Journal of Memory and Language 90 (2016) 1–13

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Evidence for the influence of syntax on prosodic parsing ☆

Andrés Buxó-Lugo *, Duane G. Watson

University of Illinois at Urbana-Champaign, United States

ARTICLE INFO

Article history:
Received 27 June 2015
revision received 25 February 2016

Keywords:
Psycholinguistics
Language processing
Prosody

ABSTRACT

We investigate whether expectations based on syntactic position influence the processing of intonational boundaries. In a boundary detection task, we manipulated (a) the strength of cues to the presence of a boundary and (b) whether or not a location in the sentence was a plausible location for an intonational boundary to occur given the syntactic structure. Listeners consistently reported hearing more boundaries at syntactically licensed locations than at syntactically unlicensed locations, even when the acoustic evidence for an intonational boundary was controlled. This suggests that the processing of an intonational boundary is a product of both acoustic cues and listener expectations.

Published by Elsevier Inc.

Introduction

In this paper, we investigate the types of information listeners use to parse prosodic structure. An important part of parsing prosodic structure is detecting intonational boundaries, which are used to group utterances into smaller constituents that sometimes reflect the syntactic structure of spoken sentences (Cooper & Paccia-Cooper, 1980; Ferreira, 1993; Watson & Gibson, 2004). These boundaries are signalled by pauses, changes in F0 contours, and pre-boundary lengthening, among other cues (e.g., Klatt, 1975; Ladd, 2008; Pierrehumbert & Hirschberg, 1990; Turk & Shattuck-Hufnagel, 2007). Listeners, in turn, can use intonational boundaries to decipher the linguistic structure of a message, as in the case of syntactically ambiguous sentences (Schafer, Speer, & Warren, 2005; Snedeker & Trueswell, 2003).

However, few studies have explored how listeners build their representation of utterances' prosodic structure. Current models that aim to shed light on the relationship between prosody and other levels of representation tend to be unidirectional, often focusing on how prosody can guide the interpretation of other constructs such as syntax (e.g., Kjelgaard & Speer, 1999; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Schafer, Speer, Warren, & White, 2000). For example, Schafer (1997) proposes the following relationship between prosody and syntax: "the prosodic representation that is constructed by the phonological component is passed on to higher-level modules in the same way that lexical information is made available to them" (p. 6) such that prosodic information is "part of the computational vocabulary of the syntactic and semantic/pragmatic processing modules" (p. 6). According to such models of prosodic parsing, listeners build prosodic representations from the acoustic cues, and then use these constructs to guide their interpretation of higher-level structures. However, it is possible that prosodic parsing is more interactive, or bi-directional. In such a model, information from higher-level structures and listener expectations, along with acoustic cues, guide the parsing of prosodic structure. This study investigates whether the detection of intonational boundaries is wholly driven by acoustic features in the speech signal, or whether input

☆ *Author's note:* This project was supported by Grant Number R01DC008774 from the National Institute on Deafness and Other Communication Disorders, Grant Number T32-HD055272 from the National Institutes of Health, and a grant from the James S. McDonnell Foundation.

* Corresponding author at: Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St., Champaign, IL 61820, United States.

E-mail address: buxo2@illinois.edu (A. Buxó-Lugo).

http://dx.doi.org/10.1016/j.jml.2016.03.001
0749-596X/Published by Elsevier Inc.

from the syntactic context influences listeners' interpretations.

Intonational boundaries provide an ideal opportunity to investigate listeners' parsing of prosodic structure because of the close link between syntactic boundaries and intonational phrasing (Nespor & Vogel, 1986). Many studies have explored the connection between syntactic and prosodic structures. For example, constraints such as Align-XP (Selkirk, 1986, 1995) and Wrap-XP (Truckenbrodt, 1999) argue that there are grammatical constraints that govern the mapping between syntactic structure and prosodic boundaries, resulting in a preference to produce intonational boundaries at syntactic boundaries. Similarly, algorithmic approaches that predict where boundaries occur make use of syntactic information, such as the length of syntactic constituents and the relationship between syntactic dependents (e.g., Cooper & Paccia-Cooper, 1980; Ferreira, 1988; Watson & Gibson, 2004).[1] Studies have also found that listeners can accurately locate syntactic boundaries based on prosodic cues alone (Beach, 1991; de Pijper & Sanderman, 1994; Streeter, 1978). Lastly, listeners use prosodic boundaries to resolve syntactic ambiguities (e.g., Kjelgaard & Speer, 1999; Kraljic & Brennan, 2005; Lehiste, 1973; Price et al., 1991; Schafer, 1997; Schafer et al., 2005; just to name a few). For example, Snedeker and Trueswell (2003) examined productions of sentences with attachment ambiguities such as: "Tap the frog with the flower," where "flower" could be used as an instrument used for tapping, or the prepositional phrase could be interpreted as a modifier of "the frog." Speakers who were aware of the ambiguity produced intonational boundaries that disambiguated the syntax (after the verb for a modifier interpretation, and after the noun "frog" for an instrument interpretation). Critically, listeners used these cues to carry out the correct instruction. This suggests that listeners can accurately parse the syntactic structure of a sentence if intonational boundary cues are provided.

Given that there is a strong correlation between intonational boundaries and syntactic structure, it is possible that listeners not only use prosodic structure to make inferences about syntactic structure, but also use syntactic structure to make inferences about prosodic structure. This type of interaction between processing systems is ubiquitous in language processing. For example, perception studies have found that syntax influences where listeners report hearing bursts of noise (Garrett, Bever, & Fodor, 1966), that morphological context affects the perception of ambiguous phonemes (Ganong, 1980), and that top-down knowledge of the speech signal affects whether degraded speech is perceived as speech at all (Remez, Rubin, Pisoni, & Carrell, 1981). More recent studies (e.g., Kim & Osterhout, 2005; Tabor & Tanenhaus, 1999) have proposed parallel-process models where processing streams for semantic interpretation and syntactic interpretation are independent but still interact through cross-talk or attraction. According to some of these models, each processing system (e.g., syntactic processing system, semantic processing system, etc.) attempts to reach likely interpretations of a stimulus based on their input; however, if a processing system does not have sufficient evidence for converging on an interpretation, it is likely to be influenced by other processing streams.

Given that interaction between levels of processing is ubiquitous in the language comprehension system, it would be surprising if listener expectations did not influence their interpretation of prosody. Some studies have found that prosodic information from earlier in an utterance influences how listeners segment words (e.g., Brown, Salverda, Dilley, & Tanenhaus, 2011; Dilley, Mattys, & Vinke, 2010) and how they interpret lexical stress (Brown, Salverda, Dilley, & Tanenhaus, 2012) later in an utterance. Also work by Bishop (2012) suggests that expectations about discourse structure can influence the perception of acoustic prominence. This is further supported by work by Cole, Mo, and Baek (2010), where untrained listeners prosodically transcribed speech from the Buckeye corpus. In their study, both vowel duration and syntactic context were correlated with boundary reports, each factor independent of the other. In fact, syntactic context was the best predictor of boundary detection, suggesting that listeners' judgments were influenced by their expectations of where boundaries should occur.

However, Cole et al. (2010) did not directly manipulate listener expectations of intonational boundaries. Corpus analyses are a useful tool for detecting correlations, such as the one found between syntactic context and boundary detection in Cole et al. (2010). However, a challenge for these approaches is controlling for other potential variables that might be confounded with the theoretical construct of interest. For example, it is possible that boundary detection was driven by acoustic cues that were not accounted for in the analyses. This makes it difficult to definitively establish that syntactic expectations are driving the detection of intonational boundaries. An advantage of investigating this issue through an experimental design is that these potential confounds can be more precisely controlled with the goal of understanding whether syntactic context alone drives the perception of prosody. That is our goal here. If prosodic parsing is guided by expectations, one would expect a greater tendency to report hearing an intonational boundary in locations in which they typically occur. In the current study, we directly manipulated the acoustic evidence for intonational boundaries and the syntactic context in which these possible boundaries were located. By manipulating word duration, F0 contour, and pause duration of potential boundary sites, we were able to make these locations sound more or less boundary-like. These manipulated words were placed at points at which boundaries were syntactically licensed and at points at which boundaries were syntactically unlicensed, allowing us to independently manipulate acoustic and syntactic cues to the presence of a boundary. Examining this question in the context of a controlled experiment allows us to see the effects of syntax on prosodic parsing while controlling for acoustic factors, and vice versa. Furthermore, by individually manipulating acoustic cues and syntactic context, we can observe how these factors interact. For example: how strong do the acoustic cues have to be for

[1] These apparent effects of constituent length have also been conceptualized as effects of the phonological length of consistuents (see Jun & Bishop, 2015 for a discussion).