



## Research Article

# Listeners use intonational phrase boundaries to project turn ends in spoken interaction

Sara Bögels<sup>1</sup>, Francisco Torreira<sup>\*,1</sup>

Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

## ARTICLE INFO

*Article history:*

Received 15 May 2014  
 Received in revised form  
 8 April 2015  
 Accepted 23 April 2015

*Keywords:*

Turn-taking  
 Turn projection  
 Intonation  
 Phrasing  
 Prosody  
 Dialogue  
 Conversation

## ABSTRACT

In conversation, turn transitions between speakers often occur smoothly, usually within a time window of a few hundred milliseconds. It has been argued, on the basis of a button-press experiment [De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3):515–535], that participants in conversation rely mainly on lexico-syntactic information when timing and producing their turns, and that they do not need to make use of intonational cues to achieve smooth transitions and avoid overlaps. In contrast to this view, but in line with previous observational studies, our results from a dialogue task and a button-press task involving questions and answers indicate that the identification of the end of intonational phrases is necessary for smooth turn-taking. In both tasks, participants never responded to questions (i.e., gave an answer or pressed a button to indicate a turn end) at turn-internal points of syntactic completion in the absence of an intonational phrase boundary. Moreover, in the button-press task, they often pressed the button at the same point of syntactic completion when the final word of an intonational phrase was cross-spliced at that location. Furthermore, truncated stimuli ending in a syntactic completion point but lacking an intonational phrase boundary led to significantly delayed button presses. In light of these results, we argue that earlier claims that intonation is not necessary for correct turn-end projection are misguided, and that research on turn-taking should continue to consider intonation as a source of turn-end cues along with other linguistic and communicative phenomena.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Everyday conversation is characterized by a regular exchange of turns between interlocutors. Although considerable variation occurs in the timing between the end of one speaker's turn and the beginning of the next (De Ruiter, Mitterer, & Enfield, 2006; Heldner & Edlund, 2010), the most frequent turn transitions occur with only slight gap or overlap, regardless of the language (Sacks, Schegloff, & Jefferson, 1974; Stivers et al., 2009). Since turn-taking appears to be smooth in many instances, the question arises how interlocutors manage to time their responses to previous turns. One possible answer is that listeners wait for the offset of speech (i.e., silence of a certain duration) to start their own production; that is, they follow a reactive strategy. However, since speech planning takes at least a few hundred milliseconds (600 ms for picture naming; Indefrey & Levelt, 2004; Jescheniak, Schriefers, & Hantsch, 2003; 1500 ms for simple sentences, Griffin & Bock, 2000), and since the minimal response time for spoken utterances (uttering a pre-planned sound) is about 200 ms (Fry, 1975), reaction to silence is unlikely to be the explanation for at least a large proportion of turn transitions. It seems therefore that listeners often project the end of a turn before its occurrence on the basis of information present in the ongoing turn. In the paper that initiated the modern study of turn-taking, Sacks et al. (1974, 722) already pointed out that it is unlikely that lexico-syntactic information alone will be sufficient for this prediction: "When it is further realized that any word can be made into a 'one-word' unit-type, via intonation, then we can appreciate the partial character of the unit-types' description in syntactic terms". Later studies seemed to substantiate that both syntactic and intonational information is used for turn-end projection (Ford & Thompson, 1996; Gravano & Hirschberg, 2011; Local & Walker, 2012; Wells & MacFarlane, 1998). Against this position, and on the basis of an online button-press study, De Ruiter et al. (2006) have claimed that lexico-syntactic information alone may be sufficient for turn-end projection, and that intonation is neither necessary nor sufficient for this task. The present study explicitly

\* Correspondence to: Language and Cognition Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD, Nijmegen, The Netherlands.

E-mail addresses: Sara.Bogels@mpi.nl (S. Bögels), Francisco.Torreira@mpi.nl (F. Torreira).

<sup>1</sup> Equal first authors.

addresses this controversy with a dialogue task and an online button-press experiment in which we control the occurrence of syntactic and intonational phrase boundaries in Dutch polar questions.

Observational studies of English dialogue have identified cues related to turn ends as opposed to turn-medial positions. [Duncan \(1972\)](#) identified several ‘turn-yielding cues’ that speakers appear to use in American English, including specific intonational inflections, body motion, stereotyped expressions, and syntactic completion, and observed that the chance for a speaker change increases with the number of cues encountered in the speaker’s turn. In a conversation-analytic study of over 400 turn transitions in conversations in American English, [Ford and Thompson \(1996\)](#) found that smooth turn transitions only occur at points of simultaneous syntactic, intonational, and pragmatic completeness. More recently, [Gravano and Hirschberg \(2011\)](#) found that both prosodic and lexico-syntactic cues can predict the yielding of the turn after a silence in a game task in American English. [Local and Walker \(2012\)](#) studied all points that were syntactically (and contextually) complete in a telephone call in British English, and focused on non-pitch phonetic cues. Among the identified cues that project the end of the current turn are final lengthening, the absence of segmental reduction, and audible outbreaths. [Koiso, Horiuchi, Tutiya, Ichikawa, and Den \(1998\)](#) found that, in Japanese, both prosodic and syntactic features were related to turn-taking. [Dombrowski and Niebuhr \(2005\)](#) found that phrase-final pitch rises differed in a number of characteristics depending on whether they were produced in a turn-yielding, or a turn-holding context. Turn ends thus appear to be characterized both by completion at the lexico-syntactic and intonational levels. Both kinds of information are therefore potentially available to listeners for projecting turn ends in turn-taking situations. But observational studies cannot provide direct evidence about what information listeners actually use. For instance, listeners might not rely on intonational cues for purposes of turn-taking even when present, and instead focus on lexico-syntactic cues only, or vice versa.

In order to investigate which information listeners use to identify turn ends, several researchers have presented participants with speech fragments, asking them to predict whether the speaker would continue after the fragment or stop speaking. In agreement with the observational studies mentioned above, some of these studies have shown that both syntactic and intonational completion ([Caspers, 1998](#); for Dutch) as well as other prosodic cues ([Hjalmarsson, 2011](#); for Swedish) are used by listeners to identify turn ends in excerpts extracted from recorded conversations. [Geluken and Swerts \(1994\)](#) presented listeners with utterances from a production experiment, and also found that listeners could distinguish between turn boundaries and non-turn boundaries on the basis of prosodic cues. In contrast, [Schaffer \(1983\)](#), using fragments from conversation, concluded that there is significant variability in listener’s use of intonation cues to turn-ends. One should bear in mind, however, that these studies used offline judgments by listeners who were not taking part in the conversation that they judged, and who were not under the timing constraints operating in conversation (i.e., minimization of long gaps, [Sacks et al., 1974](#)). For these reasons, these studies leave open the question whether the same types of cues are used online by interactants in a conversation.

[De Ruiter et al. \(2006\)](#) argued that one reason why intonational cues may not be used for turn-end projection is that they occur too late in the current speaker’s turn to allow the next speaker sufficient language planning time. To investigate the role of lexico-syntactic and intonational cues in turn-end projection, they conducted an online experiment with isolated turns from a spontaneous Dutch corpus. Participants were asked to listen to each turn and press a button in anticipation of its end. Turns were presented in three different conditions: in their original form, with flattened pitch, and in a low-pass-filtered condition that obscured lexico-syntactic information. Interestingly, it was found that button-press times for stimuli in which the pitch had been flattened were as accurate as button-press times for unmanipulated turns. In contrast, when the words of the turn were made unidentifiable, participants tended to press the button less accurately (i.e., earlier) with respect to the end of the turn than when lexical information was preserved. The authors concluded that lexico-syntactic information is necessary for turn-end projection, and that it is possibly sufficient, since intonation did not seem to be necessary for this task in their experiment.

Another recent online experimental study ([Keitel, Prinz, Friederici, von Hofsten, & Daum, 2013](#)) used similar stimulus types (dialogs with either normal or flattened pitch) to investigate the role of lexico-syntactic and intonational information in the acquisition of turn-taking skills by German children. Adult and child participants looked at videos of dialogs while their eye movements were tracked, providing potential evidence of anticipation of turn-ends by early switch of gaze to the next speaker. In concordance with [De Ruiter et al. \(2006\)](#), adults did not anticipate turns better in the dialogs with normal pitch than in the dialogs with flattened pitch. Children younger than 3;0 did not reliably anticipate turn-ends, but 3-year-olds could. However, unlike the adults, these children performed better in dialogs including pitch. The authors concluded that children anticipate turns in conversations in an adult-like manner only after they develop a sophisticated understanding of language.

If it is correct that lexico-syntactic information is sufficient for the accurate timing of responses, then the following question remains: how do listeners know, at a given syntactic completion point in the middle of a turn, whether the end of the turn has been reached or not? For example, the sequence of words ‘are you a student’ could be a complete question, but could also be the beginning of a longer question such as ‘are you a student at this university?’. This is an acute problem for lexico-syntactic models of turn-end projection, because, as [Sacks et al. \(1974\)](#) noted, any syntactic phrase can constitute an appropriate full turn, and thus there are many potential turns within actual turns (a problem parallel to the problem of word-recognition, given that most words contain other words; [Cutler, 2012](#): 48–50). One possibility is that the wider discourse context of the turn determines which syntactic completion point is a plausible turn end. Since [De Ruiter et al. \(2006\)](#) used isolated turns extracted from different conversations and presented them randomly to participants, this cannot be the case in their study. In the case of [Keitel et al. \(2013\)](#), on the other hand, this might well have been the case, since participants in this study listened to conversations.

However, it may also be that [De Ruiter and colleagues’](#) and [Keitel and colleagues’](#) conclusions about the role of intonation in turn-taking are not correct. In order to control for the role of intonation in turn-end projection, both groups of researchers flattened the pitch in their stimuli, but left all other cues to intonational phrasing intact, in particular the lengthening commonly found at the end of intonational phrases (e.g., [Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992](#); [Gussenhoven & Rietveld, 1992](#); [Turk & Shattuck-Hufnagel, 2007](#)). Since

Download English Version:

<https://daneshyari.com/en/article/7532907>

Download Persian Version:

<https://daneshyari.com/article/7532907>

[Daneshyari.com](https://daneshyari.com)