Testing the relevance of prenuclear accents for predicting intonational meaning in German

Timo B. Roettger[1,2], Michael Franke[3], & Jennifer Cole[1]

[1]Northwestern University, [2]University of Cologne, [3]University of Osnabrück
timo.b.roettger@gmail.com, michael.franke@gmail.com, jennifer.cole1@northwestern.edu

ABSTRACT

Listeners can rapidly integrate intonational information to infer a speaker's intended meaning. But not all components of an intonation contour contribute to meaning equally well. Prenuclear pitch accents, tonal events preceding the nuclear pitch accent in an utterance, have been described as not reliably mapping onto discourse meaning. We use mouse tracking to investigate whether German listeners can use prenuclear pitch accents to predict upcoming referential information in the utterance. Our results are compatible with the assumption that listeners ignore prenuclear accents when predicting speakers' intentions. All materials, data, and scripts can be retrieved here: https://osf.io/xf8be/.

Keywords: intonation, prosody, predictive processing, mouse tracking

1. INTRODUCTION

The messages that language users intend to convey do not only depend on what speakers say, but also *how* they say them. To understand communicated meaning infer beyond literal content, listeners must evaluate and simultaneously integrate information associated with the rich context and the temporally extended speech signal. What parts of the speech signal listeners use to infer a speaker's intention and how they integrate this information in real-time is, however, only poorly understood. The present paper contributes to this understanding by investigating the real-time processing of different types of intonational patterns.

Positional asymmetries in intonation contours

Intonation refers to the modulation of the fundamental frequency that signals post-lexical meaning. In many languages, intonation systematically expresses important communicative functions such as illocutionary force and information structure [e.g. 17]. For example, in West-Germanic languages like German or English, the position and form of a f0 movement can signal a referent as discourse-given or discourse-new [e.g. 22].

Traditionally, a special functional status is assigned to the last pitch accent in a phrase and the following boundary tone (i.e. the nuclear contour, e.g. 8). This focus on the nuclear contour is partly due to the belief that parts of the intonation signal preceding the nuclear contour, i.e. the prenuclear contour, is not relevant for expressing discourse functions. In line with this belief, prenuclear accents in English have been described as optional and variable in production [7]; as placed for rhythmic purposes only [6]; They have lesser acoustic prominence and are less likely to be identified as prominent than nuclear accents [11].

These findings suggest a lesser role for prenuclear accents in conveying discourse meaning. Yet this prediction is at odds with certain findings from production experiments. For instance, in English, systematic differences in the prenuclear region are associated with the distinction between broad focus and narrow object focus [e.g. 1, 4]. In German, prenuclear accents can be used to mark contrastive topics [3]. There is also a small body of evidence that listeners can attend to early intonational cues to distinguish questions from statements [e.g. 20, 21].

The studies cited above show that intonational cues in the prenuclear region may be weakly associated with discourse meaning. Yet there remain many questions about the extent to which listeners rely on early intonational cues for comprehending that meaning. Notably, previous findings are mostly based on speech production studies or offline perception / rating tasks, leaving open the question of whether prenuclear intonational cues play a role in the real-time processing of utterance meaning. The present study addresses this question by investigating the predictive use of informative prenuclear accent distinctions on the interpretation of upcoming referring expressions.

Rational processing of intonation

It has been established that listeners can use intonational cues to anticipate a likely speaker-intended referent even before encountering disambiguating lexical materials [e.g. 9, 25, 30, 31]. Moreover, listeners rapidly adapt their predictive cue interpretation based on recent exposure [16, 26, 27]. For example, Roettger and Franke [26, 27] investigated whether German listeners can anticipate referential intentions (i.e. whether the upcoming referent is discourse-given or contrastive) based on either the presence or the absence of an early pitch accent on the auxiliary verb. After hearing a polar question as in (2a), listeners heard either an answer like (2b) with a pitch accent

on the auxiliary verb, confirming the proposition under discussion and thus referring to the given referent (here 'pear') or an answer like (2c) with a pitch accent on the referent, introducing a contrastive referent (here 'violin').

(2a) Hat der Wuggy dann die Birne aufgesammelt?
'Did the wuggy then pick up the pear?'

(2b) Der Wuggy HAT dann die Birne aufgesammelt.
'The wuggy DID then pick up the pear.'

(2c) Der Wuggy hat dann die GEIGE aufgesammelt.
'The wuggy then picked up the VIOLIN.'

Using mouse tracking [e.g. 18, 29], Roettger and Franke showed that listeners exploit the early nuclear pitch accent on the verb in (2b) to anticipate the given referent. Listeners also rapidly learned to use the absence of this pitch accent on the verb in (2c) to anticipated the contrastive referent, long before the contrastive pitch accent on the object was available.

The authors argue that these results are compatible with the idea that comprehenders rationally adapt their expectations about otherwise unreliable intonational information in light of confirming evidence. In other words, for the rational comprehender it seems irrelevant what the nature of the cue is. What matters is the reliable co-occurrence of a certain interpretation and an intonational cue (here: that the absence of a pitch accent on the verb reliably co-occurs with a contrastive interpretation).

This framework predicts that listeners should also be able to exploit information about prenuclear accents as communicatively significant iff there is a reliable co-occurrence of prenuclear accent and meaning. But the same rational listener framework also considers listeners' prior knowledge of German to the experiment, in which case we may have predicted that the lack of a reliable association between prenuclear accents and referential meaning would lead listeners in an experimental setting to initially disregard the prenuclear region of the intonation contour. These prior expectations can then be rapidly adjusted based on new experiences, allowing prenuclear form-function mappings to be learned.

2. METHOD

Two of the following three experiments (experiment 2 and 3) were preregistered prior to data collection. The preregistration files can be retrieved with all materials, data, and analysis scripts from osf.io/xf8be/.

2.1. Participants and procedure

90 German listeners participated in the study (38 men, 52 women, mean age = 25.4 (SD = 3.3)). Subjects were seated in front of a Mac mini 2.5 GHz Intel Core i5. They controlled the experiment via a Logitech B100 corded USB Mouse. Cursor acceleration was linearized and cursor speed was slowed down (to 1400 sensitivity) using the CursorSense© application (version 1.32). Slowing down the cursor ensured that motor behavior was recorded in a smooth trajectory as the acoustic signal unfolded.

Subjects learned about a 'wuggy'- a fantasy creature that picks up objects. There were 12 objects to pick up (bee, chicken, diaper, fork, marble, pants, pear, rose, saw, scale, vase, violin).

Each trial exposed subjects to a context screen, shown for 2500ms and providing a specific discourse context in form of a pre-recorded question. Subjects heard either a polar question (2a), which introduced a referent as discourse-given, or a neutral question such as (3). After the context screen, participants saw a response screen with two response alternatives, each depicting the images of an object in the upper left and right corner, respectively. Via mouse click, participants initiated an audio playback of a prerecorded answer to the question, specifying which object was picked up, e.g. the violin or the pear (e.g. 2b-2c). For any answer to (2a), the referent of the answer stands in a specific discourse relation to the referent in the question, as *given* or *contrastive*. On the other hand, for any answer to (3), the object introduces *new* information, and does not stand in a specific discourse relation to the object referent of the question. If the discourse status of the object is reflected in the prosodic patterning of the prenuclear region, the listener may use early prosodic cues to anticipate the object referent. In the absence of any informative cues in the region prior to the lexical object, the listener must rely on the lexical information in the answer to disambiguate referents.

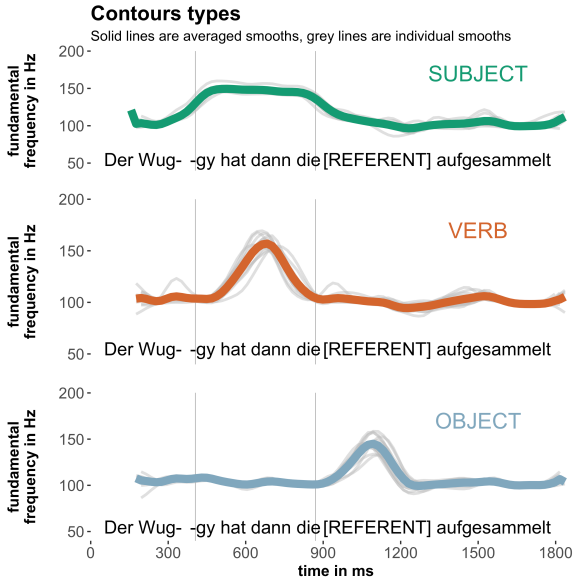
(3a) Was ist passiert?
'What happened?'

Subjects were instructed to choose a response alternative as quickly as possible by moving the mouse cursor over the correct image.

2.2. Stimuli and Materials

Acoustic stimuli were recorded by two trained phoneticians. Statements were produced with three different intonation contours (see Fig. 1). The SUBJECT contour exhibits a rising prenuclear pitch accent on the subject, a high plateau and a falling nuclear accent on the sentence object. The VERB contour exhibits an early nuclear high rising accent on the auxiliary verb "hat". This contour strongly implies a verum focus reading, indicating that the proposition under discussion is true. The OBJECT condition exhibits a high rising accent on the sentence object, commonly used to indicate that the object contrasts with a discourse-salient alternative.

Figure 1: F0 contours used in the experiments.



2.3. Design

In our experiments, these three contours systematically occurred with the referential status of the object. In experiment 1, the VERB contour occurred when the object was referentially (and lexically) given, OBJECT occurred with a contrastive referent for the object (both as expected for German listeners); In experiment 2, VERB occurred with a given object referent, and SUBJECT occurred with contrastive object referents; In experiment 3, SUBJECT occurred with given object referents, and OBJECT occurred with contrastive object referents. In all conditions the OBJECT contours are also used for the LEXICAL disambiguation (as an answer to the question in 3) which served as a control condition in which subjects have to wait for the lexical information in the signal.

Subjects were exposed to 12 blocks of 12 stimuli each. In all experiments, each block contained 4 trials referring to the given referent, 4 trials referring to the contrastive referent, and 4 lexical disambiguation trials.

2.4. Analysis and predictions

The screen coordinates of the computer mouse were sampled at 100 Hz using the mousetrap plugin [14] implemented in the experimental software OpenSesame [19]. Trajectories were processed with the package *mousetrap* [13] using *R* [24]. For each trial, we compute the turn-towards-the-target (TTT, see 26, 27 and scripts) as the latest point in time at which the trajectory did not head towards the target.

We fitted Bayesian hierarchical linear models which predict TTT values by experimental group (E1, E2, E3), discourse relation (lexical, given, contrastive), experimental block (1:12) and their three-way interaction, using the package *brms* [5] in *R*. The models include maximal random-effect structures, allowing the predictors and their interaction to vary by-subjects (discourse relation × block) and by-target referents (discourse relation × group × block). We used weakly informative Gaussian priors centered around 0 (sd = 100) (see osf.io/xf8be/). In the body of this paper, we will report the posterior distributions of relevant predictor levels or differences between levels directly. We report the posterior means alongside their 95% credible intervals (CIs) (henceforth in [square brackets]). A 95% credible interval demarcates the range of values that comprise 95% of probability mass of our posterior beliefs. For practical convenience, we consider evidence as *compelling* when the 95% CI of a difference between predictor levels does not include 0.

If listeners use prenuclear accents to anticipate the discourse status of the upcoming referent, we expect SUBJECT trials to elicit earlier TTTs than OBJECT and VERB trials. If they disregard prenuclear accents, they should be as slow as OBJECT trials. In the latter case, if listeners are rapidly adapting to the reliable co-occurrence of intonational form and meaning [26,27], we expect them to learn that the prenuclear accent is informative and thus TTTs should become faster over the course of the experiment for SUBJECT trials.

3. RESULTS AND DISCUSSION

Figure 2 and Table 1 summarize the results. Looking at the horizontal cursor positions over time, it becomes clear that there are temporal differences between conditions and experimental groups, with some conditions leading to a very early turn towards the target (i.e. y-values increase early during the heard utterance). Across groups, listeners turn towards the target between 132-170ms after the acoustic onset of the noun in the LEXICAL baseline (~872ms).

In E1, there is compelling evidence that VERB trials elicit earlier TTTs than OBJECT trials, which elicit earlier TTTs than LEXICAL trials. These patterns replicate findings by [27] and suggest that listeners use both the presence of an early pitch accent in VERB and the absence of that pitch accent in OBJECT to anticipate the discourse status of the referent. A similar pattern emerges in E2: There is compelling evidence that VERB trials elicit earlier TTTs than SUBJECT trials, which elicit earlier TTTs than LEXICAL trials. Again, listeners use the presence of the early pitch accent in VERB, but surprisingly, they do not use the even earlier prenuclear pitch accent on the subject to the same extent. There is no compelling evidence that OBJECT trials in E1 and SUBJECT trials in E2 elicit different

TTTs ($ß_{diff}$ = 22 [-36,79]), suggesting that listeners predictive behavior is similar across these conditions.

Figure 2: Horizontal cursor position of space-normalized averaged trajectories for experiment 1-3. Semitransparent lines are averaged trajectories for individual participants. Grey vertical lines indicate temporal landmarks.

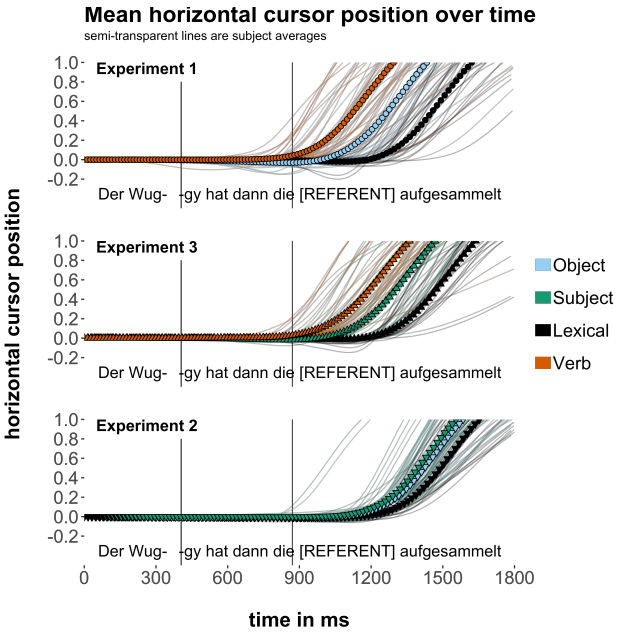


Table 1: Estimated time lag between onset of the auditory stimulus and the turn towards the target in ms. Posterior means and 95% credible intervals for conditions across experiments.

Exp. 1	**Lexical**	>	**Object**	>	**Verb**
	1021 (970;1071)		856 (807;901)		716 (662;766)
Exp. 2	**Lexical**	>	**Subject**	>	**Verb**
	1042 (995;1092)		878 (834;921)		762 (717;809)
Exp. 3	**Lexical**	>	**Object**	=	**Subject**
	1004 (953;1059)		939 (890;989)		926 (877;981)

In E3, there is compelling evidence that OBJECT trials have lower TTTs than LEXICAL trials, but the data suggest that they are slower than its counterpart in E1 ($ß_{E3\text{-}E1}$ = 83[23,150]) A similar slowing down pattern can be found for SUBJECT ($ß_{E3\text{-}E2}$ = 49, [-16, 110]), which however, is not credibly different from 0. These patterns suggest that if listeners do not have a comparison to the early nuclear pitch accent in VERB, their prediction of the upcoming referent becomes poorer.

Except for the OBJECT trials in E1 ($ß_{slope}$ = -19[-38, -1]), there is no compelling evidence that any of these patterns changes over the course of the experiment. The negative slope in OBJECT trials in E1 replicates [26]'s findings and suggests that listeners learn to use the otherwise uninformative absence of a pitch accent on the auxiliary verb over the course of the experiment.

To sum up, in line with previous studies, we can confidently say that listeners use an early nuclear pitch accent on the VERB to anticipate the discourse status of the referent. We can say that listeners alos use something in the signal in the OBJECT trials to anticipate the referent. We interpreted this as the absence of the pitch accent on the verb (in line with [26], [27]). We can say that the very early prenuclear accent on the subject is not systematically used to anticipate the referent. In fact, when the nuclear pitch accent on the verb is not available for comparison in E3, the predictive advantage of both OBJECT and SUBJECT trials decreases.

To sum up, it appears as if listeners mainly attend to what happens on the verb. A (nuclear) pitch accent leads to anticipation of the given referent, no pitch accent leads to anticipation of the contrastive referent. The latter inference, however, is only made in direct comparison to the VERB contour.

5. GENERAL DISCUSSION

The present paper demonstrates that listeners ignore the early component of a complex pitch contour when predicting the referential status of upcoming expressions. At first sight, our results are not compatible with the idea that comprehenders rationally exploit informative intonational cues to predict speaker intentions. In our experiments the prenuclear pitch accent is consistently matched with an upcoming referential interpretation. Nevertheless, neither do listeners use this cue initially, nor do they learn to use it over the course of the experiment.

This is in line with a diverse body of research suggesting that prenuclear pitch accents must play a different role in communication than nuclear pitch accents. This is also in line with evidence from a recent artificial language learning experiment [12] which suggests that prenuclear parts of the intonation contour are ignored by older children and adults, but not by younger children. [12] suggest that younger children payed attention to the holistic contour and older children had learned already that in English there is a strong positional asymmetry in intonation contours (in line with work from object recognition in vision [28] and speech sound perception [23]).

In light of the variable nature of prenuclear parts of intonation contours, we speculate that listeners selectively disregard early prenuclear information in the intonation contour. Listeners appear to allocate attentional resources to those aspects of the speech signal that they expect to be most informative for communication.

7. REFERENCES

[1] Bishop, J. 2017. Focus projection and prenuclear accents: Evidence from lexical processing. *Language, Cognition and Neuroscience*, 32(2), 236-253.

[2] Boersma, P. & Weenink, D. 2016. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.17. http://www.praat.org/.

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

[4] Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. 2010. Acoustic correlates of information structure. *Language and cognitive processes*, 25(7-9), 1044-1098.

[5] Bürkner, P.-C. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28.

[6] Calhoun, S. 2010. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, 86, 1–42.

[7] Chodroff, E., & Cole, J. 2018. Information Structure, Affect, and Prenuclear Prominence in American English. *Proc. Interspeech*, 1848-1852.

[8] CursorSense 2016. (computer software, version 1.3.2). Plentycom Systems. Retrieved from http://plentycom.jp/en/cursorsense/download.php

[9] Dahan, D., Tanenhaus, M.K., & Chambers, C.G. 2002. Accent and reference resolution in spoken-language comprehension. *JML*, 47(2), 292-314.

[10] Halliday, M.A.K (1967) *Intonation and Grammar in British English*. Den Haag.

[11] Hualde, J.I., Cole, J., Smith, C.L., Eager, C.D., Mahrt, T. Napoleão de Souza, R. 2016. The perception of phrasal prominence in English, Spanish and French conversational speech. *Proc. 8th Speech Prosody*, 459–463.

[12] Kapatsinski, V., Olejarczuk, P., & Redford, M.A. 2017. Perceptual Learning of Intonation Contour Categories in Adults and 9-to 11-Year-Old Children: Adults Are More Narrow-Minded. *Cognitive science*, 41(2), 383–415.

[13] Kieslich, P.J., & Henninger, F. 2016. *Mousetrap: Mouse-tracking plugins for OpenSesame* (Version 1.2.1).

[14] Kieslich, P.J., Wulff, D.U., Henninger, F., and Haslbeck, J.M.B. 2017. *mousetrap: Process and Analyze Mouse-Tracking Data*. R package version 3.0.0. https://CRAN.R-project.org/package=mousetrap

[15] Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M.K. 2014a. Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, 133(2), 335–342.

[16] Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. 2014b. Rapid adaptation in online pragmatic interpretation of contrastive prosody. *Proc. 36th CogSci, Quebec City.*

[17] Ladd, D.R. 2008. *Intonational phonology*. Cambridge University Press.

[18] Magnuson, J.S. 2005. Moving hand reveals dynamics of thought. *PNAS*, 102(29), 9995-9996.

[19] Mathôt, S., Schreij, D., & Theeuwes, J. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, 44(2), 314–324.

[20] Petrone, C., & D'Imperio, M. 2011. From tones to tunes: Effects of the f0 prenuclear region in the perception of Neapolitan statements and questions. In S. Frota, G. Elordieta, & P. Prieto (Eds.), *Prosodic categories: Production, perception and comprehension* (pp. 207-230). Springer, Dordrecht.

[21] Petrone, C., & Niebuhr, O. 2014. On the intonation of German intonation questions: The role of the prenuclear region. *Language and Speech*, 57(1), 108-146.

[22] Pierrehumbert, J., & Hirschberg, J.B. 1990. The meaning of intonational contours in the interpretation of discourse. *Intentions in communication*, 271-311.

[23] Pisoni, D.B., Lively, S.E., & Logan, J.S. 1994. Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In J.C. Goodman & H.C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121–166). Cambridge, MA: MIT Press.

[24] R Core Team 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[25] Roettger, T.B. & Stoeber, M. 2017. Manual Response Dynamics Reflect Rapid Integration of Intonational Information during Reference Resolution. *Proc. 39th CogSci*. London.

[26] Roettger, T.B., & Franke, M. 2018a. Dynamic speech adaptation to unreliable cues during intonational processing. *Proc. 40th CogSci,* Madison.

[27] Roettger, T.B., & Franke, M. 2018b. Evidential strength of intonational cues and rational adaptation to (un-)reliable intonation. Unpublished manuscript at PsyArXiv: https://psyarxiv.com/awp87

[28] Smith, L.B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, 96(1),125–144

[29] Spivey, M.J., Grosjean, M., & Knoblich, G. 2005. Continuous attraction toward phonological competitors. *PNAS*, 102, 10393–10398.

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

[31] 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, 367–392.