Proc. Speech Science Technology (SST 2016), Sydney, Australia, 185-188

Pitch accent type affects lexical activation in German: Evidence from eye tracking

Katharina Zahner, Muna Schönhuber, Janet Grijzenhout, Bettina Braun

Department of Linguistics, University of Konstanz

{katharina.zahner|muna.schoenhuber|janet.grijzenhout|bettina.braun}@uni-konstanz.de

Abstract

This visual world eye tracking study tests how f0 affects stress perception in online speech comprehension. The screen showed segmentally overlapping cohort pairs with different stress patterns (WSW/SWW) together with two distractors. In experimental trials, auditory stimuli referred to the WSW cohort member, which was presented with a medial-peak (L+H* L-%) or an early-peak pitch accent (H+L* L-%). Prior to segmental disambiguation, participants fixated the SWW stress competitor more when the WSW target was presented with an early-peak accent. Hence, the peak position affects lexical activation, such that pitch peaks preceding stressed syllables in WSW words temporarily activate SWW words.

Index Terms: eye tracking, stress, pitch accent type, German

1. Introduction

In languages with free stress (e.g., English, German or Dutch, cf. [1]), lexical stress distinguishes between lexical candidates (e.g., English *permit* vs. *permit*; underlining indicates lexical stress). Even when there is no exact stress-minimal pair, lexical stress reduces the number of lexical candidates [2, 3] and strongly guides lexical activation [4-7]. Cross-modal priming experiments with word fragments (cf. [5-7]) indicated that listeners are faster in recognizing words when the prime matched the stress pattern of the target (e.g., "octo" for "octopus/October", see [5]) than when it mismatched. Visual-world eye tracking studies further showed that suprasegmental information immediately modulates word recognition (cf. [4]): Dutch listeners use duration and intensity to resolve lexical competition prior to segmental disambiguation [4].

Lexical stress is an abstract notion at the word-level, which is acoustically cued by duration, intensity, and spectral energy. In German, stressed syllables are longer [8] and louder [8, 9] than unstressed ones, produced with increased vocal effort [10] and more peripheral vowel quality [11]. When accented, stressed syllables are additionally associated with a pitch movement which varies in regard to its position to the stressed syllable [12]. Many early studies have considered pitch a cue to lexical stress, with stressed syllables showing a higher f0 than unstressed ones (cf. [13, 14]). Yet, pitch movements are induced by phrase-level intonation, hence operating on a different linguistic level. Here, we investigate whether different pitch accent types (part of phrase-level phonology) affect the processing of metrical stress (word-level phonology) by studying lexical activation in German.

Theoretically speaking, it may seem odd to expect that intonation could have any effect on lexical access in an intonation language (in which pitch is not lexically contrastive); yet, previous research demonstrated that lexical processing is indeed affected by intonation. For instance, an unfamiliar intonation contour leads to slower lexical access in Dutch [15], listeners rely on the pitch contour to decide between lexical candidates in Italian, English and French [16-18] and German listeners are slower in processing isolated SW words whose f0-contours are taken from WS words [19].

More importantly, recent studies on German showed that phrase-level intonation, i.e. pitch accent type, influences infants' and adults' perception of metrical stress. (Note that in spoken communication, the choice of accent type is influenced by pragmatic considerations: While early-peak accents signal information that is accessible or predictable for listeners [20, 21], medial-peak accents are associated with information that is newly introduced into the discourse [21].) Specifically, [22] showed that German nine-month-olds rely on the position of the pitch peak as a cue to stress, such that only high-pitched stressed syllables are used for the extraction of trochees from fluent speech. High pitch even outweighs other stress cues [23]: Infants treat high-pitched unstressed syllables erroneously as stressed, taking them as the strong (stressed) element when segmenting speech. Furthermore, in an offline-paradigm, German adults have been demonstrated to make more errors in identifying the stressed syllable when the word is produced with a pitch accent that renders the stressed syllable low-pitched [24]. However, these studies do not answer the question whether and how pitch information influences German adults in online word recognition.

Taken together, we predict that phrase-level intonation modulates lexical access in online speech comprehension. More specifically, high-pitched but unstressed syllables in WSW words are expected to be perceived as stressed, leading to the temporary activation of cohort competitors with an "opposite" stress pattern, i.e. a SWW word-prosodic structure.

2. Experiment

We investigate how pitch accent type affects lexical activation in German adults by using eye tracking (visual world) with four printed words on screen [25], a paradigm that is sensitive to phonetic/phonological differences. We particularly focus on two kinds of rising-falling accents that differ in the alignment of the tonal movement with the stressed syllable: In early-peak accents, the pitch peak precedes the stressed syllable (H+L*), while in medial-peak accents (L+H*) the stressed syllable and the pitch peak coincide. Similar to [4], the screen showed two written representations of trisyllabic cohort competitors that differ in the position of stress (e.g., WSW "Libelle" 'dragonfly' and SWW "Libero" 'sweeper'). We tested whether high-pitched unstressed syllables in WSW words, i.e., syllables on which a pitch peak is realized, but which do not carry lexical stress (as in early-peak accents, H+L*), prompt adults to perceive them as stressed, thus activating the cohort member with initial stress (SWW).

2.1. Methods

2.1.1. Participants

Forty-eight German native speakers (38 female, ∅ = 23.9 years, SD = 3.1 years) with normal or corrected-to-normal vision and unimpaired hearing participated for a small fee.

2.1.2. Materials

Sixty-four trisyllabic cohort pairs that differed in the position of stress (WSW vs. SWW words) were selected. The cohort pairs were segmentally identical up to at least the onset consonant of the second syllable, e.g., "Libelle" [li'bɛlə] vs. "Libero" ['libəʁo]. Cohort members were matched for lexical frequency and number of characters across groups. For each cohort pair, we selected two semantically and phonologically unrelated distractors with comparable number of characters and frequency to be presented on screen. Distractors were SWW-, WSW- or WWS words (one third in each pattern). For 32 of the 64 cohort pairs, the auditory target was one of the cohort members (in 16 experimental trials the WSW word, in 16 distractor trials the SWW word), in the other 32 trials the auditory target referred to one of the unrelated words (filler trials). For instance, in experimental trials, the WSW cohort member "Libelle" was presented as auditory target and the SWW cohort member "Libero" was the stress competitor.

The auditory targets were embedded in a semantically non-constraining carrier sentence (e.g., "Bitte klicke Libelle an", 'Please click on dragonfly'). A female native speaker of Standard German recorded the sentences in a sound-attenuated cabin in the PhonLab at the University of Konstanz (44.1kHz, 16Bit). The auditory targets for experimental (WSW as auditory target) and distractor trials (SWW as auditory target) were produced in two intonation conditions: with a medial-peak (L+H*) and an early-peak accent (H+L*, see Figure 1). The recordings in the early- and medial-peak condition were matched along a number of acoustic variables; see Table 1 for measurements in experimental trials. Similarly, half of the auditory targets for the fillers were recorded with a medial-peak, half with an early-peak accent. To reduce distal prosodic context effects [26], stimuli were spliced into a carrier sentence "Bitte klicke". Four different carriers (same across conditions) were used for targets starting with different vowels ([a] vs. [e]), the consonant [m] or any other consonant to avoid co-articulation. Thus, the carrier was identical for each cohort pair in both intonation conditions. The cross-spliced stimuli were rated to sound natural and splicing was not noticeable. Words in the carrier were not accented (see Figure 1) to avoid metrical expectations based on the preceding f0-contour [26].

2.1.3. Procedure

Participants were tested individually in the PhonLab at the University of Konstanz. They were seated approximately 70cm in front of a LCD screen (37.5cm x 30cm). The Desktop Mount was used with a head support. The dominant eye was calibrated (pupil and corneal reflection) in an automatic procedure, using the SR Eyelink 1000 Plus tracking system at a sampling rate of 250Hz.

The experiment consisted of 64 trials, 16 experimental trials (WSW cohort member as auditory target), 16 distractor trials (SWW cohort member as auditory target), and all 32 filler trials (unrelated distractor as auditory target). In experimental and distractor trials, intonation condition (medial- vs. early-peak) was distributed in a Latin-Square Design, i.e., each subject heard both intonation conditions, but each item in only one of the conditions. Eight experimental lists were created, pseudo-randomizing the order of trials such that each half contained the same number of cohort and distractor trials, with the constraint of the experimental item (WSW) being at most the third item of the same condition in a row. Each list started with five filler and two distractor trials to familiarize participants with the task and voice. Participants were assigned randomly to one of the experimental lists (six participants per list, eight items in each condition).

Figure 1: *F0-contour of two experimental trials (medial-peak (dashed); early-peak condition (solid)).*

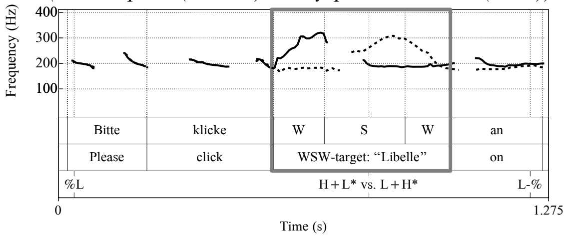


Table 1. *Acoustic realization means (and standard deviations) of WSW targets in two intonation conditions (experimental trials).*

Acoustic variable	Medial-Peak Condition	Early-Peak Condition
F0-excursion of accentual movement in st	Rise: 8.36 (0.60)	Fall: 8.43 (0.67)
Slope of accentual movement in Hz/sec	51.6 (11.7)	54.8 (9.7)
Duration of first syllable in ms	143 (34)	146 (34)
Duration of second syllable in ms	214 (48)	226 (48)
Duration of third syllable in ms	164 (34)	157 (35)
H1*-A3* ratio ([10]) in middle of first vowel in dB	27.0 (10.8)	23.2 (9.7)
H1*-A3* ratio in middle of second vowel in dB	30.8 (7.2)	31.8 (6.3)
H1*-A3* ratio in middle of third vowel in dB	27.7 (5.5)	28.7 (4.5)

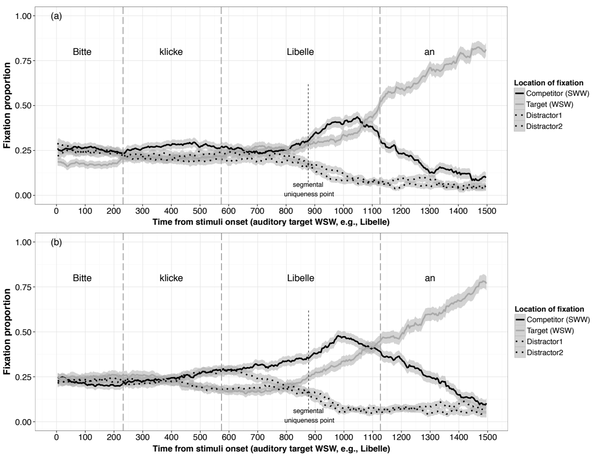
Every trial started with a fixation cross that was centered on screen and displayed until participants clicked on it. Upon clicking, the four words appeared on screen (Times New Roman Font, size 20). They were presented in the outer third of the four quadrants of the screen (to avoid peripheral looking) and framed by a rectangular box. The position of the cohort members and distractors were randomized on screen, such that in all lists the auditory target occurred equally frequent in all four possible positions. The carrier phrase started 2000ms after the words occurred on screen, leaving a preview of 1426ms on average. Participants were instructed to click on the word named in the auditory stimulus as fast as possible. Auditory stimuli were presented via headphones (Beyerdynamic DT-990 Pro, 250 OHM) at comfortable loudness. Every fifth trial, a drift correction was initiated. After half of the trials, there was an optional pause. The total duration of the experiment was approximately 15 minutes.

2.2. Results

For reasons of space, we only report on experimental trials here. In these trials, participants clicked on the auditory target (WSW) in 98.3% of all trials; there was no effect of intonation condition on error rates or reaction times (both $p > 0.3$).

The eye tracking data were extracted in 4ms steps and coded into saccades, fixations, and blinks (default settings for normal saccade sensitivity in the EyeLink 1000 software); only fixations were further processed. For experimental trials, fixations were automatically labeled as being directed to the target (e.g., "Libelle"), the stress competitor (e.g., "Libero"), or to the distractors if they fell within a square of 200x200 pixels around the respective word. Figure 2 shows the evolution of fixations in experimental trials to the words on screen in the two intonation conditions. The dashed vertical lines indicate the acoustic landmarks of the auditory stimuli.

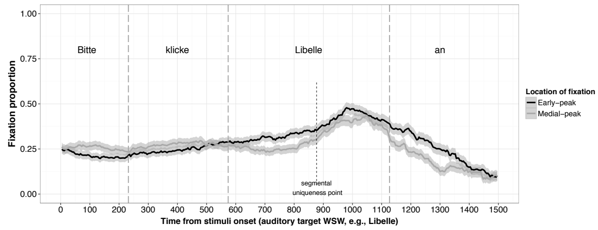
Figure 2: *Evolution of fixations to competitor, target and the two distractors in the medial-peak condition (a) and the early-peak condition (b).*



For statistical analysis of the SWW competitor activation in experimental trials, empirical logits of fixations to the competitor were calculated (ratio of the fixations to the competitor divided by the fixations directed to the three other objects or somewhere else [27]). We fitted a linear mixed effects regression model with *intonation condition* (medial-peak vs. early-peak) as fixed factor and *participants* and *items* as crossed random factors [28], allowing for adjustments of intercepts and slopes [29]. P-values were obtained using the Satterthwaite's approximation (R package lmerTest [30]).

Results showed an effect of *intonation condition* on the fixations to the stress competitor (SWW) in the window in which suprasegmental information differed across conditions but segmental information did not distinguish the lexical candidates. During segmental ambiguity, participants fixated the SWW stress competitor more often when the WSW target was realized with an early-peak (average elogs -1.14), compared to a medial-peak accent (average elogs -1.67, $ß = 0.5$ [0.02;1.04], $SE = 0.26$, $t = 2.05$, $p = 0.04$). This difference in fixations to the SWW competitor is illustrated more directly in Figure 3. Figures 2 and 3 further suggest differences in fixations to the stress competitor already before information on the auditory target became available. This difference is not significant, however ($p = 0.24$). Fixations to the target also differed as a function of intonation condition: There were significantly more fixations to the WSW target in the medial-peak, compared to the early-peak condition, but only during the processing of the early part of the first syllable (720ms-800ms; $ß = 0.5$ [0.05; 0.93], $SE = 0.22$, $t = 2.22$, $p = 0.03$).

Figure 3: *Fixations to SWW stress competitor in experimental trials (WSW auditory target) in the two intonation conditions.*



2.3. Discussion

Our results reveal intonational interference during lexical activation: There were more fixations to the stress competitor (SWW) – and consequently fewer fixations to the target (WSW) – when the WSW target was presented with an early-peak accent (H+L*) than when presented with a medial-peak accent (L+H*). This shows that high-pitched unstressed syllables temporarily activate competitors with initial stress.

Note that both medial-peak (L+H*) and early-peak accent (H+L*) naturally occur in German. The carrier sentence used in the experiment might favor a medial-peak accent, since the objects to be clicked on were new referents [21]. Yet, the repeated mention of the same carrier evokes a notion of accessibility of the objects as a whole, which makes an early-peak realization equally pragmatically appropriate [20].

3. General Discussion and Conclusion

Overall, we find that pitch accent type influences lexical activation: German adults inadvertently activate words with the wrong stress pattern (SWW) when presented with WSW words in which the pitch peak precedes the stressed syllable. During segmental ambiguity, high-pitched unstressed syllables are more often interpreted as stressed than low-pitched unstressed syllables. Note that in both intonation conditions, the prosodic stress cues suggest that the second syllable is stressed. The cue that varies across intonation conditions is the position of the pitch peak. High pitch hence seems to be a relevant cue to metrical stress for German adults. Even though theoretically different from word-level stress, pitch accent type affects the processing of metrical stress. Here, we demonstrate that pitch information influences German listeners in online word recognition, not only in offline metalinguistic stress judgments [24]. We further show that the association between high pitch and metrical stress observed in German infants [22, 23] is also found in German adults.

How can we explain that pitch accent type influences lexical processing in that way? It is striking that adults interpret high-pitched unstressed syllables erroneously as stressed, since due to phrase-level intonation this strategy is not profitable in all cases. In spoken communication, using f0 as a cue to stress might even lead to higher processing costs, as there is more lexical competition. We see three explanations why high-pitched syllables play a role in the online processing of metrical stress: First, high-pitched syllables are perceptually salient [31]. Listeners might thus equate perceived acoustical prominence with metrical prominence. Second, medial-peak are more frequent in German spontaneous speech than early-peak accents [32]. Conceivably, the frequent encounter of high-pitched stressed syllables might make the pitch peak a cue to stress. Finally, the findings in the framework of the iambic-trochaic law might shed light on the observed pattern of results: [33], for instance, showed that adults tend to group pairs of syllables alternating in pitch with initial prominence (SW). Future research with other pitch accent types and/or other languages is needed to better understand the mechanism.

Here, we show that high pitch is a cue to stress in German adults' online processing, despite the fact that pitch accent type is part of phrase- and not word-level phonology. Hence, in speech processing f0 is more intertwined with other stress cues than in phonological theory. Consequently, this finding poses questions for spoken word recognition models (e.g., Shortlist [34], Shortlist B [35], FUL [36]) that currently do not account for utterance-level intonation.

4. References

[1] U. Domahs, I. Plag, and R. Carroll, "Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited," *The Journal of Comparative Germanic Linguistics,* vol. 17, pp. 59-96, 2014.

[2] A. Cutler, *Native listening: Language experience and the recognition of spoken words*. Cambridge, Mass. [u.a.]: MIT Press, 2012.

[3] A. Cutler and D. Pasveer, "Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition," *Proceedings of 3rd Internatinal Conference on Speech Prosody*, Dresden, 2006.

[4] E. Reinisch, A. Jesse, and J. M. McQueen, "Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately," *The Quarterly Journal of Experimental Psychology,* vol. 63, pp. 772-783, 2010.

[5] M. Koster, W. van Donselaar, and A. Cutler, "Exploring the role of lexical stress in lexical recognition," *The Quarterly Journal of Experimental Psychology,* vol. 58, pp. 251-274, 2005.

[6] C. K. Friedrich, S. A. Kotz, A. D. Friederici, and T. C. Gunter, "ERPs reflect lexical identification in word fragment priming," *Journal of Cognitive Neuroscience,* vol. 16, pp. 541-552, 2004.

[7] N. Cooper, R. Wales, and A. Cutler, "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," *Language and Speech,* vol. 45, pp. 207-228, 2002.

[8] M. Jessen, K. Marasek, and K. Claßen, "Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German.," *Proceedings of the 13th International Congress of the Phonetic Sciences,* Stockholm, 1995.

[9] G. Dogil, "Phonetic correlates of word stress," *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (Univ. Stuttgart),* vol. 2, pp. 1-60, 1995.

[10] C. Mooshammer, "Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German," *Journal of Acoustical Society of America,* vol. 127, pp. 1047-1058, 2010.

[11] P. Delattre, "An acoustic and articulatory study of vowel reduction in four languages," *International Review of Applied Linguistics and Language Teaching (IRAL),* vol. 7, pp. 294-325, 1969.

[12] J. B. Pierrehumbert, "The Phonology and Phonetics of English intonation," Massachusetts Institute of Technology, Dept. of Linguistics and Philosophy, Bloomington, 1980.

[13] D. B. Fry, "Experiments in the perception of stress," *Language and Speech,* vol. 1, p. 126, 1958.

[14] I. Lehiste, *Suprasegmentals*. Cambridge, Mass. [u.a.]: MIT Press, 1970.

[15] B. Braun, A. Dainora, and M. Ernestus, "An unfamiliar intonation contour slows down online speech comprehension," *Language and Cognitive Processes,* vol. 26, pp. 350-375, 2011.

[16] M. D'Imperio, C. Petrone, and N. Nguyen, "Effects of tonal alignment on lexical identification in Italian," in *Tones and Tunes*, C. Gussenhoven and T. Riad, Eds., Berlin: Mouton de Gruyter, 2007, pp. 79-106.

[17] D. R. Ladd and A. Schepman, ""Sagging transitions" between high pitch accents in English: Experimental evidence," *Journal of Phonetics,* vol. 31, pp. 81-112, 2003.

[18] P. Welby, "The role of early fundamental frequency rises and elbows in French word segmentation," *Speech Communication,* vol. 49, pp. 28-48, 2007.

[19] C. K. Friedrich, K. Alter, and S. A. Kotz, "An electrophysiological response to different pitch contours in words," *Neuroreport,* vol. 12, pp. 3189-3191, 2001.

[20] S. Baumann and M. Grice, "The intonation of accessibility," *Journal of Pragmatics,* vol. 38, pp. 1636-1657, 2006.

[21] K. Kohler, "Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics," *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK),* vol. 25, pp. 115-185, 1991.

[22] K. Zahner, M. Schönhuber, and B. Braun, "The limits of metrical segmentation: Intonation modulates infants' extraction of embedded trochees," *Journal of Child Language,* pp. 1-27, 2015.

[23] K. Zahner, M. Schönhuber, J. Grijzenhout, and B. Braun, "High pitch signals word onsets for German 9-month-olds: Evidence for a pitch-based segmentation strategy," Presented at *Tone and Intonation in Europe (TIE)*, Canterbury, UK, 2016.

[24] S. Egger, "The impact of pitch accents on the identification of word stress in German," MA, Department of Linguistics, University of Konstanz, Konstanz, 2015.

[25] J. M. McQueen and M. Viebahn, "Tracking recognition of spoken words by tracking looks to printed words," *The Quarterly Journal of Experimental Psychology,* vol. 60, pp. 661-671, 2007.

[26] M. Brown, A. P. Salverda, L. C. Dilley, and M. K. Tanenhaus, "Metrical expectations from preceding prosody influence perception of lexical stress," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 41, pp. 306-323, 2015.

[27] D. J. Barr, T. M. Gann, and R. S. Pierce, "Anticipatory baseline effects and information integration in visual world studies," *Acta Psychologica,* vol. 137, pp. 201-207, 2011.

[28] R. H. Baayen, *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge [u.a.]: Cambridge Univ. Press, 2008.

[29] D. J. Barr, R. Levy, C. Scheepers, and H. Tily, "Random-effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of Memory and Language,* vol. 36, pp. 255-278, 2013.

[30] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen. (2013). *lmerTest: Tests for random and fixed effects for linear mixed effect models*. Available: http://cran.r-project.org/web/packages/lmerTest/index.html.

[31] S. Baumann and C. Röhr, "The perceptual prominence of pitch accent types in German," *Proceedings of the 18th International Congress of the Phonetic Sciences*, Glasgow, UK, 2015.

[32] B. Peters, K. Kohler, and T. Wesener, "Melodische Satzakzentmuster in prosodischen Phrasen deutscher Spontansprache - Statistische Verteilung und sprachliche Funktion," in *Prosodic Structures in German Spontaneous Speech (AIPUK 35a)*, K. Kohler, F. Kleber, and B. Peters, Eds., Kiel: IPDS, 2005, pp. 185-201.

[33] R. A. H. Bion, S. Benavides-Varela, and M. Nespor, "Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences," *Language and Speech,* vol. 54, pp. 123-140, 2011.

[34] D. Norris, "Shortlist: A connectionist model of continuous speech recognition," *Cognition,* vol. 52, pp. 189-234, 1994.

[35] D. Norris and J. M. McQueen, "Shortlist B: A Bayesian model of continuous speech recognition," *Psychological Review,* vol. 115, pp. 357-395, 2008.

[36] A. Lahiri and H. Reetz, "Underspecified recognition," in *Labphon*. vol. 7, C. Gussenhoven, N. Werner, and T. Rietveld, Eds., Berlin: Mouton, 2002, pp. 637-676.