The contribution of speech-rhythm to end-of-utterance detection

Franziska Schaller
Experimental Neurolinguistics Group
Faculty of Linguistics and Literary Studies
Bielefeld University
P.O. Box 100131
33501 Bielefeld, Germany
franziska.schaller@uni-bielefeld.de

Horst M. Müller
Experimental Neurolinguistics Group
Faculty of Linguistics and Literary Studies
Bielefeld University
P.O. Box 100131
33501 Bielefeld, Germany
horst.mueller@uni-bielefeld.de

Abstract

In the current study we focused on the question of how language specific properties other than syntax and semantics – especially rhythm – affect end-of-utterance detection. We compared the anticipation timing accuracy for German stimuli (mother-tongue) to the timing accuracy for foreign language items and sinusoidal tones. Subjects were more accurate when anticipating the ends of utterances in stress-timed than in syllable-timed languages or tones. We interpret these results as an indicator for rhythm being relevant in end-of-utterance anticipation.

1 Introduction

Interlocutors show accurate timing in conversation. This has already been suggested by Sacks et al. (1974), who developed an established turn-taking system. From a projectionists' point of view, recipients anticipate when a speaker's turn ends, which permits very precise speaker changes. For this anticipation process, lexico-syntactic characteristics seem to be particularly relevant (e.g. Beattie, 1981; de Ruiter et al., 2006; Magyari et al., 2011; Müller et al., 2013). Other studies suggest that prosodic and suprasegmental characteristics are important (e.g. Gravano and Hirschberg, 2009; Heldner et al., 2009; Wells and Macfarlane, 1998). Also, it is discussed whether successful turn-taking processes depend on the interplay of several aspects – including e.g. semantics, syntax, prosody, and rhythm (e.g. Ford and Thompson, 1996; Selting, 1996).

As for rhythm, Wagner et al. (in press) suggest that there is a strong connection between prosodic characteristics of an utterance and speech rhythm, insofar as the prosodic features follow a regular oscillation pattern. Interlocutors entrain their speech rhythm according to this pattern (Inden et al., in press) and could use it to adjust their turn onsets (Auer et al., 1999; Couper-Kuhlen, 1993; Wilson and Wilson, 2005). However, there is no empirical evidence for this hypothesis so far (Inden et al., in press).

The intention of the current study was to assess whether speech rhythm and general articulatory speech-specific features other than syntax and semantics are relevant for end-of-utterance anticipation. For that purpose, we surveyed how well participants were able to anticipate the ends of utterances in different languages and measured their anticipation timing accuracy (ATA) as an indicator of conscious behavioral processes.

2 Material and method

2.1 Stimuli

We used spoken sentences (161 total) as stimuli. There were 23 items in each of seven languages (German, English, Italian, Polish, Turkish, Arabic, Korean). Languages other than German (L1) and English (L2) were judged as unknown. Sinusoidal tones (10 total) were used as control items. The tones were generated at 450 Hz and matched the length of the sentences.

As for the unknown language utterances, we expected that participants would have to use speech rhythm and other suprasegmental features for a successful anticipation since they could not rely on semantic and syntactic content. If they do not use other elements besides syntax and semantics in their daily turn-taking, their ATA should not be much better for the linguistic, but incomprehensible stimuli as it is for the maxi-

mally non-linguistic sinusoidal tones that do not contain any linguistic information at all.

2.2 Procedure

The items were presented auditorily (45 to 55 dB) and subjects listened to them with headphones. They were asked to push a button on an external response box at the exact moment the utterance ended. The time span between the actual end of the utterance and the button push was defined as the ATA.

3 Results

In addition to checking for ATA differences between the languages, we also grouped them as a) either Indo-European (IE) or Non-Indo-European (Non-IE) and b) either stress-timed or (rather) syllable-timed. Both the IE and the stress-timed group contained the stimuli in the known languages.

Comparisons of the ATA of the foreign language stimuli and the tones revealed several differences. As expected, subjects reached a better ATA for the ends of German items than for any other stimulus type. Further, they anticipated the ends of tones and of Turkish stimuli equally worse than the ends of all other stimulus types. A repeated measures ANOVA ($F(3.42, 119.52) = 100.27$, $p \leq .001$) and the Bonferroni multiple comparison post-hoc test showed that there were significant differences between the ATA of almost all item types. The comparison of Polish, Turkish and Korean items to tones revealed no significant differences in the ATA.

As well, there was an overall highly significant effect ($F(1.33, 46.65) = 98.35$, $p \leq .001$) when comparing the ATA of IE languages to that of Non-IE languages and of tones. All stimulus groups differed significantly from each other. The ends of IE utterances were most accurately anticipated. Since these results suggest that there must be some suprasegmental elements relevant for end-of-utterance anticipation, we grouped the sentence types according to their stress pattern. The ATA differences between stress-timed and syllable-timed languages and tones were highly significant ($F(1.35, 47.32) = 116.61$, $p \leq .001$). The ends of stress-timed utterances were anticipated significantly better than of syllable-timed items ($p \leq .01$) and of tones ($p \leq .01$).

4 Discussion

The ATA differences between foreign language stimuli and tones were mostly not significant, which implies that anticipation performance was definitely better when subjects had access to semantics and syntax and that suprasegmentals alone were not sufficient for an adequate anticipation performance. Nevertheless, subjects anticipated the ends of the Non-IE utterances better than of tones although they did not have access to syntax and semantics. It is probable that they used language-universal linguistic properties – which we suppose to be suprasegmental in nature – to anticipate the ends of utterances in unknown languages. Possible relevant properties in this context are e.g. the last major accent and specific F_0-contours that have been discussed in a number of corpus studies (Caspers, 2003; Heldner et al., 2009; Koiso et al., 1998; Wells & Macfarlane, 1998). Further, language differs from tones in its speech specific rhythm, which might be relevant in the anticipation of utterance-ends as well (e.g. Beňuš et al., 2011). Our results support this assumption. There was a significant difference between the ATA of stress-timed vs. syllable-timed languages. Thus, participants were more accurate when anticipating the ends of Arabic items, the stress pattern of which is similar to that of German, than when anticipating the ends of e.g. Polish items that differ from German in their stress pattern. Further, there were no significant differences between the ATA of utterances with a syllable-timed rhythm and the ATA of tones. Thus, subjects' anticipation performance was inadequate when they were not able to make use of neither syntax and semantics nor a well-known stress pattern. This implies that rhythm probably is an important feature that people use when anticipating the end of an utterance.

5 Conclusion

We propose that in combination with syntax and semantics, rhythm is a relevant characteristic in the anticipation of utterance ends. As well, there are other suprasegmental characteristics which influence anticipation processes, albeit to a lesser degree than syntax and semantics.

The results of the current study thus support the view that there is a number of features that are all accounted for by conversational partners when anticipating the end of a turn.

Acknowledgments
This study was supported by the DFG (SFB 673). We thank Jan de Ruiter and Petra Wagner for their helpful comments.

References

Augustín Gravano and Julia Hirschberg. 2009. Turn-yielding cues in task-oriented dialogue, in *Proceeding of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 253-261, eds. Patrick G. T. Healey, Roberto Pieraccini, Donna K. Byron, Steve Young, and Matthew Purver. Association for Computational Linguistics, Stroudsburg, PA.

Benjamin Inden, Zofia Malisz, Petra Wagner, and Ipke Wachsmuth. In press. Timing and entrainment of multimodal backchanneling behavior for an embodied conversational agent, in *Proceedings of the 15th International Conference on Multimodal Interaction, ICMI '13*, eds. Julien Epps, Fang Chen, Sharon Oviatt, Kenji Mase, Andrew Sears, Kristina Jokinen, and Björn Schuller. ACM, New York, NY.

Bill Wells and Sarah Macfarlane. 1998. Prosody as an interactional resource. Turn-projection and overlap. *Language and Speech*, 41: 265-294.

Celia E. Ford and Sandra A. Thompson. 1996. Interactional units in conversation. Syntactic, intonational and pragmatic resources for the management of turns, in *Interaction and grammar*, 134-184, eds. Elinor Ochs, Emanuel A. Schegloff, and Sandra A. Thompson. Cambridge University Press, Cambridge, UK.

Elizabeth Couper-Kuhlen. 1993. *English speech rhythm: Form and function in everyday verbal interaction.* Benjamins, Amsterdam.

Geoffrey W. Beattie. 1981. The regulation of speaker turns in face-to-face conversation: Some implications for conversation in sound-only communication channels. *Semiotica*, 34: 55-70.

Hanae Koiso, Yasuo Horiuchi, Syun Tutiya, Akira Ichikawa, and Yasuharu Den. 1998. An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Language and Speech*, 41: 295-321.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50: 696-735.

Horst M. Müller, Stefanie Jansen, and Hendrik Wesselmeier. 2013. Anticipation of speakers' end-of-turn indicated by lateralized readiness potential. *Society for Neuroscience*, 11403.

Janneke Caspers. 2003. Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics*, 31: 251-276.

Jan P. de Ruiter, Holger Mitterer, and Nick J. Enfield. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82: 515-535.

Lilla Magyari, Marcel C. M. Bastiaansen, Jan P. de Ruiter, and Stephen Levinson. 2011. *Neuronal correlates of anticipation related to turn-taking in conversations.* Poster presented at the AMLaP 2011, Sept 1st-3rd, Paris, France.

Margret Selting. 1996. On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics*, 6: 357-388.

Margaret Wilson and Thomas P. Wilson. 2005. An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review*, 12: 957-968.

Mattias Heldner, Jens Edlund, Kornel Laskowski, and Antoine Pelcé. 2009. Prosodic features in the vicinity of silences and overlaps, in *Nordic Prosody: Proceedings of the Xth Conference, Helsinki 2008*, 95-105, eds. Martti Vainio, Reijo Aulanko, and Olli Aaltonen. Lang, Frankfurt.

Peter Auer, Elizabeth Couper-Kuhlen, and Frank Müller. 1996. *Language in time*. Oxford University Press, Oxford.

Petra Wagner, Zofia Malisz, Benjamin Inden, and Ipke Wachsmuth. In press. Interaction phonology – A temporal coordination component enabling representational alignment within a model of communication, in *Advances in interaction studies. Alignment in communication*, 109-132, eds. Ipke Wachsmuth, Jan P. de Ruiter, Petra Jaecks, and Stefan Kopp. Benjamins, Amsterdam.

Štefan Beňuš, Augustín Gravano, and Julia Hirschberg. 2011. Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, 43: 3001-3027.