this approach to the statistical analysis is not the standard approach and the results stemming from this approach should be validated on a larger set of data using ANOVA's including all the three intonation conditions.

time window. However, the increase was more substantial in the H*L and L*H conditions than in the 'deaccented' condition. We conducted two repeated measures ANOVAs to assess the difference between conditions. The variable INTONATION consisted of the H*L and 'deaccented' conditions in one analysis and the L*H and 'deaccented' conditions in the other conditions. Both analyses revealed a marginally significant main effect of INTONATION, indicating that the differences in fixation proportions between the H*L condition and the 'deaccented' condition ($F(1, 22) = 4.03$, $p = .057$) and between the L*H condition and the 'deaccented' condition ($F(1, 22) = 3.21$, $p = .087$) were relatively reliable. These results suggest that the intermediate learners tended to associate H*L and L*H with givenness but deaccentuation with newness. This stood in contrast to the results from the given competitor referent, which suggested that the intermediate learners associated L*H and deaccentuation with givenness and H*L with newness.

3.2 The advanced learners of English

3.2. J Fixations to the given competitor referent

The proportion of fixations to the given competitor increased in all three intonation conditions during the critical time window, with an even stronger trend of increase in the H*L and L*H conditions. But more fixations were launched to the given competitor in the 'deaccented' condition. The difference in the mean fixation proportion between the 'deaccented' condition and the two 'accented' conditions suggested that the advanced learners seemed to associate deaccentuation with givenness but accentuation with newness regardless of accent type, differing from the native speakers of English in Chen et al.'s (2007) study in their responses to L*H. Again, the differences in fixation proportions between the 'accented' conditions and the 'deaccented' condition did not reach statistical significance, possibly due to the combined effect of individual variation in the data and the small number of subjects (N=12).

3.2.2 Fixations to the given target referent

Figure 4 displays the changes in the mean proportion of fixations to the given target referent in the advanced learners. As can be seen, the proportion of fixations to the given target increased steadily in all three intonation conditions in the critical time window. However, the increase was more substantial in the H*L condition than in the L*H and 'deaccented' conditions. As a result, the mean proportion of fixations was higher in the H*L condition than in the other two conditions. We conducted two repeated measures ANOVAs to access the significance of the differences between conditions. In one analysis, the variable INTONATION consisted of the H*L and L*H conditions; in the other analysis, it consisted of the H*L and 'deaccented' conditions. [4] The analysis concerning the H*L and L*H conditions revealed a

The mean proportion of fixations differed between conditions also before information in the critical word was processed, which could only be triggered by intonation information in the 'now put the' sequence. This is hard to explain as the

significant main effect of INTONATION, indicating that the differences in the mean proportion of fixations between the H*L and L*H conditions (F(1, 11) = 5.65, p = .037) was reliable. The analysis concerning the H*L and 'deaccented' conditions revealed a significant interaction of INTONATION and TIME INTERVAL (F(1, 220) = 1.76, p = .026). This indicated that the difference in the increase in proportion of fixations over time between the H*L and 'deaccented' conditions was robust. These results thus suggest that the advanced learners associated H*L with givenness but L*H and deaccentuation with newness, exactly opposite to what native speakers of English did in Chen et al.'s (2007) study. Further, the results were not consistent with the results from the given competitor, which suggested that the advanced learners associated deaccentuation with givenness and H*L and L*H with newness.

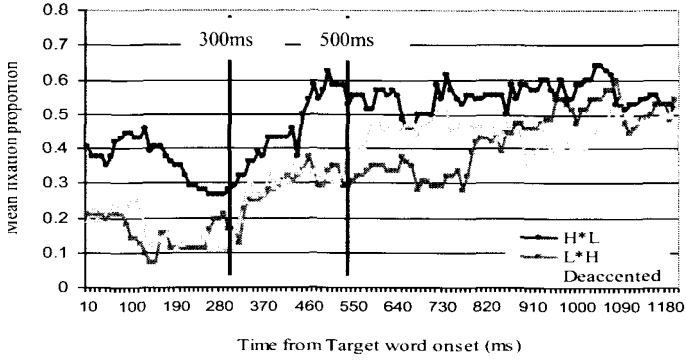

Figure 4. Mean proportion of fixations to the given target referent in the advanced learners in three intonation conditions, starting from the critical word onset till 1200 ms after the critical word onset. The two vertical dark lines indicate the begin and the end of the critical time window (300-500 ms after the critical word onset)

4 General discussion and conclusions

The overall low proportion of fixations to the new competitor referent and the new target referent during the critical time window suggested that the learners hardly considered a new referent as a candidate for the upcoming referent. They simply waited till the critical word was fully available to decide what to look at. Intonation did play a role in the learners' fixations to the given target referent and the given competitor referent. The results were however of a mixed nature. For both the intermediate and advanced learners, the patterns of fixations to the given competitor suggest a more sensible use of intonation in reference resolution. The intermediate learners appeared to associate H*L with newness and L*H and deaccentuation with givenness, like native speakers of English, in line with prediction (1) based on

'now' sequence was spoken with a high rise (H*L) across conditions and 'put' was always unaccented.

successful mastery of L2 intonation. The advanced learners appeared to associate H*L and L*H with newness and deaccentuation with givenness, in line with prediction (2) based on LI transfer. Surprisingly, the patterns of fixations to the given target suggest that both groups of learners hardly made use of the links between intonation and information status when predicting the upcoming referent. The intermediate learners associated H*L and L*H with givenness but deaccentuation with newness; the advanced learners associated H*L with givenness but L*H and deaccentuation with newness.

If the learners made use of the links between intonation and information status in online reference resolution in the case of the given competitor referent, how can we explain their responses in the case of the given target referent? The answers to this question probably lie in the fact that the critical word always represented the target referent. Segmental information such as coarticulation effects in the phonemically similar sequence could potentially reveal the lexical identity of the word. For example, the sequence /kau/ from 'comb' might bias listeners' guess to 'comb' due to nasalisation at the end of the sequence. The intermediate learners fixated the given target referent more in the H*L and L*H conditions than in the 'deaccented' condition. This may be because the phonemically similar sequence had greater spectral clarity due to accentuation (Koopmans-van Beinum and van Bergem 1989), which in turn made it easier for the learners to detect coarticulation effects and guess which word it came from. In the case of the advanced learners, the extra processing load caused by mapping the words onto the pictures for the first time during the experiment might have washed out the word recognition advantage in the L*H condition because it was an acoustically weaker accent than H*L. Intonation thus played a role in modulating the fixations to the given target referent via the effect of accentuation on spectral clarity, instead of via the interface between intonation and information status.

The two processes of predicting the upcoming referent can be reconciled in the following scenario. That is, the learners tended to stay fixating the previously-mentioned referent at the beginning of the second instruction. This bias was modulated over time by both segmental and intonational information in the critical time window. The learners relied primarily on segmental information if it pointed to the previously mentioned referent as the upcoming referent, as in the case of the given target referent. But the learners became more attentive to intonational information if segmental information did not point to the previously mentioned referent as the upcoming referent, as in the case of the given competitor referent. Since previous studies using a similar set of target-competitor word pairs (Dahan et al. 2002; Chen et al. 2007) found clear evidence for the use of the links between intonation and information status in online reference resolution in native speakers of English, this hypothetical scenario suggested a plausible difference between native listeners and L2 listeners. Specifically, the native listeners primarily relied on the links between intonation and information status to predict the upcoming referent, whereas the L2 listeners seemed to attend segmental information first and resorted to the links between intonation and information status only when segmental information went against their expectation (i.e. the previously-mentioned referent as the upcoming referent).

Regarding the learners' use of intonational information, there was a difference between the two groups of learners. The intermediate learners seemed to exploit the links between intonation and information status just like the native listeners in Chen et al.'s (2007) study. The advanced learners were less native like than the intermediated learners. They associated L*H with newness, as native speakers of Dutch do in Dutch (Braun and Chen 2010). Previous studies have also reported better performances from less proficient learners in certain tasks (e.g., van Heugten, Lodestijn, and Son 1982; Chen 2009). This could be because the intermediate learners were more on guard against L1 transfer or because they are simply better informed intonationally as a result of even more exposure to spoken English than the advanced learners.

In conclusion, our results suggest that L2 learners adopted learner-specific strategies in their use of acoustic information to predict the upcoming referent (prediction 3). The links between intonation and information status were exploited but only if the referent of interest was already mentioned and if the segmental information did not support their bias towards a previously mentioned referent.

5 Acknowledgements

We are grateful to Ton Kox and Frans Lemmerling for making the testing at Heerbeeck College possible. We thank the pupils and their teachers Karin Wensink, Annemiek Stam, Marjon van Winkelhof at Heerbeeck for their cooperation. Thanks also go to Marieke Hoetjes and Rik van den Brule for assistance in testing and data processing.

6 References

Altmann, Gerry T.M. & Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73. 247-264.

Altmann, Gerry T.M. & Yuki Kamide. 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(4). 502-518.

Akker, Evelien & Anne Cutler. 2003. Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition* 6(2). 81-96.

Braun, Bettina & Aoju Chen. In press. Now for something completely different: anticipatory effects of intonation. In O. Niebuhr, & H. Pfitzinger (eds), *Prosodies, context, function, and communication.* Berlin: de Gruyter.

Braun, Bettina & Aoju Chen. 2010. Intonation of'now' in resolving scope ambiguity in English and Dutch. *Journal of Phonetics* 38(3). 431-444.

Chen, Aoju, Els den Os, & Jan Peter de Ruiter. 2007. Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. *The Linguistic Review* 24. 317-344.

Chen, Aoju. 2009. Perception of paralinguistic intonational meaning in a second language. *Language Learning* 59(2). 367-409.

Dahan, Delphine, Michael K. Tanenhaus. & Craig G. Chambers. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47. 292-314.

Eberhard, Kathleen. M., Michael J. Spivey-Knowlton, Julie C. Sedivy, & Michael K. Tanenhaus. 1995. Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research* 24. 409-436.

Hallett, Peter. 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.

Ito, Kiwako & Sharon R. Speer. 2008. Anticipatory effect of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58, 541-573.

Koopmans-van Beinum, Florien J., & Dick R. van Bergem. 1989. The role of given' and 'new in the production and perception of vowel contrasts in read text and in spontaneous speech. *EUROSPEECH* 1989. 2113-2116

Salverda, Anna. P., Delphine Dahan, & James M. McQueen. 2003. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90, 51-89.

Salvucci, Dario D. & Joseph H. Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Eye Tracking Research and Applications Symposium.* New York, 71-78.

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

van Heugten, M., Lodestijn, A., & van Son, N. 1982. *Interference van het Nederlands bij het aanleren van Engelse intonatie patronen.* Unpublished manuscript, Radboud University Nijmegen, The Netherlands.

Viviani, Paolo. 1990. Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In *Eye Movements and their Role in Visual and Cognitive Processes,* Eileen Kowler (ed.), 353-393. Amsterdam: Elsevier.

Weber, Andrea, Martine Grice, & Matthew Crocker. 2006: The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition* 99, B63-B72.