Use of L+H* for immediate contrast resolution

Kiwako Ito and Shari R. Speer

Department of Linguistics
Ohio State University, USA
{ito; speer}@ling.ohio-state.edu

Abstract

Previous eye tracking studies [1] demonstrated that a prominent L+H* accent in English instructions evoked a contrastive relationship between discourse referents, immediately guiding listeners' eye movements to contrastive candidates in the visual scene. However, the visual layout in the previous study conflated looks to contrastive candidates with those to previously fixated object groups, thus calling into question the cause of the very early fixations to contrastive targets. In order to eliminate the possibility that L+H* evokes contrast only when the visual layout allows easy detection of the target, the current experiment used layouts where contrastive candidates had separate visual domains. The results confirm the immediate effect of L+H*, demonstrating a faster increase in fixations to contrastive targets than with H*. In addition, results show an intonational 'garden-path effect' due to the immediate integration of L+H* as contrast marking accent.

1. Introduction

Since Pierrehumbert proposed the pitch accent inventory for American English within the autosegmental metrical framework nearly three decades ago [2], most experimental research has focused on advancing our understanding of the phonetic properties of various pitch accents [3, 4, 5]. Less is known about how listeners interpret different types of pitch accent during natural discourse, and whether and when the previously proposed pragmatic distinctions among various tunes [6] affect spoken language comprehension. We investigated listeners' online reactions to a range of intonation patterns using a real-world eye movement monitoring paradigm [1].

In previous experiments [1], participants followed auditory instructions to decorate holiday trees with multi-colored ornaments laid out on a grid. Ornaments were sorted by type, so that each cell on the grid contained a set of ornaments of the same kind in different colors. The decoration sequences were carefully constructed to include contrasts between consecutively-mentioned ornaments (e.g., *Next, hang the blue ball.* → *After that, hang the green ball.*). In those contrastive sequences, the color adjective was either accented with L+H* (e.g., *After that, hang the GREEN ball.*) or with H* (e.g., *After that, hang the green ball.*). The results demonstrated the robust effect of prominent accent L+H*, which produced very early looks to the target cell (i.e., *balls*), approximately 150-200 ms faster than those in a comparable condition with H*. Because the fixations to the target cell started to increase from the *offset* of the adjective, and programming and execution of saccades generally takes at least 100-150 ms for a simple visual search task, the authors argued that listeners' eye movements were based primarily on the accentual information on the color adjective. This provides evidence for the claim that L+H* can evoke a set of alternatives for the upcoming referent in a discourse [6].

In addition to the above facilitative effect of L+H*, the previous study also demonstrated that a L+H* can act as a *false alarm* in speech comprehension. In a locally non-contrastive sequence such as "*Next, hang the blue ball* → *After that, hang the GREEN drum*", participants initially fixated the previously mentioned ornament cell (balls) due to the strong anticipation of contrast triggered by the L+H* on *GREEN*. This also resulted in delayed eye movements to the correct target, *drum*.

Although the above results suggest the immediate processing of the accentual information that conveyed contrastive status on the target discourse entity, there remains a possibility that results were due to the display setup. In the previous study, the authors argued that a L+H* on a pre-nominal modifier evoked a contrast between the accented color adjective and the color of the immediately preceding ornament, and that this contrastive link between the two color adjectives led to a mapping between the two modified nouns, allowing the anticipation of a specific candidate for the upcoming object. Thus, as soon as participants heard L+H* on the color, they expected the same object noun as that of the previous trial. Note that this expectation was reflected in the rapid returns to the cell that contained the preceding target. In other words, participants did not have to *search for* the contrastive candidate in a separate location from the just-visited cell. Thus a question arises as to whether the effect of L+H* was observed because anticipating the same type of object merely required a return to the just-visited visual domain. If the contrast-evoking effect of L+H* is genuinely facilitative, we should be able to demonstrate anticipatory eye movements due to accentual information on the pre-nominal modifier even when the target candidates appear in visual areas different from the previous target. The current experiment was designed to test the anticipatory effect of L+H* with a *true* visual search task.

2. Experiment

2.1. Design and Materials

Participants decorated holiday trees following pre-recorded auditory instructions. Each participant decorated three trees using the ornaments laid out on three separate grids. Four types of ornaments (3 targets: star, tree, ball, and 1 filler: heart) were painted in three colors (red, yellow and green), yielding 12 ornament sets that occupied 12 cells on each grid. (The size of ornaments was unified within a grid but altered across the three grids to distract participants from the experimental manipulation.) Each cell contained four identical ornaments. The three target ornaments in three colors were

distributed to occupy nine out of ten peripheral cells surrounding the two central cells. The two central cells and the remaining one peripheral cell were occupied by the filler ornaments (i.e., hearts). The locations of ornaments were altered across the three boards. An example ornament grid is shown in Figure 1.

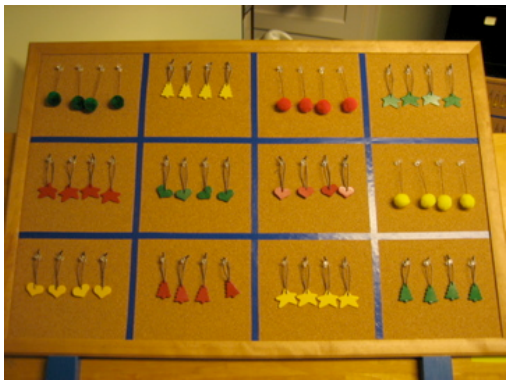

Figure 1: *Example ornament layout on a grid*

The auditory instructions were recorded by a trained female phonetician who maintained her overall pitch range and speech rate within and across conditions [1]. Each tree was decorated with 26 ornaments, and the orders of decoration were constructed to include locally contrastive sequences such as "*Hang a red star*. → Next, hang *a yellow star*.", and locally non-contrastive sequences such as "*Hang a yellow tree*. → *Next, hang a green ball*." The noun phrases in those critical utterances had one of two pitch accent patterns: (1) L+H* on the adjective, and no accent on the noun. E.g., *hang a YELLOW*$_{[L+H^*]}$ *star* $_{[no\ accent]}$, (2) H* on the adjective, and !H* on the noun. E.g., *hang a yellow*$_{[H^*]}$ *star*$_{[!H^*]}$. All the instruction utterances were ToBI transcribed by an annotator blind to the experimental design. The example F0 traces and the ToBI transcriptions are given in Figure 2.

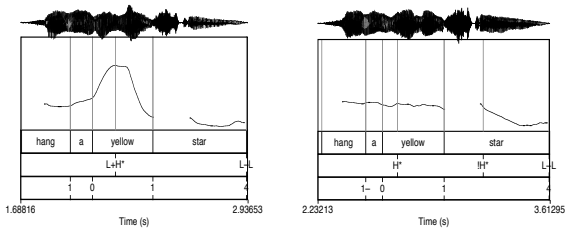

Figure 2: *Example F0 traces and ToBI annotations for [L+H* noaccent] (left) and [H* !H*] (right).*

Table 1: *Mean duration and F0 of target noun phrases.*

Sequence & Accentual Pattern	Adj dur (ms)	Adj F0 (Hz)	N dur (ms)	N F0 (Hz)
contrast [L+H* noacc]	330	299	489	148
contrast [H* !H*]	332	207	549	164
no contrast [L+H* noacc]	320	300	491	150
no contrast [H* !H*]	316	208	558	163

Table 1 summarizes the mean word duration and the mean F0 values of the adjectives and the nouns for each of the four critical conditions. On average, color adjectives' F0 were approximately 90 Hz higher when produced with L+H* than with H*. The nouns' F0 were about 14 Hz lower without an accent than with a !H*. The effect of post-focal phonetic attenuation for the nouns was observed predominantly in their much shorter duration as compared to their !H*-accented renditions (mean 490 ms vs. 554 ms).

2.2. Participants

36 undergraduate students at the Ohio State University participated in partial fulfillment of a course requirement. Data from 30 native speakers of American English are analyzed below.

2.3. Eye-tracking procedure

Participants sat in front of a drafting table with the top tilted at 35 degrees to support the ornament display board. They wore lightweight headgear fitted with an eye-camera and a magnetic transmitter that functioned to correct measured eye positions for head movement. Participants followed instructions from a pre-recorded speaker to choose an ornament from the grid and place it on a small tree located to their right. The x and y coordinates of eye-fixations on the board were recorded at 60 Hz using ASL Eye-Trac 6 data-collection system. The experimenter monitored the participant's eye locations and body orientations via a ceiling-mounted camera, and pressed a key to play each instruction when the participant had finished hanging an ornament and had faced back to the board.

3. Results

Each of the 30 participants had 9 trials in each of the four critical conditions. The fixation proportion was calculated for each time point by dividing the total number of actual fixations to the target by the total number of possible fixations. Five trials were excluded from this calculation because participants reached for an ornament not mentioned in the instruction.

A comparison between the two conditions with the contrastive discourse sequence confirmed the facilitative effect of L+H* as compared to H* in searching for a contrastive target. When participants heard a sequence such as "*Hang a red star*. → Next, hang *a YELLOW star*.", the fixations to the target ornament cell increased faster than when they heard, e.g., "→ Next, hang *a yellow star.*" As Figure 3 shows, the fixation proportions to the target started increasing about 200 ms before the onset of the noun (e.g., *star*) in both contrastive conditions, but the two lines diverged at the onset of the noun. Until about 300 ms past the noun offset, fixation proportions to the target were consistently higher for [L+H* no-accent] than for [H* !H*] trials. Since the duration of adjectives did not differ between the two conditions and the nouns were longer with !H* than without an accent, the faster increase in fixations with [L+H* no-accent] cannot be attributed to any durational advantage for planning and executing saccades. It is important to note that the fixations to the target started increasing in both conditions as soon as the participants heard the color adjectives. This contrasts with the previous study [1] where the onset of the increase in fixations occurred around the onset of the noun. This early effect indicates that with the present visual display,

which sorted ornaments by both color and object type, listeners immediately used the adjective segmental information to constrain the candidates for the target noun.

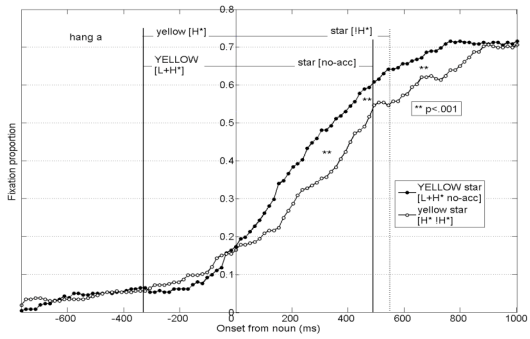

Figure 3: *Fixation proportions to the target in two contrastive sequences: red star→YELLOW/yellow star*

The fixation patterns for the non-contrastive sequence trials also confirmed the effect of L+H* in evoking contrast. When participants heard sequences such as "*Hang a yellow tree. → Next, hang a GREEN ball.*", looks to the contrastive competitor (e.g., *green tree*) kept increasing until about 200 ms into the target noun (Figure 4). Unlike the previous study [1], the fixations to the contrastive competitor did not exceed the fixations to the target. Nonetheless, the contrastive competitors (e.g., *green tree*) were fixated equally frequently as the targets (e.g., *green ball*) until enough segmental information of the noun was processed to select the correct target. Importantly, the contrastive competitors were fixated more frequently than the other same color competitors (e.g., *green star & green heart*), and contrastives' competition lasted longer than that of the other same-color competitors. Figure 4 also exhibits the lower increase and earlier decrease in the averaged fixations to the non-contrastive color competitors as compared to the contrastive competitor. These results do not merely confirm that listeners first used the segmental information of the adjective to narrow down the set of candidates, but they also demonstrate that listeners utilized the intonational information to further activate the contrastive candidate, leading to its sustained competition with the target until the segmental information of the noun singled out the target object.

Fixation patterns for the non-contrastive sequence with [H* !H*] provided further evidence for the effect of accentual difference, and it also revealed the plausible strategy that participants adopted for the present visual search task. With non-contrastive sequences such as "*Hang a yellow tree. → Next, hang a green ball.*", listeners initially fixated non-contrastive candidates (e.g., *green star & green heart*) more than the contrastive candidate (e.g., *green tree*) (Figure 5). This may indicate that the participants generally assumed a non-contrastive sequence in the experimental task, and thus directed their attention to non-repeated ornament types. (Note that the fixations to the target were expected to compete with those to the non-contrastive competitors. We do not as yet have an account for the low number of initial fixations to the target shown in Figure 5.) The tendency to initially avoid the contrastive competitor is observable also in Figure 4, where the initial fixation rise for the contrastive competitor is lower than the other same color candidates. In the absence of L+H* on the color, however, the fixations to the contrastive competitor did not rise as rapidly and never surpassed the fixations to other same color candidates. Instead, fixations to both the contrastive competitor and the other competitors decreased gradually, as the noun's segmental information guided fixations to the correct target.

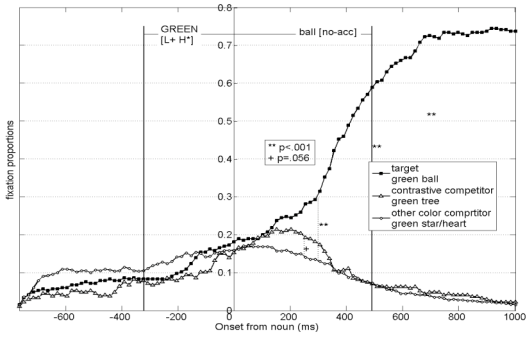

Figure 4: *Fixation proportions for a non-contrastive sequence with L+H*: yellow tree→ GREEN ball.*

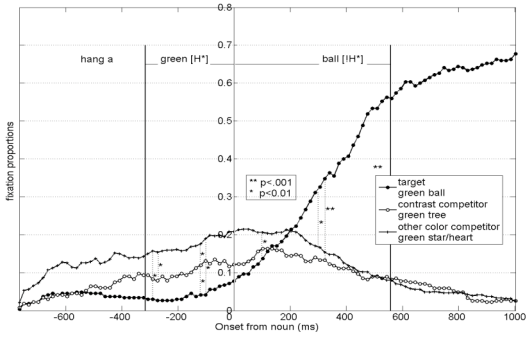

Figure 5: *Fixation proportions for a non-contrastive sequence with H*: yellow tree→ green ball*

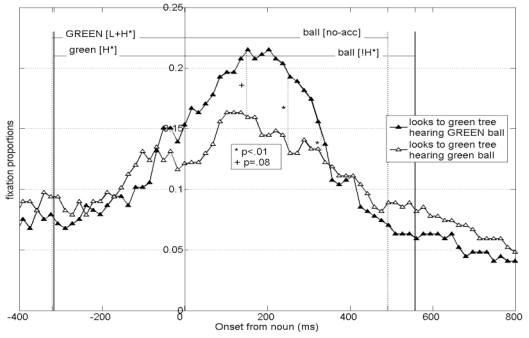

Figure 6: *Fixation proportions to the contrastive competitor: yellow tree→ GREEN/green ball*

The direct comparison of fixations to the contrastive competitor in the two non-contrastive sequences is shown in Figure 6. The two lines diverge at around 50ms before the noun onset, with fixations to the contrastive competitor rising approximately 5% more with L+H* than H*. Again, this difference confirms the effect of L+H* on the adjective in evoking a contrast set for the upcoming noun. In contrast, H*

accents did not draw additional attention to the contrastive competitor.

4. Discussion

Using visual layouts that grouped objects by both color and shape, the present eye tracking study confirmed the robust facilitative effect of the prominent accent L+H* in visual search. L+H* accent on a prenominal modifier evoked a local contrast between the consecutively mentioned discourse entities, and thus speeded the detection of the contrastive object even when the target objects were located in non-predictable areas different from the immediately preceding object cells. Since the effect appeared from the onset of the noun (Figure 3) and the duration of the adjective did not differ across the two contrastive sequence conditions, we attribute the faster increase in fixations primarily to the accentual difference in the adjectives. The results endorse the use of eye movement monitoring for investigating prosodic effects in comprehension. With a carefully designed layout of objects, unconscious eye movements can reveal the immediate use of intonational cues in speech processing.

In addition to the above effect of L+H* in true visual search, the present results demonstrated the incremental integration of modifiers' segmental information, and also a potential general strategy for the given task. First, unlike the previous study where each cell contained ornaments in multiple colors, the current layouts allowed the early selection of candidate sets according to the information of prenominal adjective. Because ornaments were sorted by color as well as by shape, hearing a particular color enabled the immediate elimination of the objects in other colors. Listeners seemed to rapidly integrate the color information to limit their choices to the four out of the twelve sets of objects, and started sending their eyes to those limited sets accordingly, as shown in the early onset of fixation rises at around 200 ms before the noun onset across Figures 3-6. Those early fixation increases confirm incremental speech processing previously reported [1, 7, 8], i.e., listeners do not wait to respond to speech input until they hear the last word of the critical phrase.

Although the color information was used promptly to narrow down the candidate sets, the data also suggest that listeners did not pay an equal amount of attention to each of the four possible candidate sets across trials. Instead, they seemed to follow the assumption that the decoration sequence would not repeat the same ornament type. This was shown in the very infrequent initial fixations to the potential contrastive competitors in non-contrastive sequences (Figure 4 & 5). In both conditions, participants fixated the other color competitors more frequently than the contrastive competitor, which would be the ornaments of the same shape as the immediately preceding ornament. An alternative account for this result is that detection of an ornament with the same shape as that mentioned on the immediately preceding trial was extremely easy, and thus participants did not have to fixate the contrastive competitors as frequently as the other competitors to spot their approximate locations. Unfortunately, the present data do distinguish between these two possibilities.

Whether it was the 'assume-no-repetition' strategy or the 'must-check-different-shapes' strategy, the fixation patterns across the present layouts indicated a bias against contrastive competitors. Note that this initial bias supports the robust impact of L+H* in evoking the contrastive interpretation of incoming speech. In spite of the initial infrequent fixations, L+H* on the color adjectives led to the immediate recovery of attention to contrastive competitors and let them compete against the actual targets until past 200 ms after the noun onset. Although the fixations to the contrastive competitor never exceeded those to the target, the size of fixation increase (approx. 5% to 20 %), or the magnitude of the L+H* garden-path effect in the present data was in fact comparable to that in the previous study (10% to 25%) [1]. Again, such a steep increase of fixations to the contrastive competitor was not observed when the color adjectives were heard with the less prominent H* accent. In sum, the present data did not simply confirm the strong intonational garden-path effect, but they also demonstrated that it could overcome the initial bias.

5. Conclusions

The controversy over the pragmatic distinctions among the various accentual types cannot be resolved without empirical evidence for the claims about the function of each accent. The current naturalistic discourse task allowed the observation of listeners' responses to different accent types during purposeful spoken language comprehension. The present study hopefully provides convincing evidence for the contrast-evoking effect of a prominent accent that has been conventionally labeled as L+H*. An ongoing experiment further investigates the impact of L+H* with subsective adjectives such as scalars (small, medium and large), which may express a more context-dependent contrastivity of the objects than intersective adjectives such as colors. The interaction between semantic properties of words and intonational marking of discourse status remains as an important theme for our future research.

6. References

[1] Ito, K. & Speer, S. R. (in press). Anticipatory effect of intonation: Eye movements during instructed visual search. *Special Issue of Journal of Memory and Language on Language-Vision Interactions in Sentence Processing*.

[2] Pierrehumbert, J. B. (1980). The phonology and phonetics of English intonation. PhD dissertation. Massachusetts Institute of Technology.

[3] Ladd, D. R (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

[4] Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The Original ToBI System and the Evolution of the ToBI Framework. In Jun, S-A. (Ed.), *Prosodic Typology The Phonology of Intonation and Phrasing*. New York: Oxford University Press Inc.

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

[6] Pierrehumbert, J., & Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 342-365). Cambridge: MIT Press.

[7] Kamide, Y., Altman, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,*133-156.

[8] Brown-Schmidt, S. & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language, 54,* 592-609.