

Online construction of implicature through contrastive prosody

Heeyeon Y. Dennison, Amy J. Schafer

Department of Linguistics, University of Hawai‘i at Mānoa, Hawai‘i

{yoon, aschafer}@hawaii.edu

ABSTRACT

Little experimental evidence exists for how prosodic/intonational information might affect the generation of an implicature. We provide online evidence that the combination of an L+H* pitch accent and an L-H% boundary tone work together to imply a contradiction, and that this contour has distinct effects from an L+H* L-L% tune. We also compare the online processing of changes in meaning suggested by prosody versus explicit negation. The results highlight the importance of intonational information in sentence understanding, and the differences in processing prosodically cued contrastive information versus lexical negation.

Keywords: visual search, contrastive prosody, implicature

1. INTRODUCTION

Recent psycholinguistic studies using an eye-tracking paradigm have provided detailed evidence on how the contrastive pitch accent L+H* is interpreted in the mind of listeners (e.g., [5, 12, 13]). These and similar studies have explored how the pitch accent delivers contrastive meaning by evoking an alternative set, when the alternative set is composed of objects depicted in a visual display (e.g., a blue star vs. a red star, both present in a display). These studies limited their investigations to the effects of pitch accents that were combined with only one type of edge tone: L-L%. Here, we consider how the type and location of pitch accents as well as the type of boundary tone affect the on-line comprehension of implied meaning.

We tested sentences such as *Lisa had the bell*, in which contrastive accent on *had* and a sentence-final rise work together to support the implicature that Lisa no longer has the bell. In this paper we call this implicature state contradiction. Our sentences were presented as part of a visual world task, in which eye movements were tracked during the comprehension of critical sentences. The task required participants to use the computer mouse to click on depictions of the mentioned objects (e.g., the bell). Here, we focus on predictions and results from the click component of the task.

One goal of the study was to test the compositionality hypothesis for tune meaning [8]. The hypothesis states that each component of prosodic units (e.g., pitch accent, boundary tones) carries particular meanings and they are combined to contribute to the interpretation of the whole tune. The particular meaning of L+H* that is discussed in [8] is to evoke a salient scale that promotes the accented item while rejecting some alternative items. Using three types of tunes—contrastive, emphatic, and neutral (see Table 1 below for experimental conditions), we examined whether an L+H* accent on the critical word was sufficient to create the state contradiction meaning, versus whether the implicature was dependent on the combination of the contrastive accent and a subsequent L-H% end contour.

We also looked at a prosodically matched set of negative sentences, such as *Bart didn’t have the bee* (see Table 1). Researchers have argued that interpreting negative

sentences requires the processing of both the negated meaning and the actual meaning [4]. For example, [6] presented experimental results showing that upon reading a sentence like *The door was not open*, people think of both an open door and a closed door. Therefore, a second goal of this study was to compare the time course of processing a type of derived meaning signaled by a lexical cue (i.e., not) versus a prosodic cue (i.e., L+H* and/or L-H%).

2. METHOD

2.1. Materials

A visual field presented on a computer monitor was split in half to depict the rooms of two cartoon characters: Lisa Simpson’s room on the left and Bart Simpson’s room on the right. Figure 1 shows an example display. Each room contained 10 to 12 objects, one of which matched a test sentence (e.g., *Lisa had the bell*). Displaying multiple objects in a visual search task avoids the closed-set issue addressed by researchers who are concerned that any effects from simple displays may be artifacts of task-specific strategies (e.g., [5, 10]). Further, our task compared conditions in which the mentioned object was or was not found in the room of the referent of the sentential subject. The visual search task thus increased the difficulty for locating the mentioned object, allowing us to evaluate whether prosodically conditioned implicature and negation influenced a shift in attention from the sentential subject’s room to the alternate room (or the whole display).

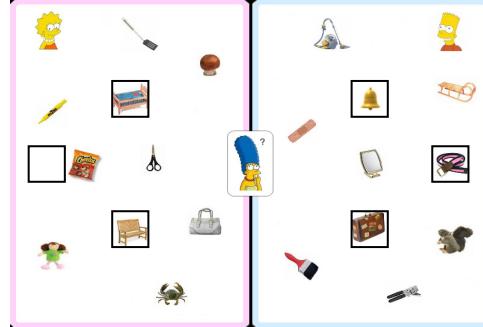


Figure 1. An example of a room display
(The six squares were not shown to participants.)

All target objects had monosyllabic names. This was to ensure participants had opportunities to perceive prosodic signals—especially the final rising or falling boundary tones—before the critical segmental information was fully disclosed. To further prevent object search purely based on segmental information, each target item (e.g., bell) was displayed with three cohort items (e.g., belt, bed, bench)—one located in the same room as the target and two positioned in the other room. To participants, the displays appeared as if objects were moved for every trial. There were six designated locations, however, where the target and cohort objects appeared. The six squares marked on Figure 1 represent the critical cells (they were not shown to participants).

2.2. Participants and Procedures

Sixty four native English speakers participated in this study and received either course credit or \$10. Participants were seated in front of a desktop computer and wore a lightweight head-mounted eye-tracker (ASL E5000).

The experiment was run by E-prime, Version 1.2 (Psychology Software Tools, Inc.). Participants were first presented with a cover story that provided a clear behavioral goal: to find and click on the objects mentioned in sentences. The story stated that some creature moves belongings of Lisa and Bart around and this upsets the kids. The mom in the story, Marge Simpson, constantly recites ‘who had what’ from her memory to aid the return of objects to the correct owner’s room. This story line ensured a felicitous use of the past tense verb ‘had’ both for ongoing possession of an object and for a switch of possession. Participants in a pilot study confirmed the naturalness of this set-up.

Each trial consisted of two visual searches directed by two separate sentences. Target sentences, however, appeared only as the first sentence in a trial. All of the second sentences were filler items. In each trial, the onset of the display was synchronized with the onset of the first sound file. Participants listened to a sentence (e.g., *Lisa had the bell*) and then clicked on the object mentioned in the sentence (e.g., the bell) as quickly as possible. Mouse click reaction times (RTs) were measured from the offset of each sound file. Upon the mouse click (i.e., the completion of the first search), another sentence was played while the visual display remained the same. Overall, each participant completed a total of 80 randomly-presented pairs of visual searches, comprised of 32 target and 128 filler sentences.

2.3. Experimental Conditions and Predictions

The reason for using the past tense was to induce one of the dominant implications—but not anymore—from the L+H* L-H% tune (i.e., state contradiction: [3]). The experimental set-up biased participants to perceive ‘Lisa had the bell ...but now Bart has it.’ That is, given the specific layout of the experiment where the possessor of an object can only be either Lisa or Bart (i.e., a limitation on the number of alternative possessors), we predicted that participants who perceived implied state contradiction would quickly switch their looks from the room of the mentioned possessor (i.e., Lisa) to the room of the implied possessor (i.e., Bart).

The experiment contained 2 sets of 4 comparable conditions: the first set with four affirmative sentences and a second set with four negative sentences. The affirmative and negative conditions used different items for both the sentential subject and the objects. The affirmative set always used the sentential subject *Lisa*, while the negative set always employed *Bart*. Filler sentences disguised this manipulation.

ToBI and acoustic analyses confirmed that the sentences had similar overall durations (mean: 1.09 seconds for affirmative sentences, 1.15 seconds for negative sentences) and were produced with the intended intonation (see Table 2 for key measurements).

Each condition in the affirmative set (C1–C4 in Table 1) had an equivalent condition in the negative set (C5–C8), using three types of tunes: contrastive, emphatic, and neutral tunes. In all conditions except C2 and C6, the object location was faithful to the sentence meaning, under the assumption that the prosodically implied state contradiction meaning would be dominant in the contrastive conditions (C1 and C5); see [3]. In C2 and C6, the object location was not supported by the sentence meaning, if we assume that the L+H* L-L% tune emphatically affirms the literal meaning stated in the sentence. These two false conditions served as critical

comparisons for the contrastive tune conditions (C1 and C5), as well as the emphatic true conditions (C3 and C7), respectively. Fillers limited the proportion of false trials to just 7% and normalized the range of tunes.

Table 1. Sentence types, tunes, and truth of the display

	Sentences and Tunes	Object location		Truth
		L's	B's	
C1	L+H* L-H% Lisa HAD the bell. (contrastive)		bell	True
C2	L+H* L-L% Lisa HAD the bell. (emphatic)		bell	<i>False</i>
C3	L+H* L-L% Lisa HAD the bell. (emphatic)	bell		True
C4	H* H*L-L% LISA had the BELL. (neutral)	bell		True
C5	L+H* L-H% Bart DIDN'T have the bee. (contrastive)		bee	True
C6	L+H* L-L% Bart DIDN'T have the bee. (emphatic)		bee	<i>False</i>
C7	L+H* L-L% Bart DIDN'T have the bee. (emphatic)	bee		True
C8	H* H* H*L-L% BART DIDN'T have the BEE. (neutral)	bee		True

Table 2. Duration (in ms) and F0 Max (in Hz) for test items

Aff. Set	Lisa		had		the		Object	
	ms	Hz	ms	Hz	ms	ms	Hz	
Cont.	264	190	332	245		48	487	217
Emph.	269	191	333	254		51	439	171
Neut.	274	194	247	187		45	495	227
Neg. set	Bart		didn't		have	the	Object	
	ms	Hz	ms	Hz	ms	ms	ms	Hz
Cont.	264	192	235	275	195	44	414	209
Emph.	262	200	210	265	211	44	392	160
Neut.	254	193	194	200	231	39	454	226

2.3.1. Hypotheses for the affirmative set

Assuming highly incremental sentence processing and a tight relationship between attention and visual behavior (e.g., [10, 11]), we predicted that the subject name *Lisa* in the auditory stimuli would direct participants’ looks to Lisa’s room in the display. Therefore, we expected that looks to and mouse clicks on the target objects would be faster in conditions C3 and C4 than in any other condition, since the rest of the linguistic information following the subject name supports the referent of the sentential subject as the possessor of the mentioned object. We also predicted that we might see shorter mouse click RTs in C4 than in C3. The direct object carries a presentational pitch accent (H*) in C4, but not in C3, and accentuation is known to facilitate information processing (e.g., [5]).

Further, we predicted that the rising edge tone L-H% in C1 would motivate a shift of looks from the mentioned subject’s room (i.e., Lisa’s) to the other room (i.e., Bart’s), if participants are able to construct implied state contradiction

from the L+H* L–H% contour. If an L+H* accent alone is sufficient to induce the implicature, participants should change their looks from Lisa's room to Bart's room also in C2 and C3, upon listening to the accented verb *had*. In this case, click times should be faster in C2 than in C3, since in C2 but not C3 the target objects are presented in Bart's room and participants should look at Bart's room as soon as they have processed the contrastive accent.

However, if the L+H* by itself is not sufficient to systematically induce a shift of looks to the other room (i.e., the L–H% is necessary for the target implicature), then click times should be shorter in C1 than in C2. The shift of looks in C1 is motivated by a linguistic signal, whereas the shift of looks in C2 is due to a search failure in the absence of any supporting linguistic signal.

2.3.2. Hypotheses for the negative set

As in the affirmative set, the sentential subject name in the negative set (Bart in this case) will initially direct participants' looks to the corresponding subject's room. Upon listening to the negation word *didn't*, participants must now process the negation. We considered two possibilities for how a negative sentence is processed and directs attention to a visual display.

First, negative sentence processing might be fundamentally similar to affirmative sentence processing. In that case, participants' behavior with respect to the negative conditions will resemble those in the affirmative conditions, except that participants will begin by looking in the contrasting room (with respect to the sentence literal interpretation: i.e., Bart's room). Upon listening to the negation word *didn't*, participants will shift their looks to Lisa's room, and the rest of the information will be processed as in the affirmative set.

Alternatively, if the comprehension of negative sentences involves activating both the stated (negated) meaning and the contradictory (ultimate) meaning (e.g., [6]), then processing negative sentences is fundamentally different from that of affirmative sentences. Since affirmative sentences mainly evoke the stated meaning, we would expect visual attention to be directed to just the sentential subject's room. But if negative sentences evoke the stated meaning and the contradictory meaning, we would see a shift in visual attention that includes the alternative room as well as the current room: an expansion of the search fields from Bart's room alone to the whole space. The expansion of visual search fields upon the word *didn't* would then greatly reduce differences in mouse click RTs across the four conditions, because participants would be rigorously scanning items in both rooms as the rest of the linguistic information unfolds, regardless of the prosodic pattern.

2.3.3. Hypotheses for the comparison of the two sets

Comparing the affirmative and negative sets together, we predicted three types of mouse click RTs overall: short, medium, and long RTs. The short mouse click RTs will result from C3 and C4 since these conditions guarantee successful object searches within one room (i.e., no need to switch rooms). The long mouse click RTs will come from C1 and C2; participants will initially search objects in one room, but because of either a prosodic cue (C1) or search failure (C2), participants will shift their searches to the other room. Since participants will not be motivated for a shift from Lisa's room until the cue arrives (C1) or the search of the first room fails (C2), their mouse click RTs on the targets in Bart's room will be much slower. The mouse click RTs from negation conditions will fall in between, on either view of how negation is processed. Participants will start their

searches in one room, but they will quickly either switch rooms, or expand their search fields to the whole space. Because of this shift, mouse click RTs in the negative set will be slower than those in C3 and C4. However, either the early looks to Bart's room or the early search-field expansion will reduce the overall search times for C5 and C6 compared to those in C1 and C2.

3. RESULTS

Data from four participants were excluded from the analysis: two for experimenter error, one for difficulty performing the task correctly, and one for mean click RTs that exceeded 2.5 SD from the overall participants' mean. This paper thus presents data from 60 valid participants. Mouse click data showed overall mean accuracies of 99.96% from the participant analysis and 96.9% from the item analysis. There were no significant differences in click accuracy across conditions.

Mouse click RTs were analyzed from accurate trials only. Click RTs were trimmed with a fixed-cut off at 6000 milliseconds. This removed 1.81% of total data. Then, long RTs exceeding 2.5 SD from each participant's mean were replaced with the value of that participant's mean plus 2.5 SD [9]. Table 3 below presents mean mouse click RTs for each condition from the participant analysis.

Table 3. Mean mouse click RTs from participants (in milliseconds, measured from the offset of the sentence)

Condition	Affirmative set	Negative set
Contrastive true	C1	1973
Emphatic false	C2	2220
Emphatic true	C3	1181
Neutral true	C4	1072
		C5
		1675
		C6
		1571
		C7
		1533
		C8
		1458

Results from the affirmative set indicate strong support for the hypothesis that the state contradiction interpretation depends on the presence of the sentence-final rise. One-way ANOVA tests for the affirmative set indicated significant differences among these four conditions: $F_1(3,236)=76.03$, $p<.000$; $F_2(3,60)=18.29$, $p<.000$. Paired t-tests confirmed that correct mouse clicks in both the contrastive true (C1) and emphatic false (C2) conditions took significantly longer than mouse clicks in the emphatic true (C3) and neutral true (C4) conditions, all at t_1 : $p<.000$, t_2 : $p<.000$. Note that this includes the critical comparison of C2 and C3; RTs were significantly shorter in C3 than in C2, not the reverse (t_1 & t_2 : $p<.000$). In addition, mouse click RTs in C1 were significantly faster than in the emphatic false condition (C2), highlighting the facilitatory role of linguistic cues in visual search (t_1 : $p<.000$, t_2 : $p=0.12^1$). Also, click RTs in the emphatic true condition (C3) versus the neutral condition (C4) approached significance by participants, although not by items (t_1 : $p=.06$, t_2 : $p=0.15^2$).

As for the negative set, one-way ANOVA tests indicated no meaningful differences across the four conditions: $F_1(3,236)=1.44$, $p=0.23$; $F_2(3,60)=0.46$, $p=0.71$. RTs in the contrastive true (C5) and emphatic false (C6) conditions were numerically longer than RTs in the emphatic true (C7) and neutral true (C8) conditions, but these differences were not statistically valid.

¹ Two items produced much lower accuracy in one of their conditions (*bell*: 67% in C2, *lime*: 73% in C1), due to cohort competition. After removing these two items, the item analysis was marginally significant: $p=.06$.

² After excluding two lowest accuracy items, the p value is .09.

We also found that mouse click RTs for each pair of comparable conditions in the affirmative and negative sets were significantly different. For the contrastive pair (C1 vs. C5), click RTs in the affirmative set were significantly slower than those in the negative set from the participant analysis: $t_1(1,59)=2.00$, $p=.004$, $t_2(1,30)=2.04$, $p=.206$. For the the emphatic false pair (C2 vs. C6), click RTs in the affirmative set were significantly slower than those in the negative set from both participant and item analyses: $t_1(1,59)=2.00$, $p<.000$, $t_2(1,30)=2.04$, $p=.007$. In contrast, click RTs for the emphatic true pair (C3 vs. C7) and the neutral true pair (C4 vs. C8) were significantly slower in the negative set than in the affirmative set: $t_1=2.001$, $p<.000$ for both pairs; for C3 vs. C7, $t_2=2.04$, $p=.014$; for C4 vs. C8, $t_2=2.04$, $p=.002$.

4. DISCUSSION

Previous studies have shown immediate and even predictive use of the contrastive pitch accent, L+H*, during online reference resolution using affirmative sentences (e.g., [5, 13]). However, the present study provides evidence that the type of accompanying boundary tones affects how and when the pitch accent information influences online processing of affirmative sentences. The results suggest that state contradiction implied from the L+H* L-H% tune requires the integration of the pitch accent (L+H*) and boundary tones (L-H%). The pitch accent L+H* alone was not strong enough to evoke the implicature.

One might expect that in the current study the L+H* alone should encourage shift of looks to the alternative subject's room. The lack of such effects might suggest that the presence of the emphatic true condition, where L+H* together with L-L% signaled emphatic affirmation of the stated meaning, weakened the alternative-set evoking function of L+H*. However, the conditions also included the emphatic false condition, which placed the target objects in the alternate subject's room. This condition could have supported the alternative-set evoking function of L+H*. In fact, out of three conditions containing L+H*, two placed the target objects in the alternate subject's room. However, the results showed that participants perceived the implied contradiction only when they encountered the rising edge tone L-H% along with the L+H*. This suggests the importance of the whole tune and its distinctive meaning (e.g., [1, 2, 7]).

More importantly, however, it suggests a clear separation between evoking alternative sets with promotion of the accented item (e.g., asserting that Lisa HAD the bell), versus making use of an alternative set in drawing an implicature (e.g., the implied state contradiction that Bart now has the bell). The L+H* accent on the verb *had* promotes the state information "had" while contrasting it to the alternative state "had not". However, it is the rising edge tones L-H% that leads participants to use this contrastive information to construct an implicature.

The current results hence provide empirical evidence for the compositionality hypothesis for tune meaning [8]. That is, the target implicature discussed in the current study (Lisa had the bell...but now Bart has it) results from the distinct functions of each prosodic element of the tune, L+H* and L-H%. We expect our eye-fixation data will be valuable in evaluating the compositionality hypothesis more closely. In particular, the proportion of looks to the alternate subject's room upon receiving L+H* versus L-H% will tell us to what extent just the former cue or the combination are responsible for the alternative-set evocation and the resulting state contradiction implicature.

Lastly, this study suggests differences in the processing time course for a contradictory meaning that is signaled by a lexical cue (i.e., *not*) versus a prosodic cue (i.e., L+H* and/or L-H%). The mouse click RT patterns suggest that the contradictory meaning perceived by the prosodic tune reflects a clear shift of attention from the literal meaning to the implied meaning. The contradictory meaning constructed by the explicit negation, however, may allow lingering activation of the literal meaning while the ultimate contradictory meaning is being constructed. Future research will be needed to clarify to what extent the persistence of the residual activation of the literal (negated) meaning is based upon task demands versus an integral part of the processing of negative sentences.

5. ACKNOWLEDGEMENTS

This work was supported by a NSF doctoral dissertation improvement grant, BCS-0921696. We thank Victoria Anderson for assistance with auditory stimuli recording.

6. REFERENCES

- [1] Cruttenden, A. (1997). *Intonation*. NY: Cambridge University Press.
- [2] Dainora, A. (2002). Does intonational meaning come from tones or tunes? Evidence against a compositional approach. Paper presented at the *Speech Prosody*, Aix-en-Provence, France.
- [3] Dennison, H. Y., Schafer, A. J., & Anderson, V. B. (2008). Perception of contrastive meaning through the L+H* L-H% contour, *Proc. JASA*, 124(4), 2497.
- [4] Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *Journal of Cognitive Systems Research* 1, 19-33.
- [5] Ito, K., & Speer, S. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, 58, 541-573.
- [6] Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, 38, 1033-1050.
- [7] Liberman, M., & Sag, I. A. (1974). Prosodic form and discourse function. *Tenth regional meeting, Chicago Linguistic Society*, 10, 416-427.
- [8] Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan & M. E. Pollack (Eds.), *Intentions in communication* (pp. 270-311). Cambridge: MIT Press.
- [9] Ratcliff, Roger. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3), 510-532.
- [10] Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 863-900). London: Academic Press (Elsevier).
- [11] Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.
- [12] Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs L+H*. *Cognitive Science*, 32(7), 1232-1244.
- [13] Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, 49(3), 367-392.