The role of intonation in early word recognition and learning

Jill C. Thorson and James L. Morgan

[1] Department of Cognitive, Linguistic and Psychological Sciences,
Brown University, Providence, USA
Jill_Thorson@brown.edu, James_Morgan@brown.edu

Abstract

The motivation for our study is to investigate how English-acquiring toddlers are guided by the mapping between intonation and information structure during on-line reference resolution and in novel word learning tasks. We ask whether specific pitch movements (*deaccented*, *monotonal*, *bitonal*) more systematically predict patterns of attention and subsequent novel word learning abilities depending on the referring or learning condition (*new, given, contrastive*). Experiment 1 examines the attentional patterns of 18-month-old toddlers when referents are either *new* or *given* in the discourse, and carry one of the three pitch movement types. Contrary to previous work, results show increased attention to the target in the deaccented condition if the referent is *new* to the discourse. Also, both monotonal and bitonal pitch movements direct attention to the target even when the target is *given*. Thus, pitch type interacts with information structure in directing toddler attention. Experiment 2 tests two-year-olds in a novel word learning task, varying pitch type and contrastiveness during learning. Preliminary results show that learning is aided when the novel word is introduced in contrast to a previous referent. Together, these two experiments demonstrate the role of pitch type and information structure in guiding attention and aiding early word learning.

Index Terms: intonation, information structure, first language acquisition

1. Introduction

From birth, infants are sensitive to native language rhythm and pitch patterns ([1], [2]). They can then approximate adult-like intonation contours from the onset of production ([3], [4], [5]) and align these contours with felicitous semantic and pragmatic intentions ([6]). However, little research has been conducted on the early comprehension of contours as they reflect information status.

Previous research shows that toddler attention to referents can be mediated by both intonation and information structure in discourse ([7]). In turn, attention to a referent is essential for making the correct word-to-object or word-to-action mappings necessary for early word learning ([8]). The motivation for our study is two-fold: 1) to investigate how American English-acquiring 18-month-olds are guided by mappings from intonation to information structure during on-line reference resolution in discourse (Experiment 1), and 2) to investigate how the pitch accent on a novel word interacts with its referential status to aid early word learning in American English-acquiring 24-month-olds (Experiment 2).

Social pragmatics and intentionality are often cited as essential in order to achieve successful word learning ([9], [10]). Our study tests outcomes when live interactions are removed from the experimental design and any degree of intentionality is only accessible through the utterances themselves and their corresponding prosody.

This study focuses on two pitch accents in American English, the simple monotonal H* and the complex bitonal L+H*. Previous work in adult speech perception shows that the simple H* pitch accent can be associated with either new or contrastive information in discourse, and the complex L+H* is more typically associated with a contrastive interpretation ([11]). Our experiments exploit these mappings in American English in order to test how these pitch accents interact with referential newness, givenness, and contrastiveness during early attentional processes and word learning.

2. Experiment 1

Previous research by Grassmann and Tomasello (2010) claimed that German-acquiring two-year-olds attend to a referent of a familiar word if and only if the word is both *stressed* and *new* to the discourse ([7]). Their experiment consisted of three conditions where the target referent could be introduced in a short discourse context with (a) stress only, (b) newness only, or (c) stress and newness. They used a live speaker to present the stimuli to each subject in order to ensure a level of intentionality on the side of the speaker and measured looking time and pointing to a referent as their dependent variables.

Our experiment expands upon the work by Grassmann and Tomasello (2010) but makes a number of methodological changes. First, we add in a deaccented condition to the design to act as a control against the accented (or 'stressed') conditions. Second, we ask whether specific pitch movements more systematically predict patterns of attention to a referent, rather than using one "stressed" category. Third, since a live speaker may have produced varying intonation contours, we control for speaker variations by using pre-recorded stimuli. Finally, we isolate the role of pitch in guiding attention, holding duration and intensity constant. This is a first step towards understanding how each of the acoustic correlates of intonation contribute and interact in guiding attention.

Specifically, we consider how unique pitch movements facilitate attention when a referent is either *new* or *given* in the discourse. The pitch types tested are a simple monotonal rise (~H*), a complex bitonal rise (~L+H*), and a deaccented control pattern. First, we predict that regardless of pitch type, newness will guide attention to a referent. Second, a semantically and pragmatically appropriate pitch accent will guide attention to both *new* and *given* referents.

2.1. Method

2.1.1. Participants

Data were analyzed for 48 American English-acquiring 18-month-old toddlers (27 female). The age of participants ranged from 529 to 589 days, with a mean age of 551 days. Thirteen additional participants were discarded due to fussiness (9), experimental error (1), or equipment malfunction (3). All participants were from Providence, RI, USA, and surrounding areas.

2.1.2. Stimuli

An adult female speaker produced all target and distractor utterances. In order to create the three different pitch types, the speaker first produced carrier sentences with H* accents on the target words using careful speech (slow and clear), but not child-directed speech. The pitch contour of only the target word was then digitally manipulated in Praat ([12]) to create the simple monotonal and complex bitonal versions. Deaccented target words were spliced into the carrier phrase and matched for duration and intensity to the two accented versions. All test stimuli were resynthesized. Naïve listeners judged the resynthesized target speech sounds as natural.

There were 6 (C)CVC monosyllabic target words and 18 (C)CVC monosyllabic distractors used in the procedure (See Table 1 for a full list of stimuli by trial). All target words were phonologically distinct within a trial and primarily sonorant in nature. All of the target and distractor words are commonly known by 18-month-olds. Previous knowledge of the words was confirmed by a vocabulary questionnaire completed by the caregiver. If the toddler was not familiar with a particular word, then that particular trial was eliminated during analysis (this was not a common occurrence, and it only affected at most one trial per participant).

Table 1. *Stimuli for Experiment 1.*

Trial Number	Target Word	Distractors/Fillers
Practice	spoon	cake
		bear
		fish
1	ball	sock
		lamb
		cat
2	moon	dog
		shoe
		book
3	cow	pig
		tree
		duck
4	doll	bus
		cup
		sun
5	plane	star
		truck
		dress

2.1.3. Design and procedure

We used a 2x3-mixed design to test toddlers' responses to changes in information status and intonation, isolating the specific role that pitch plays in directing attention to *new* or *given* referents. The independent variable of Information Structure *(new* vs. *given)* was manipulated within subjects, whereas the independent variable of Pitch Type *(deaccented, simple, complex)* was manipulated between subjects.

Each trial consisted of a Context Phase and a Test Phase. During the Test Phase, the test utterance played and the *proportion of looking time (PLT) to the target* was collected using the SMI iView X™ RED eye tracker. The Context Phase consisted of two parts (or slides), which established the Prior Discourse Context before the Test Phase was introduced (See Figure 1). Importantly, the target referent in the Test Phase was either *new* or *given* to the discourse as well as in contrast

to an item in the previous slide, making both the simple (~H*) and the complex (~L+H*) movements acceptable in this location.

There were 5 familiarization trials and 2 test blocks per between-subjects condition. A test block included 1 practice trial and 5 test trials for a total of 6 trials in each block. With two blocks per condition (one *new* and one *given* condition), there were a total of 2 practice trials and 10 test trials (12 total trials). Trial order within a block was randomized and block order was counter-balanced across participants. The location of the target items on the screen (left or right) was also counter-balanced within and across conditions.
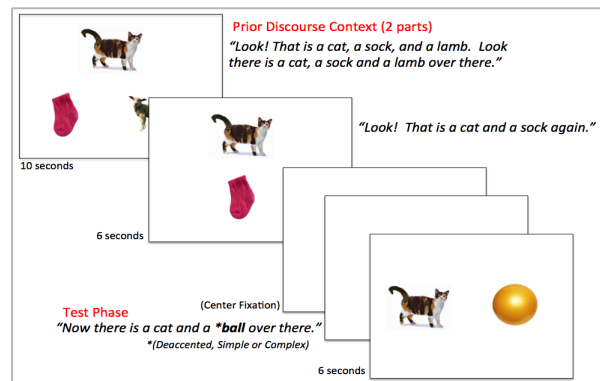

Figure 1: *Example new trial from Experiment 1.*

2.2. Results

A 2x3 repeated measures ANOVA shows a significant main effect of Information Structure ($F(1,45) = 32.36$, $p < .001$, $\eta_p^2 = .418$) and a significant main effect of Pitch Type ($F(2,45) = 6.34$, $p = .004$, $\eta_p^2 = .220$). The two-way interaction of Information Structure by Pitch Type approached but did not quite reach significance ($F(2,45) = 3.01$, $p = .059$).
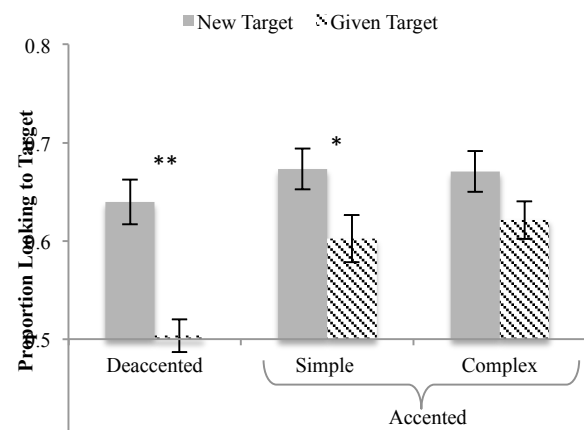

Figure 2: *Bar graph of proportion looking to a new or given target referent for each pitch type condition. Error bars show +/- SE. *: p < .05, **: p < .01.*

Planned comparisons between groups reveal more looking to a target than a distractor when the referent is *new* to the discourse context, regardless of pitch accent type (See Figure 2). Additionally, both types of accentuation (*simple* and

complex) guide more looks to the target than to the distractor when the target referent is either *new* or *given* to the discourse context. In the *simple* condition, there was significantly longer looking to the target when it is *new* to the discourse than when it is *given*. This difference between *new* and *given* target referents does not reach significance in the *complex* condition. Crucially, there are more looks to the target than the distractor referent in all conditions except when the target is *deaccented* and *given* in the discourse.

2.3. Discussion

Contrary to previous literature, *newness* is sufficient to draw 18-month-olds' attention to a referent in a discourse, even without pitch accentuation. A preference for the novel (or *new*) stimulus item over a familiar (or *given*) one suggests a more mature level processing by 18-months-olds ([13], [14]). Toddlers prefer to look at the more prominent or salient item in the discourse, where salience in this case is achieved through pitch movement and newness effects.

Additionally, even in the case of a target referent that is *given* in the discourse, both *simple* and *complex* pitch movements guide attention to this referent. Thus, the presence of either *newness* or a *pitch accent* shifts attention to a target referent in a discourse, regardless of pitch movement type. This suggests that the more salient or prominent the stimulus item, the more a toddler will look. When a target is both new and carries a pitch movement, the result is even greater looking to the target. For *given* information, attention is being driven by the pitch movement on the referent word.

Importantly, we observe robust effects for the variable of pitch type even when only the acoustic dimension of pitch is manipulated. With intensity and duration held constant, any observed pitch accent effects were a consequence of pitch (f0) manipulation. Thus, even with a more complex discourse phase and limited acoustic cues to stress, we find a very different pattern of results from previous research.

In this experiment, the two pitch types analyzed were a simple monotonal and a complex bitonal pitch movement. Future work will extend analyses to other types of pitch accents, discourse contexts, as well isolate the roles of other acoustic cues to prosody (i.e. duration and intensity). Critically, significant methodological differences change the complexion of results in comparison to previous work. Understanding the mechanisms for guiding attention is essential for subsequent early word learning.

3. Experiment 2

As demonstrated in Experiment 1, toddler attention to a referent is mediated by both the intonation and the information structure of the discourse. Experiment 2 extends these findings to investigate how pitch type interacts with contrastiveness during a novel word learning task.

Both newness and pitch were shown to aid in guiding attention during a discourse context in Experiment 1, where the target item was also in contrast to a previous referent. Contrastiveness here is defined as introducing a referent in direct opposition to a one that is previously mentioned. This type of contrast without pitch accentuation on the target word was not sufficient to guide attention to the intended referent. Interestingly, with the addition of a context appropriate pitch accent movement, attention increased to a target referent over a distractor. The goal of experiment 2 is to explore how discourse contrastiveness interacts with intonation to aid in early word learning.

First, we predict that contrastive learning situations are more likely to shift attention to a referent and aid in word learning, particularly when paired with the more prominent contrastive accent ([15]). Second, from Experiment 1 we know that accentuation facilitates attention to a referent. Thus, we predict that contrastiveness and the presence of a pitch accent will aid in the learning of a novel word.

3.1. Method

3.1.1. Participants

Preliminary data were analyzed for 12 American English-acquiring 24-month-old toddlers (7 female). The age of participants ranged from 722-767 days, with a mean age of 743 days. Four additional participants were discarded due to fussiness (3) or inattentiveness (1). All participants were from Providence, RI, USA, and surrounding areas.

3.1.2. Stimuli

The same female speaker who recorded the stimuli for Experiment 1 produced all utterances using careful speech (slow and clear), but not child-directed speech. The speaker was trained in intonational phonology and was able to produce H* and L+H* accents consistently across the different stimulus items. Each target word was produced naturally in an utterance with an H* accent, a L+H* accent, or as deaccented. The target word was then spliced out of the original utterance and into a carrier sentence. This ensured that the only difference between the different pitch type conditions was the pitch accent of the target word, and not the other parts of the sentence. All splices were made at zero-crossings.

Stimuli included two CVC monosyllabic target novel words and 4 CVC monosyllabic distractors (See Table 2 for a full list of stimuli by learning condition). All target words were phonologically distinct. All of the distractor words were animals commonly known by 24-month-olds. Previous knowledge of the words was confirmed by a vocabulary questionnaire completed by the caregiver. Novel words were associated with novel animals designed for this experiment. There are four novel animals, two for each learning condition.

Table 2. *Audio stimuli for Experiment 1.*

Learning Condition	Novel Target Word	Distractors/Fillers
Noncontrastive	*wug* (IPA: /wʌg/)	sheep bear pig
Contrastive	*neem* (IPA: /nim/)	duck bear pig

3.1.3. Design and procedure

We tested participants using a 2x3 mixed design with *contrastiveness* (within-subjects) and *pitch type* (between-subjects) as independent variables. *Contrastiveness* varied in how the target word was presented during learning, either *contrastive* or *noncontrastive* in relation to a preceding discourse element/referent (see Figure 3 for an example of a

noncontrastive learning condition). The *pitch type* variable consisted of three levels. A novel target word could bear a *simple* monotonal (H*) pitch accent, a *complex* bitonal (L+H*) pitch accent, or a *deaccented* pattern.

In a *contrastive* learning condition, the target word was always presented in opposition to one referent in the preceding discourse element (e.g. *Look! There is a bear and pig over there. Oh! There is a bear and a WUG_contrastive over there.*). In a *noncontrastive* learning condition, the target novel word was presented as presentationally new in relation to the previous discourse referents (e.g. *Look! There is a bear and pig over there. Oh! There is a sheep and a wug_noncontrastive over there.*) (See Figure 3). In this condition, the novel target was never presented in contrast to a previous referent.

An experimental condition consisted of a two-part Learning Phase and a Test Phase. The Learning Phase consisted of two 12-trial learning blocks (one *contrastive* and one *noncontrastive*). A distinct novel word was presented in each of these learning blocks. Each block included 4 target trials (where the novel word-animal pairing was presented), 4 distractor trials (where a second unnamed novel animal was presented), and 4 filler trials of familiar animals. The Test Phase consisted of 10 trials: 6 test trials and 4 control trials (See Figures 3 and 4). During a control trial, two familiar animals were presented and the toddler was asked to identify one of the referents. During a test trial, the target novel animal was presented with the distractor novel animal. The *proportion of looking time (PLT) to the target* was collected during the Test Phase to assess whether or not the child learned the novel word-animal pairings. We used the same eye-tracking paradigm as Experiment 1.

Block order and novel animal assignments were counterbalanced across conditions. Test trials were randomized and the location of the target item (left or right) was counterbalanced within and across conditions.
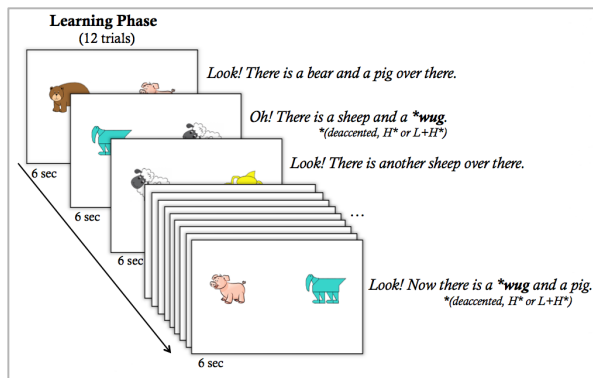

Figure 3: *Example noncontrastive learning phase from Experiment 2. The novel target word 'wug' could bear one of three pitch accent types depending on the condition.*
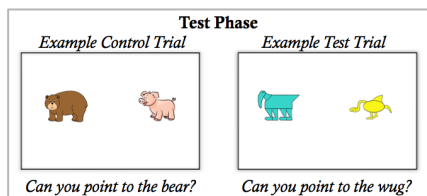

Figure 4*: Example control and test trials from the Test Phase of Experiment 2.*

3.2. Results

Preliminary results exhibit two primary patterns (See Figure 2). The conditions that show substantially more looking to the target novel animal at test are the *contrastive* accented conditions, both *simple* and *complex*. In addition, there is also more looking to the target novel animals in the *noncontrastive-complex* condition.
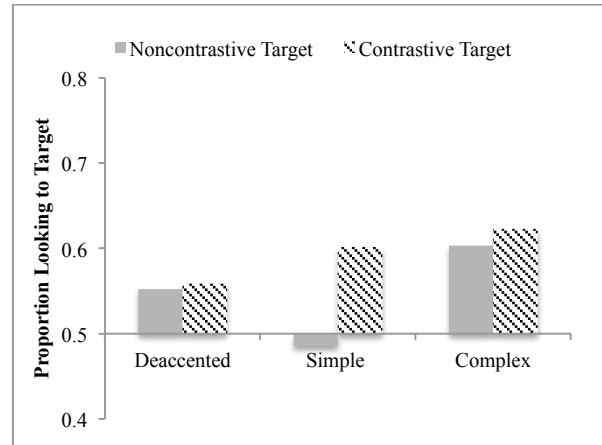

Figure 5: *Bar graph of proportion looking to a contrastive or noncontrastive target novel word referent for each pitch type condition.*

3.3. Discussion

As predicted, the initial pattern of results suggest that contrastiveness, with either a simple or a complex pitch accent, guides more looks to the target novel animal at test. Additionally, the use of a complex pitch accent during the noncontrastive learning situation also directs more looks to the target novel animal than the distractor. These results suggest learning of the novel word-animal association in these conditions, which we predicted to be the most salient during learning. Overall, these data show a preliminary pattern for how intonation and contrastiveness interact to aid in early word learning. Data collection is ongoing.

4. General Discussion

Intonation and information structure both play a role in directing toddler attention and facilitating early word learning. Experiment 1 demonstrates that either newness or the presence of a pitch movement (i.e. only f0 variation) guide 18-month-olds' attention during reference resolution. Future work will test other acoustic correlates of intonation (intensity and duration), and analyze how they interact to guide attention. Preliminary results from Experiment 2 show that contrastiveness and/or a complex pitch accent aid 24-month-olds in learning a novel word. Even with live interactions removed from both methodologies, this set of experiments show the interacting effects of prosody and information structure on attention as well as in successful word learning.

5. Acknowledgements

6. References

[1] Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 756-766.

[2] Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development, 21*(4), 779-784.

[3] Prieto, P., & Vanrell, M.M. (2007). Early intonational development in Catalan. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVI International Congress of Phonetic Sciences* (ICPhS) (pp. 309-314). Saarbrücken, Germany.

[4] Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVI International Congress of Phonetic Sciences* (ICPhS) (pp. 315-320). Saarbrücken, Germany.

[5] Frota, S., & M. Vigário. (2008, August). The intonation of one-word and first two-word utterances in European Portuguese. Paper presented at the *XI International Conference for the Study of Child Language Conference* (IASCL), Edinburgh, Scotland.

[6] Prieto, P., Estrella, A., Thorson, J. & Vanrell, M.M. (2012). Is prosodic development correlated with grammatical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*, *39*(2).

[7] Grassmann, S., & Tomasello, M. (2010). Prosodic stress on a word directs 24-month-olds' attention to a contextually new referent. *Journal of Pragmatics*, *42*(11), 3098-3105.

[8] Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development, 10*, 201-224.

[9] Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child development, 67*, 635-645.

[10] Grassmann, S. & Tomasello, M. (2007). Two-year-olds use primary sentence accent to learn new words. *Journal of Child Language, 14*, 23-45.

[11] Watson, D., Gunlogson, C., & Tanenhaus, M. (2008). Interpreting pitch accents in on-line comprehension: H* vs L+H*. *Cognitive Science, 32*, 1232-1244.

[12] Boersma, P. & Weenink, D. (2012). *Praat: A system for doing phonetics by computer*, available at http://praat.org.

[13] Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146,* 668-670.

[14] Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology, 18*, 704-713.

[15] Katz, J. & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language, 87*(4), 771-816.